Complexity of regulatory logic and dynamics of Boolean models of biological networks

By
Priyotosh Sil
LIFE10202104002

The Institute of Mathematical Sciences
Chennai

A thesis submitted to the
Board of Studies in Life Sciences
In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE



July 2026

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Priyotosh Sil** entitled: "*Complexity of regulatory logic and dynamics of Boolean models of biological networks*" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Chair - Prof. Sitabhra Sinha	Date: 30/07/2026	In-Person
Supervisor/Convener - Prof. Areejit Samal	Date: 30/07/2026	In-Person
Member 1 - Prof. Rahul Siddharthan	Date: 30/07/2026	In-Person
Member 2 - Prof. Sandeep Choubey	Date: 30/07/2026	In-Person
Member 3 - Prof. Olivier C. Martin	Date: 30/07/2026	Online
External Examiner - Prof. R.K. Brojen Singh	Date: 30/07/2026	In-Person

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 30/07/2026

Place: Chennai

Supervisor

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



Priyotosh Sil

Declaration

I, hereby declare that the investigation presented in this thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University.



Priyotosh Sil

CERTIFICATION ON ACADEMIC INTEGRITY

Undertaking by the Student

1. I **Priyotosh Sil**, HBNI Enrolment No. **LIFE10202104002** hereby undertake that the Thesis, titled "***Complexity of regulatory logic and dynamics of Boolean models of biological networks***" is prepared by me and is the original work undertaken by me.

2. I also hereby undertake that this document has been duly checked through a plagiarism detection tool and the document is found to be plagiarism free as per the guidelines of the Institute/UGC.

3. I am aware and undertake that if plagiarism is detected in my thesis at any stage in the future, suitable penalty will be imposed as per the guidelines of the Institute/ UGC.

Priyotosh Sil

(30-07-2026)
Signature of the Student with date

Endorsed by the Thesis Supervisor:

I certify that the thesis written by the researcher is plagiarism free as mentioned above by the student.

Areejit Samal

प्रोफेसर
PROFESSOR
गणितीय विज्ञान संस्थान, चेन्नई-600 113
The Institute of Mathematical Sciences, Chennai-600 113

(30/07/2026)

Signature of the Thesis Supervisor with Date and Name: Areejit Samal
Designation: Professor G
Department/ Centre: Computational Biology group
Name of the CI/OCC: The Institute of Mathematical Sciences (IMSc), Chennai

List of Publications arising from the thesis

Journals

Published

1. *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph*,

   P. Sil[†], A. Subbaroyan[†], S. Kulkarni, O. C. Martin[⋆] and A. Samal[⋆],

   Briefings in Bioinformatics, 25: bbae150 (2024).

   https://doi.org/10.1093/bib/bbae150

2. *Choices of regulatory logic class modulate the dynamical regime in random Boolean networks*,

   P. Sil, S. Mitra, O. C. Martin[⋆] and A. Samal[⋆],

   Chaos, Solitons & Fractals, 195: 116231 (2025).

   https://doi.org/10.1016/j.chaos.2025.116231

3. *A systematic evaluation of simple threshold-based Boolean rules for modeling gene regulatory networks*,

   P. Sil, O. C. Martin[⋆] and A. Samal[⋆],

   Journal of Statistical Mechanics: Theory and Experiment, 023502 (2026).

   https://doi.org/10.1088/1742-5468/ae4588

[[†] Joint-first authors; [⋆] Corresponding author(s)]

List of Publications not included in the thesis

Journals

Published

1. *Leveraging developmental landscapes for model selection in Boolean gene regulatory networks*,

   A. Subbaroyan, **P. Sil**, O. C. Martin[⋆] and A. Samal[⋆],

   Briefings in Bioinformatics, 24: bbad160 (2023).

   https://doi.org/10.1093/bib/bbad160

2. *Preponderance of generalized chain functions in reconstructed Boolean models of biological networks*,

   S. Mitra[†], **P. Sil**[†], A. Subbaroyan[†], O. C. Martin[⋆] and A. Samal[⋆],

   Scientific Reports, 14: 6734 (2024).

   https://doi.org/10.1038/s41598-024-57086-y

3. *IMPPAT: A digital phytochemical atlas enabling computational approaches to traditional Indian knowledge-centered herbal drug discovery*,

   P. Sil, R. Tiwari, S. P. Baskaran and A. Samal[⋆],

   International Journal of Ayurveda Research, 6: 387-398 (2025).

   https://doi.org/10.4103/ijar.ijar_328_25

4. *Computational investigation of single herbal drugs for diabetes and obesity using knowledge graph and network pharmacology*,

   P. Sil[†], R. Tiwari[†], V. Garisetti, S. P. Baskaran, F. H. Dhanaseelan, S. Srivastava and A. Samal[⋆],

   Computational Biology and Chemistry, 124: 109194 (2026).

   https://doi.org/10.1016/j.compbiolchem.2026.109194

Submitted

5. *An integrated resource for systems-level analysis of aging hallmarks and associated genes*,

   R. Tiwari, M. Balaji, N. Chivukula, **P. Sil** and A. Samal[⋆],

   bioRxiv 2026.05.29.728838 (2026)

   https://doi.org/10.64898/2026.05.29.728838

[[†] Joint-first authors; [⋆] Corresponding author(s)]



Priyotosh Sil

Oral or Poster presentations

1. Poster presentation titled *Critical assessment of measures to quantify relative stability of states in Boolean models of developmental gene regulatory networks* at the Conference on Nonlinear Systems and Dynamics (CNSD) 2022 held at the Indian Institute of Science Education and Research (IISER) Pune, India, from 15-18 December 2022.
2. Poster presentation titled *Leveraging developmental landscapes for model selection in Boolean gene regulatory networks* at the IMSc60 - Celebrating 60 Years of Creative Science held at the Institute of Mathematical Sciences (IMSc), Chennai, India, from 2-5 January 2023.
3. Poster presentation titled *Leveraging developmental landscapes for model selection in Boolean gene regulatory networks* at the Big Data Algorithms for Biology (BD-Bio) 2023 held at the Indian Institute of Science (IISc), Bengaluru, India, from 2-3 June 2023.
4. Poster presentation titled *Leveraging developmental landscapes for model selection in Boolean gene regulatory networks* at the Network Biology Day held at the Institute of Mathematical Sciences (IMSc), Chennai, India, on 20 July 2023.
5. Poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at the Perspectives in Nonlinear Dynamics (PNLD) 2023 held at the Indian Institute of Technology (IIT) Madras, India, from 1-4 August 2023.
6. Oral presentation titled *Adapting concepts from cellular automata to explore bushiness and convergence in Boolean network dynamics* and poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph* at the Modelling and Tackling Complex Biological Systems held at the Institute of Mathematical Sciences (IMSc), Chennai, India, from 13-14 October 2023.
7. Poster presentation titled *Biologically meaningful regulatory logic enhances the*

convergence rate in Boolean networks and bushiness of their state transition graphs at the Stochasticity and Plasticity in living Systems held at Sakleshpur, Karnataka, India (organized by National Centre for Biological Sciences (NCBS), Bengaluru), from 14-18 November 2023.

8. Poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at the Contemporary Perspectives in Computational Biology held at the Institute of Mathematical Sciences (IMSc), Chennai, India, from 19-20 February 2024.
9. Oral presentation titled *Adapting concepts from cellular automata to explore bushiness and convergence in Boolean network dynamics* and poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at the Theoretical Approaches in Cancer Progression and Treatment held at the International Centre for Theoretical Sciences (ICTS), Bengaluru, India, from 11-22 March 2024.
10. Oral presentation titled *Effect of biologically meaningful functions on the bushiness and convergence of Boolean state transition graphs* at the Symposium on Synthetic and Systems Biology (BioSynSys) 2024 held at the University of Montpellier, Montpellier, France, from 5-7 June 2024.
11. Poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at the European Conference on Mathematical and Theoretical Biology (ECMTB'24) held at the University of Castilla La Mancha (UCLM), Toledo, Spain, from 22-26 July 2024.
12. Oral presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at the NetSciX 2025 International School and Conference on Network Science held at the Indian Institute of Technology (IIT) Indore, Indore, India, from 15-17 January 2025.

13. Flash talk titled *Not all logic rules are equal: Biologically meaningful rules shape dynamics differently from random ones* and poster presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph* at the 47^{th} Indian Biophysical Society Meeting (IBS 2025) held at the Indian Institute of Technology (IIT) Madras, India, from 6-9 March 2025. (**Received a Flash talk prize**).
14. Poster presentation titled *Simple Threshold-based Boolean Rules Fall Short in Capturing Biological Regulatory Network Dynamics* at Modelling and Tackling Complex Biological Systems 2.0 held at he Institute of Mathematical Sciences (IMSc), Chennai, India, from 16-18 February 2026.
15. Poster presentation titled *Biologically Meaningful Regulatory Logic Enhances the Convergence Rate in Boolean Networks and Bushiness of Their State Transition Graph* at 2nd Bonn Conference on Mathematical Life Sciences held at Bonn Center for Mathematical Life Sciences (BCML), Bonn, Germany, from 16-19 March 2026.
16. Oral presentation titled *Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graphs* at 4th International Conference on Recent Advances in Applied Mathematics held at the Indian Institute of Technology (IIT) Hyderabad, India, from 6-8 July 2026.



Priyotosh Sil

Acknowledgements

I would like to begin by thanking my advisor, Prof. Areejit Samal, for his guidance, support, and unwavering encouragement throughout my PhD. Over the years, his mentorship has been central to my growth as a researcher, shaping not only the direction of this thesis but also the way I think about and approach scientific problems. I am particularly grateful to him for introducing me to a rich and diverse range of research questions, many of which form the core of this work. I also thank him for the many discussions and constructive feedback that helped refine these ideas over time. His emphasis on clarity, rigor, and curiosity has left a lasting impression on how I engage with research. I am also deeply thankful for the intellectual freedom he provided me and for the opportunities he created—from introducing me to Prof. Olivier Martin, to enabling me to present my work at conferences, visit the University of Paris-Saclay, and interact with the broader scientific community. His dedication to research and commitment to mentoring students have been a constant source of inspiration to me.

I would also like to express my sincere gratitude to Prof. Olivier Martin for his collaboration and support. It has been a privilege to collaborate with him on the research papers that form the basis of this thesis as well as on other projects beyond its scope, and I have greatly benefited from his insightful suggestions and thoughtful discussions. I am especially grateful for his warm hospitality during my research visit to the University of Paris-Saclay, where he ensured that my stay was both productive and enjoyable and was always willing to help whenever needed.

I have also been fortunate to work with several colleagues whose contributions have been instrumental to the research presented in this thesis as well as to other publications beyond its scope. I am grateful to A. Subbaroyan for his support during the early stages of my PhD, when he introduced me to many important aspects of research and computation. I thank him for his exceptional help in designing projects and for proposing a useful local measure used in Chapter 5. We have collaborated on three papers together, one of which forms Chapter 5 of this thesis. I would also like to thank S. Kulkarni for his exceptional

and efficient computational support, which significantly accelerated the study reported in Chapter 5. I also thank S. Mitra for collaborating on two research projects, including the work presented in Chapter 4, and another collaborative study.

I also worked on projects beyond the main scope of this thesis, which broadened my exposure to different areas and approaches. I would like to thank Prof. Smita Srivastava for her collaboration on one of these projects and for the fruitful discussions we had. I also acknowledge Dr. V. Garisetti and F. H. Dhanaseelan for their collaboration on this work. I am particularly grateful to R. Tiwari, with whom I worked closely on these projects and shared many productive discussions during the course of the work. I also thank S. P. Baskaran for her collaboration on these studies.

I would also like to express my sincere thanks to my doctoral committee members– Prof. Sitabhra Sinha, Prof. Areejit Samal, Prof. Rahul Siddharthan, Prof. Sandeep Choubey, and Prof. Olivier C. Martin–for their valuable comments and constructive suggestions during the course of my PhD. I also wish to thank everyone who taught me during the PhD coursework, including Prof. Sandeep Choubey, Prof. Vasudharani Devanathan, Prof. Prabuddha Ganguli, Prof. Sankaran Krishnaswamy, Prof. Basanta Kumar Nayak, Prof. Areejit Samal, Prof. Rahul Siddharthan, Dr. Vipin Singh, Prof. Vasudeva Siruguri, Prof. Sitabhra Sinha, Prof. B. S. Tomar, and Prof. Satyavani Vemparala.

I gratefully acknowledge IMSc and the Department of Atomic Energy (DAE) for providing financial support and an excellent research environment during my PhD. I would also like to thank the administrative staff and members of the computer committee for their constant assistance. I also sincerely appreciate the efforts of everyone working in the canteen, housekeeping, civil, security, and electrical departments, whose dedication makes IMSc a pleasant place to work and made my stay in the IMSc hostel comfortable and homely.

I would also like to sincerely thank several researchers with whom I had the opportunity to interact during conferences, visits to IMSc, and other academic occasions. In particular, I am grateful to Prof. Deepak K. Agrawal, Prof. Amit Apte, Prof. Sangram

Bagh, Prof. Anshu Bhardwaj, Prof. Brato Chakrabarti, Prof. Claudine Chaouiya, Prof. Samrat Chatterjee, Prof. Riddhiman Dhar, Prof. Madhulika Dixit, Prof. Chetan Gadgil, Prof. Tomas Gedeon, Prof. Gourab Ghoshal, Prof. Silvia Grigolon, Prof. Stephan Grill, Prof. Sanjay Jain, Prof. Claus Kadelka, Prof. Mustafa Khammash, Prof. Martin Lenz, Prof. Herbert Levine, Prof. James Locke, Prof. Philip Maini, Prof. Adilson E. Motter, Prof. Mitali Mukerji, Prof. Manikandan Narayanan, Prof. Debnath Pal, Prof. Sagar Pandit, Prof. Dinakar Ramakrishnan, Prof. Karthik Raman, Prof. Ramakrishna Ramaswamy, Prof. Ram Rup Sarkar, Prof. Anupam Kumar Singh, and Prof. Denis Thieffry for their insightful discussions and interactions. I was also fortunate to present my research to several other faculty members during their visits to IMSc and at conferences elsewhere. I would also like to thank my master's thesis advisor, Prof. Nikhil Pal, who first introduced me to the experience of doing research and motivated me to pursue a research career. I am grateful to all of them, as these interactions have influenced the way I think about science.

I would like to thank all the current and past lab members at IMSc for creating a supportive and stimulating environment that made my PhD journey both enjoyable and memorable. I am grateful to my past lab mates Ajay, Ajaya, BS Karthikeyan, Geetha, Gokul, Ishwarya, Janani, Jayalakshmi, Kaustav, Kishan, Kundhanathan, Pavithra, Pritam, Priya Darshini, Priya Shree, Sabrina, Shashank, Shreelekshmi, Suchetana, Sudharsan, Vinayak, Vivek, and Yasharth for the many conversations, shared experiences, and moments that made life in the lab vibrant. Special thanks to Ajaya for his friendly, big-brother-like attitude that made it easy to approach him anytime. I deeply appreciated his enthusiasm in discussing a wide range of topics and for helping me learn several new computational tools and techniques. I also thank current lab members Camelia, Harshini, Jaisubasri, Madhumita, Nikhil, Priya, Rahul, Shrish, Shreyes, and Vasavi for creating a friendly and encouraging atmosphere in the lab. I will always be grateful to Madhumita, Nikhil, Priya, and Shreyes for their unwavering support, for creating many fun moments, and for the numerous academic and non-academic discussions we shared. Special thanks to Camelia, Madhumita, Mridhula, and Vasavi for their help with proofreading the final version of my

thesis. I would also like to acknowledge Rohit, Vishwajeet, and Sushant for the many late-night discussions and arguments over science, music, and philosophy, as well as for playing badminton together. Finally, I thank Velu for being one of the best roommates during the early period of my PhD and for a discussion on a mathematical problem that eventually led to a proof included in Appendix B of this thesis. I would also like to thank my friends from my undergraduate and master's days–Abhishek, Amal, Asim, Bikash, Bhaskar, Bikram, Debjani, Madhumita, Rachana, Ranjana, Santu, Shankha, Shubhajit, Somnath, Soumya, Souvik, Sourav, Subhashis, Suhita, Taniya, and Vivek. I am equally grateful to my childhood friends Bijay, Prasun, Suman, and Shubhajit.

Lastly, I would like to express my deepest gratitude to my family for their constant love, support, and encouragement throughout this journey. I owe the most to my father, Prasanta Sil, and my mother, Kanika Sil, whose sacrifices, patience, and faith in me have been the driving force behind my journey. I also remember with great affection my late grandfather, Trilochan Sil, who showered me with immense love and encouragement. I would like to thank my younger brother, Pranotosh Sil, with whom I have shared countless joyful moments over the years. I am also deeply grateful to my uncles, aunts, and cousins, who have always encouraged and supported me. I would also like to thank Madhumita, who has been a constant source of support for nearly eight years. From sharing the same classroom during our university days to working together as colleagues in the lab, her encouragement and understanding have helped me navigate many challenging phases of this journey. I am also grateful to her parents for their kindness and encouragement. To conclude, I express my heartfelt gratitude to all those who have supported me along the way. I am also grateful to Lord Gopinath Jiu for his blessings and guidance throughout this journey.



Priyotosh Sil

Contents

List of Figures **i**

List of Tables **xii**

Abstract **xv**

1 Background and motivation **1**

1.1 Thesis organization 10

2 Boolean modeling of gene regulatory networks and associated logical update rules **15**

2.1 Boolean modeling of gene regulatory networks 16

2.2 Random Boolean networks 17

2.3 Reconstructed Boolean models of biological systems 19

2.4 Background on Boolean functions 22

2.4.1 Negation of Boolean functions 24

2.4.2 Permutations of Boolean functions 25

2.5 Biologically meaningful classes of Boolean functions 26

2.5.1 Effective functions (EFs) 26

2.5.2 Effective and unate functions (EUFs) 27

2.5.3 Read-once functions (RoFs) 27

2.5.4 Nested canalyzing functions (NCFs) 27

2.5.5 Link operator functions (LOFs) 29

2.6 Complexity of Boolean functions 29

2.6.1 Boolean complexity 29
2.6.2 Average sensitivity 30
2.7 Sampling different types of Boolean functions 31
2.7.1 Effective functions (EFs) 31
2.7.2 Effective and unate functions (EUFs) 32
2.7.3 Read-once functions (RoFs) 38
2.7.4 Nested canalyzing functions (NCFs) 39
2.7.5 Link operator functions (LOFs) 39

3 Stability measures for Boolean network dynamics 41
3.1 Derrida's damage spreading and generalizations 42
3.2 Stability measures based on properties of the attractors 43
3.3 From polynomial approximations of regulatory rules to the "approximability" of network dynamics 45
3.4 Cellular automata inspired stability measures for Boolean networks 46
3.4.1 Global measures of bushiness and convergence of the STG 47
3.4.2 The Z-parameter in Cellular Automata 50

4 Choices of regulatory logic class modulate the dynamical regime in random Boolean networks 53
4.1 Ensembles of RBNs: constraining network topologies and restricting Boolean functions to a class 55
4.2 The Jensen–Shannon distance and statistical tests for comparing distributions 56
4.3 Stability is enhanced along our hierarchy of Boolean function classes . . . 58
4.4 The dependence of stability on network average degree is specific to the different classes of functions 63
4.5 The stability hierarchy is robustly maintained across network size 67
4.6 Discussion 68

5 Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph 71

5.1 Generating ensembles of biologically plausible models of ten published Boolean GRNs for living systems 73

5.2 Biologically meaningful functions lead to highly bushy and convergent STGs 75

5.3 Descriptors of local dynamics 79

5.3.1 Adaptation of Z-parameter in cellular automata to Boolean networks 82

5.3.2 Network sensitivity as a descriptor of local dynamics 90

5.4 Biologically meaningful BFs have lower local dynamics descriptor values 90

5.5 Descriptors of local dynamics can serve as good proxies of bushiness of STGs 91

5.5.1 Correlation between the descriptors of local dynamics and G-density 91

5.5.2 Performance evaluation using AUROC and AUPRC 93

5.6 The average convergence rate of trajectories originating at random states is a good proxy for the average convergence rate of trajectories originating at GoE states 95

5.7 Discussion 96

6 A systematic evaluation of simple threshold-based Boolean rules for modeling gene regulatory networks 99

6.1 Threshold function and its subtypes 101

6.2 Representation of threshold majority rules as Boolean functions and their comparison to nested canalyzing functions 104

6.3 Threshold majority rules exhibit higher averaged Boolean complexity and sensitivity than nested canalyzing functions 107

6.4 Enrichment analysis of threshold majority rules in the empirical dataset . 111

6.4.1 Empirical dataset 111

6.4.2 Relative enrichment and statistical significance 111

6.4.3 Threshold majority rules are underrepresented in the BBM empirical dataset . . . 112
6.5 Choosing threshold majority rules over nested canalyzing functions impairs recovery of the dynamical outcomes in biological Boolean networks 116
6.5.1 Imposing structural and dynamical constraints on biological models 116
6.5.2 An attractor-based scoring framework to compare a published model to an alternative model . . . 116
6.5.3 Threshold majority rules perform poorly in recovering the biologically meaningful attractors . . . 119
6.5.4 The basin fractions of biologically meaningful attractors when using threshold majority rules are far from the desired target . . . 120
6.5.5 Substituting non-nested canalyzing functions with suitable nested canalyzing functions in published models can yield superior dynamical outcomes. . . . 123
6.6 Comparison of dynamical properties in random networks when using different subtypes of threshold functions . . . 124
6.6.1 Ising majority rules drive network dynamics toward a chaotic regime 124
6.7 Discussion . . . 126

7 Summary and future outlook **131**
7.1 Summary . . . 131
7.2 Future outlook . . . 135

A Supplementary Material for Chapter 4 **139**

B Supplementary Material for Chapter 5 **211**
B.1 A brief overview of the ten published Boolean network models used in this study . . . 211
B.1.1 Root Stem Cell Niche GRN (RSCN-GRN) . . . 211
B.1.2 Epithelial-Mesenchymal Transition GRN (EMT-GRN) . . . 212

B.1.3 NeoCortex Developmental GRN (NCD-GRN) 212
B.1.4 Early Heart Development GRN (EHD-GRN) 212
B.1.5 Myeloid Differentiation GRN (MD-GRN) 213
B.1.6 T helper cell differentiation GRN (THC-GRN) 213
B.1.7 EGFR signalling pathway (EGFR-GRN) 213
B.1.8 Epithelial-Mesenchymal Transition (with Senescence) (EMT-Senescence-GRN) 214
B.1.9 Flower Organ Specification GRN (FOS-GRN) 214
B.1.10 Gonadal Sex Determination GRN (GSD-GRN) 214
B.2 Properties of Z-parameters 215
B.3 Results on non-equivalent permutations of the $k!$ permutations of Boolean functions 220
B.3.1 Preliminaries 220
B.3.2 Two theorems on non-equivalent permutations of the $k!$ permutations of BFs and their proofs 221
B.4 Supplementary tables for Chapter 5 226
B.5 Supplementary figures for chapter 5 257

C Supplementary Material for Chapter 6 **305**
C.1 Properties of the IMR and BMR sets 305
C.2 Generation of models with nested canalyzing functions preserving biological attractors 328
C.3 Supplementary tables for Chapter 6 330
C.4 Supplementary figures for Chapter 6 336

References **344**

List of Figures

2.1 Different representations of Boolean functions 23

2.2 Visual illustration of how monotonicity forces propagation of the color on vertices of the Boolean hypercube 35

2.3 The distribution of positive monotone effective and unate functions (EUFs) sampled using the proposed EUF sampling algorithm 36

2.4 Generating a random UF with a given sign combination that satisfies given fixed point constraints 37

3.1 Measures of 'bushiness' and 'convergence' of STGs 48

4.1 Summary figure of the design for the study in Chapter 4 57

4.2 Comparison of mean values of h^{∞} when using different classes of BFs on identical networks 60

4.3 Distributions of stability observables when using different classes of BFs . 61

4.4 Mean values of 12 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 12$ 64

4.5 Distributions of L_W (the weighted average attractor length) for R-R . . . 65

4.6 Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 4$ 69

5.1 STGs for two different choices of Boolean rules for the biological network MD-GRN 78

5.2 Derrida plot showing the progression of various Hamming distances over one time step for Models A and B shown in Figure 5.1 80

5.3 Distribution of global measure values for different ensembles generated using network structures from four published Boolean GRNs 81
5.4 Calculation of Z-parameters of Boolean functions in BN models 85
5.5 The fraction of non-equivalent permutations for different types of BFs . . 89
5.6 Distribution of the values of the descriptors of local dynamics for different ensembles generated using network structures from four published Boolean GRNs 92
5.7 Correlation heat map between descriptors of local dynamics and global measures (bushiness and convergence rate) for ensembles generated using network structures from four published Boolean GRNs 94

6.1 The true regulators used by a TMR can differ from the explicit ones, leading to failures in modeling of self regulation 108
6.2 TMRs have high averaged Boolean complexity and averaged sensitivity compared to NCFs 110
6.3 Enrichment of EThFs and NCFs, and depletion of IMRs and non-NCF BMRs, in the BBM dataset 113
6.4 Flowchart of the multi-step filtration pipeline applied to the models in the BBM dataset 115
6.5 TMR-based models generally fail to recover biological attractors 121
6.6 JS distance from gold standard reveals poor preservation of basin size distributions of biological attractors by TMR-based models 122
6.7 BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the P-P network topology with $N = 12$. . . . 127

7.1 Trends of different stability measures with varying N and k 134

A.1 Comparison of mean values of δ when using different classes of BFs on identical networks 140
A.2 Comparing the distributions of h^{∞} against δ and h^{∞} against L_w 141

A.3 The distributions of h^{∞} against δ for R-P networks 142
A.4 The distributions of h^{∞} against δ for R-R networks 143
A.5 The distributions of h^{∞} against δ for P-P networks 144
A.6 The distributions of ϕ against δ for R-P networks 145
A.7 The distributions of ϕ against δ for R-R networks 146
A.8 The distributions of ϕ against δ for P-P networks 147
A.9 Distributions of number of attractors for R-P networks 148
A.10 Comparing distributions for the number of attractors when using different classes of BFs within the R-P networks 149
A.11 Distributions of fraction of fixed point attractors for R-P networks 150
A.12 Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the R-P networks 151
A.13 Distributions of fraction of states in the basins of fixed points for R-P networks 152
A.14 Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the R-P networks 153
A.15 Distributions of average attractor lengths for R-P networks 154
A.16 Comparing distributions for the average attractor lengths when using different classes of BFs within the R-P networks 155
A.17 Distributions of maximum attractor lengths for R-P networks 156
A.18 Comparing distributions for the maximum attractor lengths when using different classes of BFs within the R-P networks 157
A.19 Distributions of weighted average attractor lengths for R-P networks . . . 158
A.20 Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the R-P networks 159
A.21 Distributions of fraction of states on attractors for R-P networks 160
A.22 Comparing distributions for the fraction of states on attractors when using different classes of BFs within the R-P networks 161

A.23 Distributions of number of attractors for R-R networks 162
A.24 Comparing distributions for the number of attractors when using different classes of BFs within the R-R networks 163
A.25 Distributions of fraction of fixed point attractors for R-R networks 164
A.26 Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the R-R networks 165
A.27 Distributions of fraction of states in the basins of fixed points for R-R networks 166
A.28 Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the R-R networks 167
A.29 Distributions of average attractor lengths for R-R networks 168
A.30 Comparing distributions for the average attractor lengths when using different classes of BFs within the R-R networks 169
A.31 Distributions of maximum attractor lengths for R-R networks 170
A.32 Comparing distributions for the maximum attractor lengths when using different classes of BFs within the R-R networks 171
A.33 Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the R-R networks 172
A.34 Distributions of fraction of states on attractors for R-R networks 173
A.35 Comparing distributions for the fraction of states on attractors when using different classes of BFs within the R-R networks 174
A.36 Distributions of number of attractors for P-P networks 175
A.37 Comparing distributions for the number of attractors when using different classes of BFs within the P-P networks 176
A.38 Distributions of fraction of fixed point attractors for P-P networks 177
A.39 Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the P-P networks 178

A.40 Distributions of fraction of states in the basins of fixed points for P-P networks 179
A.41 Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the P-P networks 180
A.42 Distributions of average attractor lengths for P-P networks 181
A.43 Comparing distributions for the average attractor lengths when using different classes of BFs within the P-P networks 182
A.44 Distributions of maximum attractor lengths for P-P networks 183
A.45 Comparing distributions for the maximum attractor lengths when using different classes of BFs within the P-P networks 184
A.46 Distributions of weighted average attractor lengths for P-P networks . . . 185
A.47 Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the P-P networks 186
A.48 Distributions of fraction of states on attractors for P-P networks 187
A.49 Comparing distributions for the fraction of states on attractors when using different classes of BFs within the P-P networks 188
A.50 Distributions of mean approximation error (MAE) for R-P networks . . . 189
A.51 Distributions of mean approximation error (MAE) for R-R networks . . . 190
A.52 Distributions of mean approximation error (MAE) for P-P networks . . . 191
A.53 Mean values of 9 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 16$ 192
A.54 Mean values of 9 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 20$ 193
A.55 Mean values of our 12 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 12$ 194
A.56 Mean values of 9 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 16$ 195

A.57 Mean values of 9 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 20$ 196
A.58 Mean values of our 12 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 12$ 197
A.59 Mean values of 9 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 16$ 198
A.60 Mean values of 9 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 20$ 199
A.61 Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 2$ 200
A.62 Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 3$ 201
A.63 Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 5$ 202
A.64 Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 2$ 203
A.65 Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 3$ 204
A.66 Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 4$ 205
A.67 Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 5$ 206
A.68 Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 2$ 207
A.69 Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 3$ 208
A.70 Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 4$ 209

A.71 Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 5$ 210

B.1 Distribution of $\boldsymbol{G}$-density values for different ensembles generated using network structures from six published Boolean GRNs 258

B.2 Distribution of average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) for different ensembles generated using network structures from four published Boolean GRNs 259

B.3 Distribution of average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) for different ensembles generated using network structures from six published Boolean GRNs 260

B.4 Distribution of average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) for different ensembles generated using network structures from six published Boolean GRNs 261

B.5 Scatter plot between the G-density and average convergence rate of trajectories originating at *GoE* states for four published models 262

B.6 Scatter plot between the G-density and average convergence rate of trajectories originating at *GoE* states for six published models 263

B.7 Distribution of $\overline{Z}_{max}$ values for different ensembles generated using network structures from four published Boolean GRNs 264

B.8 Distribution of $\overline{Z}_{max}$ values for different ensembles generated using network structures from six published Boolean GRNs 265

B.9 Distribution of $\overline{Z}_{min}$ values for different ensembles generated using network structures from four published Boolean GRNs 266

B.10 Distribution of $\overline{Z}_{min}$ values for different ensembles generated using network structures from six published Boolean GRNs 267

B.11 Distribution of $\overline{Z}_{ave}$ values for different ensembles generated using network structures from four published Boolean GRNs 268

B.12 Distribution of $\overline{Z}_{ave}$ values for different ensembles generated using network structures from six published Boolean GRNs 269

B.13 Distribution of $\overline{Z}_{mid}$ values for different ensembles generated using network structures from six published Boolean GRNs 270

B.14 Distribution of $\overline{s}$ values for different ensembles generated using network structures from six published Boolean GRNs 271

B.15 Correlation heat map between the various descriptors of local dynamics on ensembles generated using network structures from four published Boolean GRNs 272

B.16 Correlation heat map between the various descriptors of local dynamics on ensembles generated using network structures from six published Boolean GRNs. 273

B.17 Scatter plot between the $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ on ensembles generated using network structures from four published Boolean GRNs 274

B.18 Scatter plot between the $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ on ensembles generated using network structures from six published Boolean GRNs 275

B.19 Scatter plot between the $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) on ensembles generated using network structures from four published Boolean GRNs . 276

B.20 Scatter plot between the $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) on ensembles generated using network structures from six published Boolean GRNs . . 277

B.21 Correlation heat map between descriptors of local dynamics and global measures of bushiness in ensembles generated using network structures from six published Boolean GRNs 278

B.22 Scatter plot between the $\overline{Z}_{max}$ and G-density on ensembles generated using network structures from four published Boolean GRNs 279

B.23 Scatter plot between the $\overline{Z}_{max}$ and G-density on ensembles generated using network structures from six published Boolean GRNs 280

B.24 Scatter plot between the $\overline{Z}_{min}$ and G-density on ensembles generated using network structures from four published Boolean GRNs 281

B.25 Scatter plot between the $\overline{Z}_{min}$ and G-density on ensembles generated using network structures from six published Boolean GRNs 282

B.26 Scatter plot between the $\overline{Z}_{ave}$ and G-density on ensembles generated using network structures from four published Boolean GRNs 283

B.27 Scatter plot between the $\overline{Z}_{ave}$ and G-density on ensembles generated using network structures from six published Boolean GRNs 284

B.28 Scatter plot between the $\overline{Z}_{mid}$ and G-density on ensembles generated using network structures from four published Boolean GRNs 285

B.29 Scatter plot between the $\overline{Z}_{mid}$ and G-density on ensembles generated using network structures from six published Boolean GRNs 286

B.30 Scatter plot between the network sensitivity ($\overline{s}$) and G-density on ensembles generated using network structures from four published Boolean GRNs 287

B.31 Scatter plot between the network sensitivity ($\overline{s}$) and G-density on ensembles generated using network structures from six published Boolean GRNs 288

B.32 Predictive power of $\overline{Z}_{mid}$ for G-density in ensembles generated using network structures from four published Boolean GRNs 289

B.33 Predictive power of $\overline{Z}_{mid}$ for G-density in ensembles generated using network structures from six published Boolean GRNs 290

B.34 Scatter plot between the $\overline{Z}_{max}$ and average convergence rate of trajectories originating at GoE states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 291

B.35 Scatter plot between the $\overline{Z}_{max}$ and average convergence rate of trajectories originating at GoE states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 292

B.36 Scatter plot between the $\overline{Z}_{min}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 293

B.37 Scatter plot between the $\overline{Z}_{min}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 294

B.38 Scatter plot between the $\overline{Z}_{ave}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 295

B.39 Scatter plot between the $\overline{Z}_{ave}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 296

B.40 Scatter plot between the $\overline{Z}_{mid}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 297

B.41 Scatter plot between the $\overline{Z}_{mid}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 298

B.42 Scatter plot between the network sensitivity ($\overline{s}$) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 299

B.43 Scatter plot between the network sensitivity ($\overline{s}$) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 300

B.44 Scatter plot between the average convergence rate of trajectories originating at *all* states (λ_{all}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 301

B.45 Scatter plot between the average convergence rate of trajectories originating at *all* states (λ_{all}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 302

B.46 Scatter plot between the average convergence rate of trajectories originating at *random* states (λ_{random}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs 303

B.47 Scatter plot between the average convergence rate of trajectories originating at *random* states (λ_{random}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs 304

C.1 TMRs generally exhibit substantially higher Boolean complexity than NCFs across different sign combinations of regulators 337

C.2 IMRs attain the highest average sensitivity, and for most sign combinations of regulators there exist NCFs with lower sensitivity than BMRs . . 338

C.3 BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the P-P network topology with $N = 16$ . . . 339

C.4 BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the R-R network topology with $N = 12$ . . . 340

C.5 BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the R-R network topology with $N = 16$ . . . 341

C.6 Modified TMR-based (Ising* and 01*) models generally fail to recover biological attractors 342

C.7 JS distance from gold standard reveals poor preservation of basin size distributions of biological attractors by modified TMR-based (Ising* and 01*) models 343

List of Tables

5.1 Ten published reconstructed Boolean GRN models 74

B.1 Boolean functions of the given RSCN-GRN Boolean model in BoolNet format 226

B.2 Boolean functions of the given EMT-GRN Boolean model in BoolNet format 227

B.3 Boolean functions of the given NCD-GRN Boolean model in BoolNet format 228

B.4 Boolean functions of the given EHD-GRN Boolean model in BoolNet format 229

B.5 Boolean functions of the given MD-GRN Boolean model in BoolNet format 230

B.6 Boolean functions of the given THC-GRN Boolean model in BoolNet format 231

B.7 Boolean functions of the given EGFR-GRN Boolean model in BoolNet format 232

B.8 Boolean functions of the given EMT-Senescence-GRN Boolean model in BoolNet format 232

B.9 Boolean functions of the given FOS-GRN Boolean model in BoolNet format 233

B.10 Boolean functions of the given GSD-GRN Boolean model in BoolNet format 234

B.11 Biological fixed points recovered by the RSCN-GRN Boolean model . . . 235

B.12 Biological fixed points recovered by the EMT-GRN Boolean model 235

B.13 Biological fixed points recovered by the NCD-GRN Boolean model . . . 235

B.14 Biological fixed points recovered by the EHD-GRN Boolean model . . . 236
B.15 Biological fixed points recovered by the MD-GRN Boolean model 236
B.16 Biological fixed points recovered by the THC-GRN Boolean model 236
B.17 Biological fixed points recovered by the EGFR-GRN Boolean model . . . 237
B.18 Biological fixed points recovered by the EMT-Senescence-GRN Boolean model 237
B.19 Biological fixed points recovered by the FOS-GRN Boolean model 237
B.20 Biological fixed points recovered by the GSD-GRN Boolean model 238
B.21 Type of the BF at each node of the given RSCN-GRN Boolean model . . 238
B.22 Type of the BF at each node of the given EMT-GRN Boolean model . . . 239
B.23 Type of the BF at each node of the given NCD-GRN Boolean model . . . 239
B.24 Type of the BF at each node of the given EHD-GRN Boolean model . . . 240
B.25 Type of the BF at each node of the given MD-GRN Boolean model 241
B.26 Type of the BF at each node of the given THC-GRN Boolean model . . . 242
B.27 Type of the BF at each node of the given EGFR-GRN Boolean model . . 243
B.28 Type of the BF at each node of the given EMT-Senescence-GRN Boolean model 243
B.29 Type of the BF at each node of the given FOS-GRN Boolean model . . . 244
B.30 Type of the BF at each node of the given GSD-GRN Boolean model . . . 245
B.31 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the RSCN-GRN model 246
B.32 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EMT-GRN model 247
B.33 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the NCD-GRN model 248
B.34 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EHD-GRN model 249

B.35 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the MD-GRN model 250
B.36 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the THC-GRN model 251
B.37 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EGFR-GRN model 252
B.38 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EMT-Senescence-GRN model 253
B.39 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the FOS-GRN model 254
B.40 Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the GSD-GRN model 255
B.41 Performance evaluation for using $1 - \overline{Z}_{mid}$ as a proxy for G-density, based on average AUROC and AUPRC scores 256

C.1 Numbers of various subtypes of the EThFs in the theoretical space 330
C.2 Numbers and proportions of various subtypes of the EThFs in the BBM dataset 331
C.3 Relative enrichment ratios (E_R) and statistical significance (p-values) for EThF subtypes in the BBM dataset 332
C.4 Relative enrichment ratios (E_R) and statistical significance (p-values) for the EThF subtypes ‘non-BMR NCF’, ‘NCF BMR’ and ‘non-NCF BMR’ in the BBM dataset 333
C.5 24 selected models from the BBM dataset 334
C.6 Summary of models following replacement of non-NCF BFs with NCFs . 335

Abstract

Gene regulatory networks (GRNs) orchestrate cellular behavior through complex combinatorial interactions among genes. Boolean models provide a tractable framework for studying GRN dynamics and have been widely used to analyze stability, multistability, and attractor-based representations of cellular phenotypes. While extensive research has established that biological GRNs are highly constrained in both network topology and regulatory logic, the specific effects of different classes of regulatory logic on global network dynamics have not been systematically characterized. The key objective of this thesis is to develop a systematic understanding of how different classes of Boolean functions (BFs) influence the dynamics of Boolean GRN models across varying network topologies and levels of abstraction, thereby informing principled choices of regulatory logic in Boolean modeling of biological systems. Using both random Boolean networks (RBNs) and reconstructed biological GRNs, we show that regulatory logic plays a central role in determining network stability and dynamical organization.

First, we analyze ensembles of RBNs to disentangle the effects of regulatory logic from those of network topology. By systematically varying network size, connectivity, and degree distribution, and by employing multiple notions of dynamical stability, we show that biologically meaningful classes of BFs consistently steer network dynamics toward ordered or near-critical regimes. In contrast, networks governed by random BFs exhibit strong sensitivity to topological parameters and frequently transition toward chaotic behavior with increasing connectivity or size. Across diverse stability measures, biologically meaningful BFs display reduced variability and enhanced robustness.

Second, we analyze the effect of BF classes on the stability of reconstructed Boolean GRNs using ensembles in which network structure and biological attractors are held fixed while regulatory logic is varied. Adopting a global state-space perspective, we systematically examine how different classes of BFs shape the structure of the state transition graph (STG). To characterize global dynamical organization, we adapt concepts originally developed in the study of cellular automata (CA) and introduce CA-inspired measures of

bushiness and convergence. We find that biologically meaningful BFs lead to faster convergence and more contractive STG structures than random logic.

Finally, we critically evaluate threshold majority rules (TMRs), a subclass of threshold functions frequently employed in GRN modeling due to their implementation simplicity. Through a comparative analysis with nested canalyzing functions (NCFs), we show that TMRs exhibit higher complexity and sensitivity, are underrepresented in empirical datasets of regulatory logic, and often fail to recover biological attractors and associated basin sizes. These results indicate that, despite their convenience, TMRs are poorly suited as a modeling choice for biological GRNs.

Altogether, this thesis establishes regulatory logic as a key determinant of Boolean GRN dynamics and provides systematic guidance for selecting biologically grounded BFs in the modeling of GRNs.

Chapter 1

Background and motivation

> *Deep in the chaotic regime, slight changes in structure almost always cause vast changes in behavior. Complex controllable behavior seems precluded.*
>
> – Stuart Kauffman

Cells within a multicellular organism share the same genome, yet they exhibit a remarkable diversity of phenotypes and functional behaviors. This diversity does not stem from differences in genetic content, but from finely tuned regulatory programs that govern when, where, and to what extent genes are expressed [1]. Gene regulatory networks (GRNs), comprising intricate webs of genes, transcription factors, and their regulatory interactions, lie at the heart of cellular decision-making processes that underlie development, differentiation, and adaptive responses to a changing environment [2]. Through nonlinear and combinatorial interactions, GRNs integrate multiple molecular signals and translate them into coherent cellular behaviors, while simultaneously ensuring robustness to intrinsic noise and environmental perturbations [1, 3, 4]. Over the past few decades, researchers have made significant strides in unraveling the structure and dynamics of GRNs, highlighting how biological regulation emerges from the coordinated action of many interacting components [1, 5–9]. Understanding how such networks collectively give rise to stable yet adaptable behavior remains a central challenge in systems and computational biology [3, 10–13].

To investigate how GRNs give rise to complex cellular behaviors, a variety of mathematical and computational modeling frameworks have been developed, each differing in their level of detail, underlying assumptions, and data requirements [14]. At one end of the spectrum, continuous models based on ordinary or partial differential equations describe gene expression dynamics using concentration levels and kinetic parameters, enabling detailed quantitative analyses when sufficient experimental data are available [1, 15]. Stochastic extensions of these models explicitly account for molecular noise and are particularly relevant when regulatory components are present in low copy numbers [16–18]. However, the applicability of such detailed approaches is often limited by parameter uncertainty and the rapid growth of model complexity with increasing network size. These challenges have motivated the development of more coarse-grained frameworks, including discrete and logical models, which emphasize regulatory interactions rather than precise kinetics [19–22]. In these approaches, genes are represented by a small number of activity states and regulatory effects are captured through rule-based formalisms [19–23]. As a result, such models offer scalability, interpretability, and robustness to incomplete data, making them particularly suitable for studying moderate- to large-scale GRNs. Consequently, discrete modeling frameworks have become indispensable tools for uncovering general principles governing GRN dynamics.

Building on these considerations, Boolean modeling has emerged as a particularly powerful and enduring framework for studying GRN dynamics. The origins of Boolean approaches to gene regulation can be traced back to early efforts to formalize regulatory logic in biological systems. Inspired by the operon model of Jacob and Monod [24], Sugita was among the first to employ logical circuits and Boolean formalisms to describe gene regulatory interactions [25, 26]. These foundational ideas were later developed and popularized by Stuart Kauffman, who introduced Boolean networks (BNs) as abstract models of genetic regulation and proposed that key features of cellular organization and differentiation could arise from simple binary rules acting on large networks [19, 27]. Around the same time, René Thomas developed a complementary framework that em-

phasized multistability, qualitative dynamical behavior, the role of feedback circuits in regulatory networks, and highlighted the importance of update schemes for capturing biologically realistic dynamics [28–30].

In its most common formulation, Boolean modeling represents each gene or regulatory component as a binary variable, active or inactive ("on" or "off"), and encodes regulatory influences through Boolean functions (BFs). Since its inception, the Boolean modeling framework has proven remarkably effective in capturing essential dynamical features of diverse biological processes. With the advent of high-throughput experimental techniques and the increasing availability of regulatory interaction data, this framework has become a standard tool for reconstructing and analyzing GRNs underlying processes such as differentiation, metabolism, apoptosis, proliferation, and cell-cycle control [7, 31–39]. Together, these developments have established Boolean modeling as a powerful framework that balances biological interpretability with mathematical and computational tractability, enabling the systematic analysis of large-scale regulatory networks that would otherwise remain challenging to address within more detailed quantitative frameworks.

In the early stages of systems-level thinking about gene regulation, long before the advent of high-throughput experimental techniques, detailed knowledge of regulatory interactions was largely unavailable. This scarcity of large-scale molecular data motivated the use of abstract and idealized modeling approaches aimed at uncovering general organizing principles of genetic regulation rather than reproducing specific biological systems [5, 28, 40]. Within this pre-genomic context, Stuart Kauffman introduced random Boolean networks (RBNs) as a conceptual model for gene regulation [5, 19, 27]. In RBNs, nodes (genes) are represented as binary variables whose regulatory inputs and Boolean update rules are assigned randomly. Despite their idealized nature, RBNs proved remarkably powerful in revealing fundamental features of collective gene dynamics. Systematic exploration of RBN behavior demonstrated that large networks governed by simple Boolean rules can display qualitatively distinct dynamical regimes, spanning ordered and chaotic

behavior, with a critical regime emerging at their boundary [5]. Importantly, the attractors of RBN dynamics were interpreted as stable gene expression patterns corresponding to cellular phenotypes, providing an early dynamical explanation for differentiation and robustness in biological systems. Subsequent studies on RBNs highlighted how global dynamical properties depend on the network topology and regulatory logic [41–43]. These insights motivated a shift away from purely random Boolean rules toward the systematic investigation of biologically meaningful classes of BFs, laying the conceptual foundation for later studies of reconstructed GRNs and for a central question addressed in this thesis. Although RBNs do not aim to faithfully represent real gene regulatory architectures, they established a foundational theoretical baseline that continues to inform our understanding of stability, robustness, and criticality in Boolean models of gene regulation.

The emergence of high-throughput experimental technologies in the post-genomic era fundamentally transformed the study of GRNs. These advancements enabled the systematic reconstruction of GRNs that were increasingly grounded in experimental evidence rather than theoretical abstraction. Extensive analyses of such reconstructed networks have consistently shown that their organization departs strongly from randomness, instead exhibiting structured topologies shaped by evolutionary and functional constraints [1, 8, 44, 45]. In parallel, the Boolean framework has seen a marked resurgence as a practical modeling paradigm for GRNs, owing to its ability to integrate heterogeneous data while remaining computationally tractable [7, 14, 31–33]. Over the past two decades, a growing number of Boolean dynamical models have been constructed to reproduce experimentally observed gene expression patterns and cellular phenotypes [46–48]. This shift has led to the establishment of publicly available repositories and platforms such as Cell Collective [49], GINSIM [50], and benchmark datasets including MCBF [51] and BBM [52], which collectively provide a diverse corpus of reconstructed Boolean GRN models. Together, these advances transformed Boolean modeling into a data-driven framework for system-specific GRN analysis and revealed that reconstructed biological networks deviate substantially from the random assumptions of early theoretical mod-

els, thereby motivating a more detailed examination of their structural organization and regulatory logic.

Detailed analyses of experimentally reconstructed GRNs have revealed that their structural organization is shaped by distinct organizational principles rather than random connectivity [1, 8]. Biological GRNs are typically organized into hierarchical and modular structures, often characterized by heterogeneous or heavy-tailed degree distributions, where a small number of regulatory hubs exert broad control over downstream targets, while localized subnetworks implement specific functional programs [1, 8]. Such architectures give rise to recurring patterns of interactions, including feed-forward loops, feedback motifs, and densely connected regulatory modules, which have been shown to perform distinct information processing and control functions in cellular systems [1, 44]. From a dynamical perspective, these non-random topological features contribute to the robustness of gene expression states by buffering against fluctuations and enabling compensatory pathways in response to perturbations [45]. Notably, dynamical analyses of Boolean models built on such reconstructed topologies indicate that structured architectures can bias networks toward regimes characterized by ordered or near-critical behavior, in contrast to the broad dynamical variability observed in RBN ensembles [53, 54]. Together, these findings establish that the topology of real GRNs is highly constrained and plays a central role in shaping their dynamical properties. However, while network structure provides an essential scaffold for regulatory dynamics, it alone does not determine system behavior, motivating a closer examination of how constraints on regulatory logic complement network topology in shaping the global dynamics of Boolean GRNs.

Alongside insights into network architecture, growing evidence indicates that the regulatory logic in biological GRNs is itself highly constrained and departs markedly from random expectations [7, 33, 34]. A central argument for the biological relevance of such constrained regulatory logic was articulated by Kauffman, who proposed that canalyzing functions (CFs), capture a form of 'chemical simplicity' in gene regulation, whereby specific regulatory inputs exert overriding control on gene expression outcomes [5]. Large

scale analyses of reconstructed Boolean GRNs have further demonstrated that the BFs employed in biological models are not uniformly distributed across the space of all possible BFs, but instead show strong enrichment for specific biologically meaningful classes, including unate functions (UFs) [55], canalyzing functions (CFs) [5], and nested canalyzing functions (NCFs) [7, 51, 56]. These BF classes encode regulatory logic that is constrained either through dominance and hierarchical control of the inputs, or through monotonic regulatory effects, reflecting fundamental asymmetries in the underlying molecular regulatory mechanisms. Importantly, some of the most highly enriched logic classes, such as NCFs and read-once functions (RoFs), have been shown to minimize well-established measures of Boolean complexity [57] and sensitivity [58], thereby linking regulatory logic to the robustness of network dynamics [51]. Taken together, these observations suggest that the choice of regulatory logic is a key determinant of collective network dynamics, motivating a systematic investigation of how different classes of BFs shape the behavior of Boolean GRNs.

Although extensive research has conclusively established that biological GRNs exhibit strong constraints at two levels, namely, network topology and regulatory logic, there was an unmet need for a systematic analysis of the specific role played by different classes of regulatory logic in shaping global network dynamics. Moreover, while meta-analysis based studies have revealed enrichment of biologically meaningful BFs in biological GRNs [51, 56], it remained underexplored whether and how these constraints translate into differences in global stability and state transition structure. Furthermore, the extent to which changes in network topology modulate dynamical behavior across different classes of regulatory logic had not been systematically characterized. This thesis addresses these gaps by explicitly disentangling the contribution of regulatory logic from that of network structure through a systematic analysis across two complementary settings: (a) abstract RBNs, wherein the impact of logic classes is examined across varying network architectures, and (b) reconstructed biological GRNs, wherein network structure and biological attractors are held fixed while regulatory logic is varied. Together, these

analyses provide a unified framework for understanding how choices of BF classes shape the dynamical behavior and stability of Boolean GRN models.

Assessing how regulatory logic shapes network behavior requires comparing the resulting dynamics through appropriate notions of stability. Within the Boolean network (BN) framework, stability has traditionally been examined using perturbation-based approaches, such as damage spreading measures, which probe whether small differences between network states amplify or decay over time. In particular, such measures have been widely used to characterize dynamical regimes in abstract settings such as RBNs [59, 60]. In BN framework, stability can also be characterized using properties of attractors and their basins of attraction, including their number, average length, and the fraction of fixed point attractors, since attractors capture the asymptotic behavior of the system and are often interpreted as stable cellular phenotypes [5]. However, attractor properties alone provide an incomplete picture of BN behavior, as they do not capture the transient processes through which network states evolve. A more global perspective on BN dynamics is obtained by considering the organization of trajectories in state space, which is naturally captured by the structure of the state transition graph (STG). The STG encodes all possible state transitions and thereby integrates both transient and asymptotic behavior. While perturbation-based analyses have provided important insights into the role of regulatory logic in shaping ordered and chaotic regimes, global STG-based characterizations remain largely unexplored in the context of reconstructed biological Boolean GRN models, despite their prominence in the study of cellular automata (CA) [61–63]. Accordingly, this thesis employs complementary dynamical perspectives to examine the consequences of the choice of regulatory logic on BN dynamics.

Despite the success of Boolean modeling in capturing key aspects of GRN dynamics, a critical modeling choice often receives less attention than it warrants: the choice of BFs used to define the regulatory logic at individual nodes. Given that different BFs encode qualitatively distinct regulatory mechanisms, ranging from additive and monotonic regulation to strongly hierarchical control, it is essential to systematically assess how such

choices shape the dynamical behavior and stability of BN models. The key objective of this thesis is to develop a systematic understanding of the influence of different BF classes on the dynamics of Boolean GRNs across varying network topologies and levels of abstraction, and thereby inform suitable choices of BFs when modeling biological regulatory networks within the Boolean framework. To achieve this, the thesis undertakes a multi-pronged investigation spanning both abstract network ensembles and reconstructed biological GRNs.

The **first objective** is to systematically investigate how the choice of regulatory logic class modulates the dynamical regime of BNs in abstract settings. Using RBNs as theoretical testbeds, we systematically examine how different classes of BFs influence network stability and order-chaos behavior. By analyzing ensembles of RBNs across a range of network sizes, connectivities, and degree distributions, this objective aims to disentangle the contribution of regulatory logic from that of network topology. In particular, we assess whether biologically meaningful classes of BFs consistently promote more stable dynamics than random BFs across multiple notions of stability, and whether changes in network size or connectivity affect different logic classes in distinct ways.

The **second objective** shifts the focus from abstract RBN ensembles to reconstructed Boolean GRNs. In the ensembles generated for our analyses, network structure and biological attractors are kept fixed while only the regulatory logic is varied. While RBNs provide valuable conceptual insights, they remain stylized and divorced from the architecture of real biological networks. Accordingly, this objective adopts a global, state space perspective to characterize the dynamics of reconstructed GRNs and examines how different classes of BFs shape the structure of the STG. In particular, we analyze global dynamical properties such as bushiness and convergence, which reflect the organization of transient trajectories. Since such STG-based analyses have not previously been carried out in the context of reconstructed Boolean GRNs, we adapt concepts and measures originally developed in the study of cellular automata (CA) [61–63], introducing CA-inspired descriptors tailored to BN models. These analyses allow us to systematically compare

how biologically meaningful and random BFs differ in their impact on the global organization of network dynamics.

While the analyses presented in the first two objectives establish that regulatory logic plays a decisive role in shaping dynamical behavior of Boolean GRN models, a broader practical challenge remains inherent to Boolean modeling. In realistic biological settings, the reconstruction of Boolean GRNs is fundamentally underdetermined due to limited quantitative data on regulatory mechanisms. Even when network topology and regulatory signs are known, multiple candidate BFs may be consistent with the available information, making the selection of update rules a nontrivial modeling decision [64]. As a result, practical considerations such as simplicity, uniformity, and computational convenience often influence the choice of regulatory logic, sometimes in the absence of strong biological justification.

Within this context, some threshold-based Boolean rules have gained popularity as a pragmatic modeling strategy [34, 65–67]. In particular, threshold majority rules (TMRs), which update a node's state based on the signed sum of its regulators and its own activity, offer an elegant simplification that circumvents the need to estimate interaction strengths or activation thresholds [34, 65–67]. However, a systematic evaluation of whether such simplicity comes at an acceptable expense of biological realism, as well as a meaningful comparative analysis with empirically enriched logic classes, was lacking prior to our work. This gap is further compounded by the absence of a straightforward mapping between TMRs and established classes of biologically meaningful BFs, which had hindered their placement within existing theoretical frameworks of BF classes.

The **third objective** of this thesis is therefore to critically assess the suitability of TMRs as a modeling choice for Boolean GRNs. This objective examines TMRs at both the function and network levels by comparing them against NCFs [7, 68, 69], an empirically enriched subclass of threshold functions [51, 56, 70–75], across structural, empirical, and dynamical criteria. Specifically, we evaluate whether TMRs capture key features observed in empirical datasets of regulatory logic, including canalyzation, bias patterns,

intrinsic complexity, robustness to perturbations, and the ability to recover known biological attractors and their associated basin sizes. Through this analysis, the thesis aims to determine whether the convenience of TMRs justifies their continued use in Boolean GRN modeling, or whether alternative, biologically grounded classes of regulatory logic such as NCFs provide a more principled foundation.

The objectives outlined above are addressed through a sequence of chapters that progressively develop the theoretical background, methodological framework, and systematic analyses described below.

1.1 Thesis organization

The remaining six chapters of the thesis progressively develop a framework for understanding how regulatory logic shapes the dynamics and stability of Boolean GRNs, from foundational concepts to systematic analyses of random and reconstructed models.

Chapter 2 introduces the theoretical and methodological foundations of Boolean modeling of GRNs and their associated logical update rules. We begin by presenting the basic elements of Boolean modeling, including network states, state transition graphs (STGs), attractors, and update schemes such as synchronous [5] and asynchronous [29] dynamics. The classical framework of random Boolean networks (RBNs) is then reviewed. Next, we discuss reconstructed Boolean models of biological systems that emerged in the post-genomic era, highlighting data-driven approaches and community-curated resources such as Cell Collective [49], GINSIM [50], MCBF [51] and BBM dataset [52]. The chapter further provides a detailed background on Boolean functions (BFs), including their representations, permutations, and negations, and formally defines several classes of biologically meaningful BFs, namely effective functions (EFs), effective and unate functions (EUFs), read-once functions (RoFs), nested canalyzing functions (NCFs), and link operator functions (LOFs). We then dicuss two complexity measures of BFs, Boolean complexity and average sensitivity. Finally, the chapter presents strategies employed for sampling Boolean models for a given GRN under different constraints, including attrac-

tors, sign-conformity and the type of regulatory logic, and introduces a hypercube-based iterative sampling method that enables scalable sampling of EUFs for nodes with higher input degrees.

Chapter 3 provides a systematic overview of stability measures for Boolean network (BN) dynamics. Stability is a central yet multifaceted concept in the analysis of Boolean models, and this chapter organizes the diverse notions of stability into a coherent framework. We first review perturbation-based measures derived from damage spreading arguments, including the Derrida coefficient (δ), the final Hamming distance (h^{∞}), and fragility (ϕ) [60]. We then describe stability measures based on properties of attractors and their basins of attraction, such as the number and length of attractors and the fraction of states converging to fixed points. Next, we discuss measures rooted in polynomial approximations of regulatory rules and the resulting notion of 'approximability' of network dynamics [76]. Finally, we introduce stability measures inspired by the theory of cellular automata (CA) [61–63], which characterize global properties of the STG, such as bushiness and convergence. This chapter establishes the conceptual and methodological groundwork for the analyses developed in the subsequent chapters.

Chapter 4 investigates how choices of BF class modulate the dynamical regime in RBNs. RBNs have been widely explored as models of complex systems and as simplified representations of GRNs. As in many areas of dynamical systems research, the interplay between stability and chaos has been a central theme in RBN research. Although network topology is known to influence network dynamics, the BFs assigned to each node play an equally crucial role. In this chapter, we systematically examine the influence of five different classes of BFs on RBN dynamics using a range of dynamical stability measures, including damage spreading and attractor-based properties. We show that, compared to random BFs, biologically meaningful BFs consistently drive the dynamics toward the ordered regime, regardless of network size, connectivity, or degree distribution. Moreover, as network connectivity or size increases, random BFs typically push the dynamics toward a more chaotic regime, whereas biologically meaningful functions exhibit minimal fluc-

tuations in most stability measures. These results highlight the importance of regulatory logic choices in shaping network dynamics and emphasize the advantage of restricting to biologically meaningful classes in the reconstruction and modeling of biological systems within the Boolean framework. The work reported in this chapter is based on the published manuscript [77].

Chapter 5 shifts the focus from RBNs to reconstructed Boolean GRNs, where we investigate how different classes of BFs influence the structural features of the STG. Although networks with the same topology and biological attractors can exhibit drastically different STGs depending on the type of BFs employed, such effects have not been systematically characterized. Using ten reconstructed Boolean GRNs, we generate ensembles that differ only in the types of BFs, while keeping the network structure and biological attractors fixed. We compute the rate of contraction (bushiness) and the rate of convergence from their STGs, quantified using measures inspired by CA that are based on garden-of-Eden (*GoE*) states. We find that biologically meaningful BFs lead to higher bushiness and faster convergence than random EFs. Since computing these global measures becomes computationally expensive for larger networks, we adapt Wuensche's Z-parameter [78] from CA to BFs in BNs and introduce four natural variants. Together with network sensitivity [53, 58], these four Z-parameter variants constitute the descriptors of local dynamics, and some of them are found to be useful proxies for bushiness. Finally, we provide an excellent proxy for convergence based on computing transient lengths originating from randomly sampled states rather than *GoE* states. The work reported in this chapter is based on the submitted manuscript [79].

Chapter 6 turns attention to a critical evaluation of the suitability of two simple threshold-based Boolean rules for modeling GRN dynamics. These two subtypes of threshold functions, which we collectively term threshold majority rules (TMRs), update a gene's state based on the signed sum of its regulators and its own activity, offering an elegant mathematical simplification. Here, we assess whether this simplicity of TMRs compromises biological realism. We rigorously evaluate the two standard TMR variants

with respect to their suitability for GRN modeling. First, we show that they exhibit limited canalyzation, discordant bias patterns, and often eliminate self-inhibitory interactions present in the GRN. Furthermore, TMRs exhibit elevated complexity and heightened sensitivity to perturbations. Moreover, these rules are underrepresented in empirical datasets of regulatory logic rules, contrary to what would be expected for biologically relevant functions. At the network level, we show that TMRs frequently fail to recover biological attractors and their associated basin size distributions. In contrast, models incorporating suitable NCFs consistently exhibit superior performance across these criteria. These findings invite a thoughtful reevaluation of TMRs as an appropriate logic for modeling GRNs despite their theoretical appeal. The work reported in this chapter is based on the published manuscript [80].

Chapter 7 concludes the thesis with a summary of the main findings and contributions presented across the different chapters and discusses potential avenues for future work.

Chapter 2

Boolean modeling of gene regulatory networks and associated logical update rules

Chapter overview. This chapter provides the theoretical and methodological background required for the work presented in this thesis. It begins with an introduction to the Boolean modeling framework for gene regulatory networks (GRNs), followed by a review of the classical random Boolean network (RBN) paradigm. It then discusses reconstructed Boolean models of biological systems, highlighting how experimental data are integrated to construct system-specific Boolean models. The chapter then introduces Boolean functions (BFs), including their representations and basic operations such as permutations and negations. It subsequently presents several biologically meaningful classes of BFs that frequently occur in reconstructed Boolean models of regulatory networks. It then introduces the measures used throughout this thesis to characterize the complexity of BFs. Finally, it describes the procedures used to sample different classes of BFs under biological and topological constraints, including a sampling algorithm for effective and unate functions (EUFs) developed as part of this thesis.

GRNs play a crucial role in orchestrating cellular decision-making processes [2]. Over the

past few decades, researchers have made significant strides in unraveling the complexity of these biological networks, shedding light on both their structural organization and dynamical behavior [1,5–9]. Among the various approaches developed to model GRNs, the *logical modeling* framework has emerged as a particularly powerful tool. Sugita [25,26], inspired by Jacob and Monod [24], was among the earliest to formalize logical circuits and utilize the Boolean formalism to model GRNs. Soon thereafter, Stuart Kauffman [19,27] and René Thomas [28,29] solidified this framework and brought significant attention towards this approach to study the relationships between structure and dynamics of GRNs. In its most common form, logical modeling utilizes Boolean constraints, where genes or their products are considered to be in one of two states: active (*on*) or inactive (*off*). Since its inception, this framework has been widely successful in capturing diverse biological processes. With advancements in sequencing technologies and the availability of large-scale data, this framework has become instrumental in modeling key cellular processes including differentiation, metabolism, apoptosis, proliferation and cell-cycle regulation [7,31–39].

2.1 Boolean modeling of gene regulatory networks

A Boolean model of a GRN consists of a set of nodes and directed edges that correspond to genes and the interactions between them respectively [5, 19, 27]. We denote the state of the network with N nodes at time t by the vector $\mathbf{X}(t)$ whose i^{th} entry is given by $x_i(t)$, where $i \in \{1, 2, \ldots, N\}$ and $x_i \in \{0, 1\}$. Each node of the network is assigned a BF (or *logical update rule* or *regulatory logic*) that governs the local dynamics. That is, $x_i(t+1) = f_i(x_i^1(t), x_i^2(t), \ldots, x_i^k(t))$, where f_i is the BF that acts on the k inputs to node x_i, $x_i^m(t)$ ($m \in \{1, 2, \ldots, k\}$), to return the value $x_i(t+1)$.

The number of input combinations $(x_i^1(t), x_i^2(t), \ldots, x_i^k(t))$ for which $x_i(t+1) = 1$ is the bias P of the BF f_i. Of all possible BFs, some classes are *biologically meaningful* in that they satisfy properties that 'real' regulatory logic are expected to possess such as effectiveness, unateness and canalyzation [5,55,81]. In Section 2.5, we provide a detailed

description of different classes of BFs used in the studies presented in this thesis.

If all BFs $\mathbf{F} = (f_1, f_2, \ldots, f_N)$ act on all nodes simultaneously, the update is said to be *synchronous* [5], otherwise it is *asynchronous* [29]. When the system is updated using synchronous scheme, then it can be expressed with the following equation:

$$\mathbf{X}(t+1) = \mathbf{F}\left[\mathbf{X}(t)\right] \tag{2.1}$$

Trajectories in the state space can be traced out by the recursive application of Eq. (2.1) to any given initial condition $\mathbf{X}(0)$.

The state transition graph (STG) of the Boolean network (BN) is the directed graph in state space where each edge connects a state $\mathbf{X}$ to its image under Eq. (2.1). Under synchronous dynamics, a trajectory can meet with two possible fates: (a) it reaches a *fixed point attractor*, *i.e.*, a state which on further update remains unaltered ($\mathbf{X}(t) = \mathbf{X}(t+1)$), or (b) it ends up in a *cyclic attractor* (or *cycle*), in which case the network keeps cycling through a fixed set of states indefinitely on further updates. Furthermore, the set of states in the STG that cannot be obtained by the recursive application of Eq. (2.1) to any state of the network constitutes its *garden-of-Eden* (*GoE*) states - these states lack predecessors (or pre-images). In addition, all states that are not members of the attractor are called 'transient' states.

Asynchronous updates lead to the same set of fixed points and may be considered more realistic but they are much more difficult to analyze [82]. In this thesis, we have used synchronous dynamics for all our investigations.

2.2 Random Boolean networks

For several decades, tools from the study of complex systems have played a central role in the quantitative analysis of biological systems, particularly in understanding principles of self-organization [1, 5, 8, 14, 44, 83–85]. Some of the conceptual foundations for such approaches were developed earlier in the field of cellular automata (CA), where automata

and computational theories were applied to the study of living systems [86–89]. Over time, a wide range of automata models have been investigated, including two-dimensional cellular automata [86], one-dimensional elementary cellular automata [87, 90, 91], non-uniform cellular automata [92], and sequential dynamical systems [93]. These studies provided early evidence that complex global behavior can emerge from simple local interaction rules.

In this context, Kauffman proposed that GRNs could be modeled as generalizations of automata in the form of BNs [5, 19, 27]. In these models, genes assume one of two possible states, *on* or *off*, representing whether a gene is expressed (see Section 2.1). Given the limited knowledge of regulatory interactions available at the time, Kauffman introduced RBNs, in which both the regulatory topology and the Boolean update rules are assigned randomly.

In an RBN consisting of N nodes, each node receives inputs from k other nodes selected randomly from the network and is assigned a BF that determines its state from the states of its regulators. Such networks are not intended to represent a specific biological system; rather, they form ensembles of networks whose statistical dynamical properties can be studied. RBNs were originally introduced as simple models for exploring generic dynamical properties of gene regulatory systems [19, 27]. A central hypothesis arising from this work was that attractors of the network dynamics correspond to stable gene expression patterns associated with cellular phenotypes, implying that distinct attractors may represent different cell types [19, 27]. In this interpretation, cellular differentiation can be viewed as transitions between attractors triggered by perturbations or stochastic fluctuations [19]. Kauffman also suggested that the number of attractors in an RBN scales approximately as $\sqrt{N}$, which was argued to be broadly consistent with the limited number of cell types observed in biological organisms [19]. It was further conjectured that the typical attractor lengths in RBNs increase asymptotically as a power of N or slower. Although later studies refined these scaling estimates [94–96], these early conjectures stimulated extensive research into the attractor structure and dynamics of BNs.

Another key insight obtained from studies of RBN dynamics is the existence of distinct dynamical regimes depending on network connectivity. Early studies using this framework suggested that networks with small connectivity ($k \approx 2$ or 3) tend to operate in an ordered regime where perturbations decay, whereas highly connected networks display chaotic behavior in which perturbations spread through the system. Between these regimes lies a critical regime separating order from chaos. Kauffman hypothesized that biological regulatory systems may operate near this boundary, often referred to as the *edge of chaos*, where both robustness and adaptability can be maintained [5].

Further studies on RBNs emphasized that overall dynamical behavior emerges from the interplay between network topology and regulatory logic [41–43]. For example, Kauffman *et al.* [41] showed that networks composed of canalyzing BFs exhibit stable dynamics across a range of network architectures. Several measures developed in the study of RBN dynamics have proven useful for quantifying BN behavior. For instance, the notions of *activity* and *average sensitivity* [58] establish a connection between the structure of BFs and the dynamical regime of BNs.

Together, these studies established RBNs as a suitable theoretical framework for investigating the dynamics of GRNs. They highlight that, in addition to network topology, the choice of BFs governing regulatory interactions can strongly influence network behavior. This observation motivates several of the investigations carried out in this thesis, where we systematically examine how different classes of BFs affect the dynamical behavior and stability of BN models.

2.3 Reconstructed Boolean models of biological systems

Over the past two decades, the development of experimentally grounded Boolean models has shifted the field from the theoretical abstractions of early RBN studies toward system-specific, data-constrained representations of gene regulatory dynamics. In these models, nodes represent biological entities such as genes, proteins, or signaling molecules, while logical update rules encode regulatory interactions derived from experimental observa-

tions. Unlike RBNs, where both topology and regulatory functions are assigned randomly, reconstructed Boolean models are designed to capture known molecular interactions and reproduce experimentally observed cellular behaviors.

Constructing a Boolean model of a real GRN involves data integration and biological abstraction [32, 97]. The process begins by defining the system boundary, i.e., selecting relevant genes, proteins, or signaling molecules as nodes based on prior knowledge and the biological process of interest. Regulatory interactions are then curated from diverse sources including transcriptomic datasets, chromatin immunoprecipitation (ChIP) assays, and reporter gene studies as well as curated databases such as TRRUST [98], JASPAR [99], and TRANSFAC [100].

Once the regulatory graph is assembled, a BF f_i must be assigned for each node x_i with in-degree k_i. Several strategies are employed in practice in this context. The most common is expert curation, where logical rules are constructed from experimental evidence by identifying known activators and repressors and combining them using logical operators such as AND, OR, and NOT [23, 31, 32]. Alternatively, data-driven inference methods attempt to reconstruct network topology and update functions from binarized gene expression time-series. Early approaches such as REVEAL [101] and Best-Fit Extension [102], as well as more recent tools such as GABNI [103], infer Boolean rules consistent with observed dynamics. Another class of approaches formulates rule assignment as an optimization or constraint satisfaction problem, where a prior knowledge network is fitted to experimental perturbation data to select Boolean rules that best reproduce observed signaling responses [104, 105]. In practice, reconstructed models are often hybrid constructions. The regulatory graph is typically curated from literature, while update rules are assigned through expert reasoning guided by parsimony, favoring simple logical structures unless stronger mechanistic evidence exists. For nodes with multiple regulators, knowledge of activatory and inhibitory edges alone does not uniquely determine the logical dependencies among regulators [106]. For example, if a gene is regulated by two activators, it may require both regulators for activation (AND logic) or only one of

them (OR logic). Experimental perturbation data can sometimes resolve this ambiguity: if knockout of either regulator abolishes expression, an AND rule is appropriate, whereas if only the simultaneous knockout of both regulators suppresses expression, an OR rule is consistent with the observations. When such mechanistic information is unavailable, several candidate logical rules may be constructed or a default rule can be chosen [106]. In many Boolean modeling studies, the lack of detailed regulatory information motivates the use of simple threshold-based rules due to their conceptual simplicity and ease of implementation. The suitability of such rules for modeling GRNs is discussed in Chapter 6.

Boolean models have been successfully applied to a wide range of biological systems. Early examples include the Boolean model of *A. thaliana* floral development [31] and the *D. melanogaster* segment polarity network [32]. Since then, this framework has been used to study diverse regulatory systems including models of the mammalian cell cycle [23], apoptosis signaling [107], the regulatory network of *E. coli* metabolism [33], cancer signaling [35], and transcriptional regulation in hematopoietic differentiation [37].

The expansion of reconstructed models has also led to the development of curated model repositories and benchmark datasets. Repositories and platforms such as Cell Collective [49] and GINsim [50] provide collections of published models that enable systematic investigation of the regulatory logic employed in Boolean models of living systems. For comparative studies, curated datasets have become particularly valuable. The MCBF dataset [51] compiles BFs extracted from published models, categorizes them into different BF classes, and annotates them with properties such as bias and canalization depth, while the BBM dataset [52] aggregates published Boolean GRN models from multiple sources. Together, these resources have made it possible to move from isolated case studies toward population-level, statistically grounded analyses of the structural and functional organization of Boolean GRN models.

Prior meta-analyses of reconstructed BNs have revealed that the logical update rules differ markedly from the random BFs typically assumed in classical RBN studies [51,56].

In particular, BFs in reconstructed networks are strongly biased and exhibit high levels of canalyzation, with specific subclasses significantly enriched relative to random expectations [51]. Understanding the dynamical consequences of this enrichment constitutes a central theme of this thesis.

2.4 Background on Boolean functions

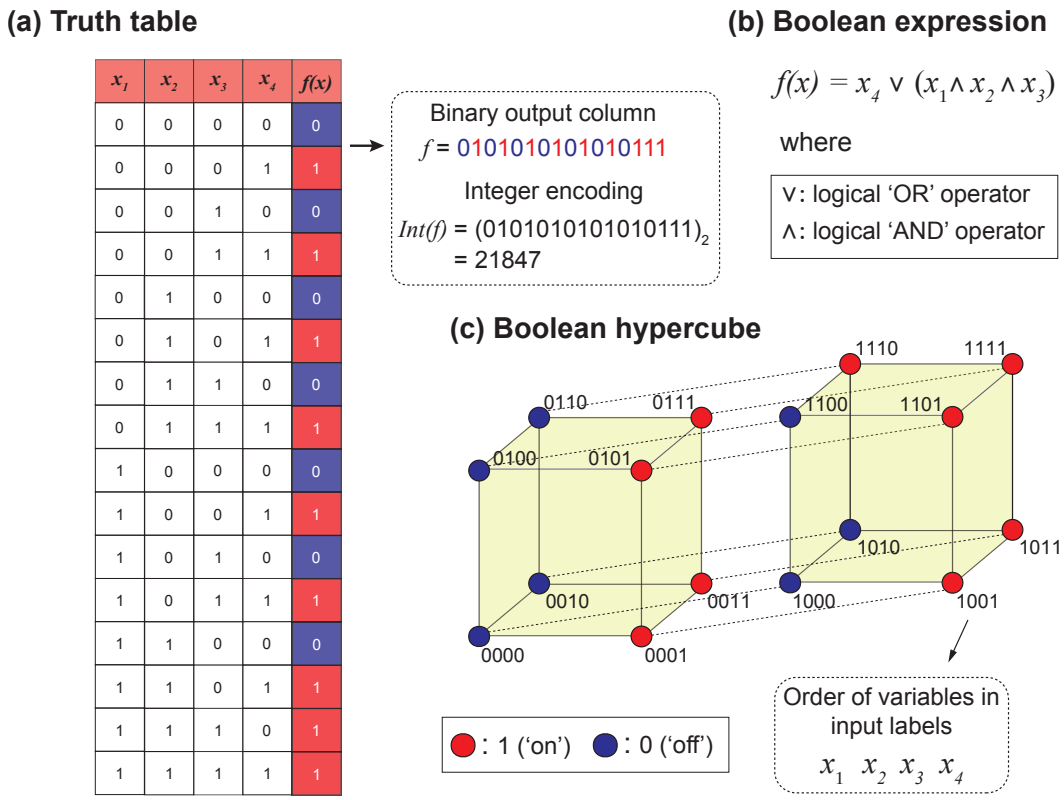
Mathematically, a BF can be represented as a mapping:

$$f : \{0,1\}^k \to \{0,1\},$$

where $\{0,1\}$ denotes the Boolean domain and k is the number of inputs of the function f. Every k-input BF can be represented in several equivalent ways, including a truth table, a Boolean expression, or a graphical representation on a Boolean hypercube. Figure 2.1 illustrates these different representations using an example with four input variables.

A BF can be specified by a truth table, which lists all 2^k possible combinations of the input variables together with the corresponding output value of the function. In this thesis, we adopt a fixed convention for ordering the rows and columns of the truth table. If $x_1, x_2, \ldots, x_{k-1}$, and x_k denote the input to the BF f, then the same order (from left to right) determines the variables appearing in the successive columns of the truth table (see Figure 2.1(a)). The rows enumerate all possible input configurations in lexicographic order, starting from the configuration where all inputs are 0 and ending with the configuration where all inputs are 1.

For convenience, we also use an integer encoding for the BF inferred from the truth table. This encoding is obtained by reading the output column of the truth table as a binary string ordered from top to bottom. The leftmost (most significant) bit corresponds to the output of the first row of the truth table (all input bits equal to 0), while the rightmost (least significant) bit corresponds to the output of the last row (all input bits equal to 1). The resulting binary string uniquely determines the BF and can be converted to its decimal



x_1	x_2	x_3	x_4	$f(x)$
0	0	0	0	0
0	0	0	1	1
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	1
0	1	1	0	0
0	1	1	1	1
1	0	0	0	0
1	0	0	1	1
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

Figure 2.1: Different representations of Boolean functions.

representation, which we refer to as the integer encoding of the BF (Figure 2.1(a)). Note that this encoding is unique for a fixed k; otherwise, the same integer may correspond to different BFs with different numbers of inputs.

A BF may also be written as a Boolean expression involving its input variables and logical operators such as OR ($\vee$) and AND ($\wedge$). An example of such a representation is shown in Figure 2.1(b), where the same function represented in the truth table is expressed as a propositional formula of the input variables.

Another useful representation is obtained by considering the BF as a Boolean hypercube. In this representation, each vertex corresponds to one possible input configuration, and edges connect vertices whose input vectors differ in exactly one variable. The value of the BF assigns a state to each vertex, which can be interpreted as "on" (1) or "off" (0). Figure 2.1(c) illustrates this representation for the example function, where red vertices correspond to output value 1 and blue vertices correspond to output value 0.

2.4.1 Negation of Boolean functions

An input x_i of a BF f is referred to as negated if the roles of 0 and 1 for that variable are exchanged. This operation generally changes the BF. It is not difficult to see that the table of this new BF can be obtained by swapping the values of the initial table's *output* column in all pairs of rows that differ only by the value of x_i. Equivalently, on the hypercube, this operation corresponds to swapping the hyperplanes $x_i = 0$ with $x_i = 1$ where i is the index of the input to be negated. For example, consider a k-input BF f where the order of the input variables in the associated truth table is $x_1, ..., x_i, ..., x_k$ and say

$$f(x_1, ..., x_i, ..., x_k) = T_1 \text{ and } f(x_1, ..., \overline{x_i}, ..., x_k) = T_2$$

where $T_1, T_2 \in \{0, 1\}$.

Note that for two different sets of values for the inputs $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{k-1}, x_k$, the two corresponding output values for $f(x_1, ..., x_i, ..., x_k) = T_1$ and $f(x_1, ..., \overline{x_i}, ..., x_k) = T_2$ may be different. Then, the negation of the i^{th} input leads to the new BF f':

$$f'(x_1, ..., \overline{x_i}, ..., x_k) = T_1 \text{ and } f'(x_1, ..., x_i, ..., x_k) = T_2.$$

As an example, consider $f = x_1 \wedge (\overline{x}_2 \vee x_3)$. Negation of the variable x_1 leads to the BF $f = \overline{x}_1 \wedge (\overline{x}_2 \vee x_3)$. Note that for any BF, there are 2^k negations possible, two per input, if we allow each input to be negated or not.

Using both permutations and negations, there are a total of $k!\ 2^k$ possible combinations of permutations and negations for a k-input BF. We can represent these permutations with negations as a matrix with rows corresponding to permutations and columns to nega-

tions. For illustration, consider the example provided above where $f_1 = x_1 \wedge (\overline{x}_2 \vee x_3)$:

$$\mathbf{M} = \begin{array}{c} \begin{array}{cccccc} \texttt{+ + +} & \texttt{+ + -} & & \texttt{...} & \texttt{- - +} & \texttt{- - -} \end{array} \\ \begin{bmatrix} x_1 \wedge (x_2 \vee x_3) & x_1 \wedge (x_2 \vee \overline{x}_3) & \dots & \overline{x}_1 \wedge (\overline{x}_2 \vee x_3) & \overline{x}_1 \wedge (\overline{x}_2 \vee \overline{x}_3) \\ x_1 \wedge (x_3 \vee x_2) & x_1 \wedge (x_3 \vee \overline{x}_2) & \dots & \overline{x}_1 \wedge (\overline{x}_3 \vee x_2) & \overline{x}_1 \wedge (\overline{x}_3 \vee \overline{x}_2) \\ x_2 \wedge (x_1 \vee x_3) & x_2 \wedge (x_1 \vee \overline{x}_3) & \dots & \overline{x}_2 \wedge (\overline{x}_1 \vee x_3) & \overline{x}_2 \wedge (\overline{x}_1 \vee \overline{x}_3) \\ x_2 \wedge (x_3 \vee x_1) & x_2 \wedge (x_3 \vee \overline{x}_1) & \dots & \overline{x}_2 \wedge (\overline{x}_3 \vee x_1) & \overline{x}_2 \wedge (\overline{x}_3 \vee \overline{x}_1) \\ x_3 \wedge (x_1 \vee x_2) & x_3 \wedge (x_1 \vee \overline{x}_2) & \dots & \overline{x}_3 \wedge (\overline{x}_1 \vee x_2) & \overline{x}_3 \wedge (\overline{x}_1 \vee \overline{x}_2) \\ x_3 \wedge (x_2 \vee x_1) & x_3 \wedge (x_2 \vee \overline{x}_1) & \dots & \overline{x}_3 \wedge (\overline{x}_2 \vee x_1) & \overline{x}_3 \wedge (\overline{x}_2 \vee \overline{x}_1) \end{bmatrix} \end{array}$$

One may encounter interesting properties when considering the permutations and negations of BFs within specific classes. For example, if one starts with a nested canalyzing function (NCF; see next Section 2.5 for the definition of this class of BFs), these operations generate BFs that are themselves NCFs, and in fact NCFs with the same bias (see Section 2.1 for the definition of bias of a BF) as the original NCF.

2.4.2 Permutations of Boolean functions

Given a k-input BF, one can apply a permutation of the k input variables in its propositional formula to produce a modified BF. There are $k!$ such permutations. Consider for instance the 3-input BF $f = x_1 \wedge (\overline{x}_2 \vee x_3)$. The $3! = 6$ permutations on its inputs correspond to (x_1, x_2, x_3), (x_1, x_3, x_2), (x_2, x_1, x_3), (x_2, x_3, x_1), (x_3, x_1, x_2) and (x_3, x_2, x_1). The associated permutations on the BFs are $f_1 = x_1 \wedge (\overline{x}_2 \vee x_3)$, $f_2 = x_1 \wedge (\overline{x}_3 \vee x_2)$, $f_3 = x_2 \wedge (\overline{x}_1 \vee x_3)$, $f_4 = x_2 \wedge (\overline{x}_3 \vee x_1)$, $f_5 = x_3 \wedge (\overline{x}_1 \vee x_2)$ and $f_6 = x_3 \wedge (\overline{x}_2 \vee x_1)$. Note that in this particular example, the six permutations lead to six distinct BFs, hereafter referred to as *non-equivalent* BFs. But that is not always the case. As an example, consider the permutations of the BF $f = x_1 \wedge (x_2 \vee x_3)$: $f_1 = x_1 \wedge (x_2 \vee x_3)$, $f_2 = x_1 \wedge (x_3 \vee x_2)$, $f_3 = x_2 \wedge (x_1 \vee x_3)$, $f_4 = x_2 \wedge (x_3 \vee x_1)$, $f_5 = x_3 \wedge (x_1 \vee x_2)$ and $f_6 = x_3 \wedge (x_2 \vee x_1)$. Here, there are only three non-equivalent permuted BFs because $f_1 = f_2$, $f_3 = f_4$ and $f_5 = f_6$.

The operation of permuting variables in the Boolean expression of a BF is equivalent to permuting the corresponding columns of the truth table and then reordering the rows (which include the output of the BF) in lexicographical order to bring the table into its canonical form. In the case

of a 3-input BF, one is to apply a permutation $\pi \in S_3$ (the symmetric group on three elements) and that will generate different orderings of the variables. The identity element of S_3 (absence of a permutation), e, does not change the order of the inputs so it corresponds to the ordering (x_1, x_2, x_3). The transposition (1 2) interchanges the 1st and 2nd variables keeping the 3rd variable fixed and generates the ordering: (x_2, x_1, x_3). The transposition (1 3) interchanges the 1st and 3rd variables keeping the 2nd variable fixed and generates the ordering: (x_3, x_2, x_1). The transposition (2 3) interchanges the 2nd and 3rd variables keeping the 1st variable fixed and generates the ordering: (x_1, x_3, x_2). The three-cycle (1 2 3) replaces the 1st variable by the 2nd, the 2nd by the 3rd and the 3rd by the 1st and generates the ordering: (x_2, x_3, x_1). The three-cycle (1 3 2) replaces the 1st variable by the 3rd, the 2nd by the 1st and the 3rd by the 2nd and generates the ordering: (x_3, x_1, x_2).

2.5 Biologically meaningful classes of Boolean functions

Throughout this thesis, we analyze the effects of different classes of BFs on BN dynamics and related properties. In particular, we focus on BFs that are known to be biologically meaningful [51]. Not surprisingly, there are a number of properties that random BFs do not often possess but are typical of biologically meaningful BFs (as arising in reconstructed BNs of biological systems). This section describes several classes of BFs that are considered in the studies presented in this thesis.

2.5.1 Effective functions (EFs)

A BF f with k inputs is said to be an *effective function* (EF) [81] if each input x_i produces a change in the output under at least one specific context formed by all other inputs. Mathematically,

$$\forall i \in \{1, 2, \ldots, k\}, \exists\, \mathbf{x} \in \{0,1\}^k \text{ with } x_i = 0, f(\mathbf{x}) \neq f(\mathbf{x} + \mathbf{e}_i). \tag{2.2}$$

Here $\mathbf{e}_i \in \{0,1\}^k$ denotes the unit vector associated with the component of index i having all entries set to 0 except for entry i.

The effectiveness of a BF in all its inputs is a necessary criterion for being biologically meaningful but not a sufficient one. Indeed, as k increases, the fraction of BFs that are effective goes to 1 [51]. Hence in practice EFs can be considered to be 'random', as we do in the studies presented

in this thesis.

2.5.2 Effective and unate functions (EUFs)

A BF f with k inputs is a *unate function* (UF) [55] if each input x_i to the BF is either monotone increasing (activatory) or monotone decreasing (inhibitory). An input x_i is increasing monotone (or activatory) if:

$$\forall\ \mathbf{x} \in \{0,1\}^k \text{ with } x_i = 0,\ f(\mathbf{x}) \leq f(\mathbf{x} + \mathbf{e}_i), \tag{2.3}$$

and decreasing monotone (or inhibitory) if:

$$\forall\ \mathbf{x} \in \{0,1\}^k \text{ with } x_i = 0,\ f(\mathbf{x}) \geq f(\mathbf{x} + \mathbf{e}_i). \tag{2.4}$$

If all inputs to a BF are positive monotone, the BF is said to be a positive monotone BF. *Effective and unate functions* (EUFs) are EFs that are also unate.

2.5.3 Read-once functions (RoFs)

A k-input BF is said to be a *read-once function* (RoF) [108] if by using only the operations of disjunction, conjunction and negation, the BF can be written so that each of the k variables appears exactly once in that expression. Mathematically, this may be succinctly stated as:

$$f(\mathbf{x}) = X_{\pi(1)} \odot X_{\pi(2)} \odot X_{\pi(3)} \ldots \odot X_{\pi(k)}, \tag{2.5}$$

where $X_{\pi(i)} \in \{x_{\pi(i)}, \overline{x}_{\pi(i)}\}$, $\odot$ represents the AND ($\wedge$) or OR ($\vee$) operator, π represents the permutations on the set $\{1, 2, \ldots, k\}$ and $\pi(i)$ is the i^{th} element of the set. Note that, for legibility of Eq. (2.5), we have omitted all parentheses necessary to completely specify the BF. But given that expression, clearly RoFs are a subset of EUFs.

2.5.4 Nested canalyzing functions (NCFs)

A BF f with k inputs is said to be canalyzing with respect to an input x_i if there exists values $a, b \in \{0,1\}$ such that fixing the variable x_i to the value a forces the output of the functions to be b,

regardless of the values taken by other input variables. Formally, this condition can be written as

$$f(x_1, x_2, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_k) = b \quad \forall x_j,\ j \neq i$$

Here, a is referred to as the canalyzing input value and b as the corresponding canalyzed output value for the input x_i.

A BF is called *canalyzing function* (CF) if at least one of the input variables satisfies the canalyzing property [5]. Moreover, CFs can exhibit multiple canalyzing inputs in a hierarchical fashion. The total number of distinct input variables that possess the canalyzing property in such a hierarchy is termed the *canalyzing depth* of the CF [109]. For a CF with k inputs, the canalyzing depth lies between 1 and k. When the canalyzing depth equals k, the CF is referred to as a *nested canalyzing function* (NCF).

Formally, a BF f with k inputs is an NCF [7, 68, 69] if there exists a permutation π of $\{1, 2, \ldots, k\}$ such that:

$$f(\mathbf{x}) = \begin{cases} b_1, & x_{\pi(1)} = a_1, \\ b_2, & x_{\pi(1)} \neq a_1,\ x_{\pi(2)} = a_2, \\ b_3, & x_{\pi(1)} \neq a_1,\ x_{\pi(2)} \neq a_2,\ x_{\pi(3)} = a_3, \\ \vdots & \\ b_k, & x_{\pi(1)} \neq a_1,\ \ldots,\ x_{\pi(k-1)} \neq a_{k-1},\ x_{\pi(k)} = a_k, \\ \overline{b}_k, & x_{\pi(1)} \neq a_1,\ \ldots,\ x_{\pi(k-1)} \neq a_{k-1},\ x_{\pi(k)} = \overline{a}_k. \end{cases} \tag{2.6}$$

In the above equation, $a_1, a_2, \ldots, a_k$ are the canalyzing input values and $b_1, b_2, \ldots, b_k$ are the canalyzed output values for inputs $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)}$ in the permutation π of the k inputs. Here, $\overline{a}_k$ and $\overline{b}_k$ are the complements of the Boolean values a_k and b_k, respectively. Alternatively, an NCF f with k inputs can be written as

$$f(\mathbf{x}) = X_{\pi(1)} \odot_1 \left(X_{\pi(2)} \odot_2 \left(X_{\pi(3)} \odot_3 \left(\cdots \left(X_{\pi(k-1)} \odot_{k-1} X_{\pi(k)}\right) \cdots\right)\right)\right), \tag{2.7}$$

where $X_{\pi(i)} \in \{x_{\pi(i)}, \overline{x}_{\pi(i)}\}$. Note that NCFs form a specialized subset of RoFs [51], which is also

evident from Eq. (2.7).

2.5.5 Link operator functions (LOFs)

A BF f with k inputs is a *link operator function* (LOF) [110–112] if it can be represented in one of the following four forms:

$$\begin{aligned} f(\mathbf{x}) &= (x_1 \vee x_2 \vee \ldots \vee x_m) \wedge \sim (y_1 \vee y_2 \vee \ldots \vee y_n) \\ f(\mathbf{x}) &= (x_1 \vee x_2 \vee \ldots \vee x_m) \vee \sim (y_1 \vee y_2 \vee \ldots \vee y_n) \\ f(\mathbf{x}) &= (x_1 \vee x_2 \vee \ldots \vee x_m) \wedge (\overline{y}_1 \vee \overline{y}_2 \vee \ldots \vee \overline{y}_n) \\ f(\mathbf{x}) &= (x_1 \wedge x_2 \wedge \ldots \wedge x_m) \vee (\overline{y}_1 \wedge \overline{y}_2 \wedge \ldots \wedge \overline{y}_n) \end{aligned} \tag{2.8}$$

Here among the k input variables, m are activatory $(x_1, x_2, \ldots, x_m)$ and n are inhibitory $(y_1, y_2, \ldots y_n)$, with $k = m + n$. The four forms of LOFs in (2.8) are referred to as AND-NOT, OR-NOT, AND-pairs and OR-pairs respectively. Note that LOFs are also a subset of RoFs but are neither a subset nor a superset of NCFs.

For the five classes discussed here, the following hierarchies hold: EF $\supset$ EUF $\supset$ RoF, NCF $\subset$ RoF and LOF $\subset$ RoF.

2.6 Complexity of Boolean functions

A range of complexity metrics for BFs has been explored in the computer science literature [113, 114]. In this thesis, we consider two measures: Boolean complexity [57] and average sensitivity [58].

2.6.1 Boolean complexity

One way to quantify a BF's succinctness is via its so-called Boolean complexity. This measure, formulated by Feldman [57], reflects the minimal descriptive length of a BF by counting the number of literals in its shortest equivalent form. While canonical forms like conjunctive normal form (CNF) or disjunctive normal form (DNF) uniquely represent a BF, they often contain redundant terms that can be eliminated using Boolean algebra. A minimal formula is obtained by systematically applying these simplifications to collapse multiple product or sum terms into a more compact

form. For instance, the 3-input BF

$$f(x_1, x_2, x_3) = x_1x_2x_3 + x_1x_2\overline{x}_3 + x_1\overline{x}_2x_3$$

initially appears with nine literals in its canonical DNF form. However, applying Boolean algebra, one can derive a minimal equivalent expression

$$f(x_1, x_2, x_3) = x_1(x_2 + x_3)$$

which uses only three literals.

Determining an absolute minimal formula is a Σ_2^P-complete problem [115, 116], so in practice one employs a heuristic logic-synthesis approach. As described in Subbaroyan *et al.* [51], in our pipeline, each BF is first converted into four candidate representations: full DNF, full CNF, Quine-McCluskey minimized DNF and Quine-McCluskey minimized CNF [117, 118]. We then pass each candidate to the "ABC" logic-synthesis tool [119] which performs further algebraic optimization. Among the four outputs, the one with the fewest literal counts serves as a proxy for the BF's true Boolean complexity.

2.6.2 Average sensitivity

The average sensitivity [58] of a BF quantifies its propensity to change output in response to single-bit perturbations of its inputs. Intuitively, it measures how frequently the output of the function changes when a single input variable is flipped, averaged over all input variables and all possible input states. Formally, for a BF f with k inputs, the average sensitivity is defined as

$$s = \left\langle \sum_{i=1}^{k} f(\mathbf{x} \oplus e_i) \oplus f(\mathbf{x}) \right\rangle_{\mathbf{x}} \tag{2.9}$$

where $\oplus$ denotes the logical XOR operator, $e_i \in \{0, 1\}^k$ is the unit vector with a 1 in the i^{th} position and 0 elsewhere, and $\langle\cdot\rangle_{\mathbf{x}}$ denotes the average over all 2^k binary inputs $\mathbf{x}$.

For a BN, the network sensitivity, $\overline{s}$, is the mean, over all nodes of the network, of the average sensitivities of the associated BFs [53, 58].

2.7 Sampling different types of Boolean functions

For a given network structure, the total number of Boolean models that satisfy the attractor constraints – even after imposing that the BFs be biologically meaningful (biologically plausible models) – is usually astronomical [64]. Therefore, it is usually computationally infeasible to carry out exhaustive analysis on all the models of a biologically plausible ensemble, barring a few cases. In view of this fact, we resort to *sampling* the space of different biologically plausible ensembles (each one constrained by a different type of BF) in order to study the effect of the type of BF on the BN dynamics. To sample models uniformly from the space of a biologically plausible ensemble, it is sufficient to uniformly sample BFs at all nodes of the network. In other words, combining BFs that are sampled uniformly from the space of a particular type of BF (satisfying the signs of inputs and/or the attractor constraints when necessary) results in a model that is uniformly drawn from the space of all possible models. So, to sample the space of a given biologically plausible ensemble, we first need to generate at each node, the exhaustive set or a sampled set of allowed BFs. We now explain how we do so for the different types of BFs.

2.7.1 Effective functions (EFs)

For nodes with up to four inputs, it is possible to list all possible EFs. For any such list, we check which ones satisfy the fixed point constraints. The resulting list is then the set of allowed BFs in this ensemble. When confronted with genes having more than four inputs, it is infeasible to generate the exhaustive list of EFs so we sample them instead. To do so, we leverage the fact that most BFs are EFs for $k \geq 4$ [51]. Specifically, we generate a random k-input EF by randomly assigning the output of rows of the truth table to either 0 or 1, except for the rows associated with the biological attractors that are fixed by precisely those constraints. We then check whether the resulting BF is EF. If it is, we are done. If it is not, we try again until obtaining an effective BF. This rejection-based random generation of BFs samples the space of EFs uniformly, and all the more efficiently that the fraction of EFs in the space of all BFs is close to 1 (as occurs with increasing k).

2.7.2 Effective and unate functions (EUFs)

For nodes with up to six inputs, an exhaustive approach is possible. That is done by first generating the exhaustive list of all the UFs for which all inputs are activatory. Such a list is far smaller than if we allow both signs for each input. We then step through this list of UFs and for each we generate the corresponding UF with the desired sign for each input. Specifically,

(a) we flip the signs of the inputs of the UF so as to match the sign combination provided for that node in the original Boolean model.

(b) We check whether the UF obtained is effective in all of its inputs.

(c) If (b) is True, we check whether that function satisfies the biological fixed point constraints. If (b) is False, we skip (d) and go to (a) for the next BF in the list.

(d) If (c) is True, we store the resulting EUF. If (c) is False, we discard the EUF and go to (a) for the next BF in the list.

The obtained list of EUFs constitutes the allowed EUFs for the considered node.

Beyond six inputs, it is computationally infeasible to generate the exhaustive list of UFs, and consequently we require a different algorithm based on a sampling approach. The problem of sampling a UF with a given sign combination can easily be mapped to the problem of sampling a positive monotone BF following the same logic as in the enumeration case described previously. Next, we describe how we sample positive monotone BFs.

2.7.2.1 Sampling positive monotone functions

We developed a sampling algorithm for positive monotone BFs based on the observation that, on a k-dimensional hypercube, positive monotonicity forces the 'propagation' of the color of a red or blue vertex (corresponding to output values 1 or 0) to its 1-Hamming neighbors that are 'ahead' or 'behind' it, respectively. We now describe such a propagation of vertex colors on the Boolean hypercube.

The propagation of the color of a vertex due to monotonicity is illustrated in Figure 2.2. In Figure 2.2(a), the vertex 4(0100) is colored red (assigned the value 1) while the other vertices are not yet colored. The 1-Hamming neighbors of vertex 4 are (0(0000), 5(0101), 6(0110), 12(1100)). If any of the vertices 5, 6 or 12 were colored blue, then the corresponding inputs x_4, x_2 or x_1 respectively would violate positive monotonicity. But vertex 0(0000) can be colored either red or

blue without violating the positive monotonicity condition on input x_3. Thus, the vertices 5, 6 or 12 are forced to be red. Note that the 1-Hamming neighbors of vertex 4 labelled with the integers 5, 6 and 12 are "larger" than 4, i.e. can be reached from 4 using positive increases to the variables. Thus if a vertex is colored red, all "larger" nodes should also be colored red, and in particular the ones that are 1-Hamming neighbors. This ordering is easily identified provided that the vertices of Boolean hypercube are labeled by the integer equivalent of the binary string associated with the vertices. We then say that the vertices 5, 6 and 12 are 'ahead' of vertex 4 which then can propagate its color to them. This propagation is continued iteratively until no nodes are ahead of the ones colored (see bottom right panel of Figure 2.2(a)).

A similar argument of propagation of color holds if, in the initial configuration of the BF, the vertex 4 was colored blue instead of red (see Figure 2.2(b)). The only difference being that the propagation of the blue color of vertex 4 is now towards the 1-Hamming neighbors (vertex 0) that have an integer label smaller than that of node under consideration. In this case, that vertex is 0. Now suppose vertex 0 was colored red, then the input x_3 violates the condition of positive monotonicity. Thus, the vertex 0 is forced to be blue. We say that the vertex 0 is 'behind' the vertex 4. The vertex 0 does not have any vertex 'behind' it, hence the propagation of color is terminated here as shown in the bottom panel of Figure 2.2(b). In sum, if we demand that a BF is positive monotone in all its inputs, then any vertex colored red (respectively blue) can iteratively propagate its color to 1-Hamming neighbors 'ahead' of it (respectively 'behind' it).

In order to sample positive monotone BFs, we begin with a hypercube in which the vertices are not colored. We then choose a random vertex from the set of uncolored vertices (all vertices are equally probable) and color it randomly (either red or blue). The color of this vertex is then iteratively propagated till no further propagation of color is possible. If any uncolored vertices remain on the hypercube, a vertex is randomly chosen from the set of uncolored vertices, colored randomly and iteratively propagated till no further propagation of color is possible. This procedure is repeated till all the vertices of the hypercube are colored. The final outcome is a positive monotone BF. Note that this BF may or may not be effective in all its inputs. To sample positive monotone BFs that are effective in all inputs, we simply sample a positive monotone BF and check if it is an EF. If it is not, we reject the BF and start again until we obtain an EF.

Notably, our sampling algorithm can generate all the positive monotone EUFs. An immediate

question is whether the above algorithm samples positive monotone EUFs *uniformly*. To answer this question, we sampled $n = 10000$, $n = 20000$ and $n = 1000000$ positive monotone EUFs with $k = 3$, $k = 4$ and $k = 5$ inputs respectively (see Figure 2.3). The sampling is not exactly uniform but it is close enough to uniform for our purposes.

2.7.2.2 Sampling unate functions with a given sign combination that satisfies fixed point constraints

The above mentioned algorithm can be easily extended to generate UFs that satisfy predefined constraints on the BF, namely, biological fixed point constraints. We now show how this sampling can be achieved using the logic provided in Figure 2.4 on a BF for which we assume that the signs of the four inputs x_1, x_2, x_3 and x_4 are '$+--+$'. The next assumption is that the fixed points constrain the vertices 7 and 10 to have the output values 0 and 1 respectively. Note that the fixed point constraints have to be consistent with the given signs, otherwise there is no UF that will satisfy the fixed point constraints. Since the BF in Figure 2.4(a) is consistent in the manner described above, we can now begin the sampling procedure. We first flip all the inhibitory inputs to make them activatory. This is achieved by swapping the hyperplanes $x_2 = 0$ with $x_2 = 1$, and $x_3 = 0$ with $x_3 = 1$ as shown in Figure 2.4(b), thereby resulting in a positive monotone UF. We now use the method described in the previous section to sample positive monotone UFs as shown in Figures 2.4(c)-(g). Once a positive monotone BF is obtained as shown in Figure 2.4(g), the signs of the inputs are returned to the initial sign combination by re-swapping the hyperplanes $x_2 = 0$ with $x_2 = 1$, and $x_3 = 0$ with $x_3 = 1$. Note that on reswapping these hyperplanes, the fixed point constraints in Figure 2.4(h) return to their original configuration as shown in Figure 2.4(a). The final BF is thus a random UF with the desired signs for the inputs and that satisfies the fixed point constraints. As before, the BF generated may be ineffective in some inputs. We thus again check if the sampled UF is an EF. If not, we keep sampling till we find an UF that is an EF.

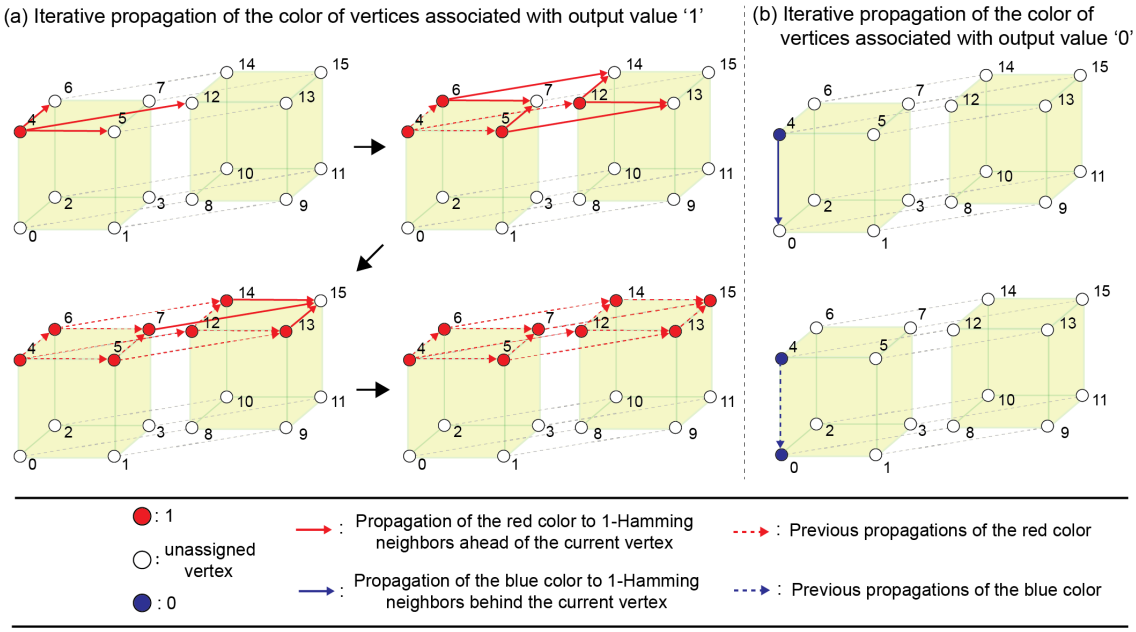


Figure 2.2: Visual illustration of how monotonicity forces propagation of the color on vertices of the Boolean hypercube. We assume that the BF is positive monotone in this illustration. We begin with a configuration on the hypercube where the vertex 4 is colored either red or blue (corresponding to output values 1 or 0 respectively) and other vertices uncolored. **(a)** Left panel: in the top left sub-figure, vertex 4 is colored red and its color is propagated in the 'forward' direction via the red solid arrows to its 1-Hamming neighbors which are 'ahead' of it (vertices 5, 6 and 12). Note that if any of the vertices 5, 6 or 12 was colored blue, the BF would violate positive monotonicity. In the sub-figure to the right (following the arrow), each of the vertices 5, 6 and 12 propagate their color to their 1-Hamming neighbors 'ahead' of them, (7, 13), (7, 14), and (13, 14) respectively. In the bottom left sub-figure, each of the vertices, 7, 13 and 14 propagate their color to their 1-Hamming neighbors, vertex 15. In the bottom right sub-figure, vertex 15 is colored red and no further propagation is possible, thereby terminating the iterative propagation. **(b)** Right panel: in the top sub-figure, vertex 4 is colored blue and its color is propagated in the backward direction via the blue solid arrows to its 1-Hamming neighbors which are 'behind' it (vertex 0). Note that if the vertex 0 was colored red, the BF would violate positive monotonicity. In the bottom sub-figure, vertex 0 is colored blue and no further propagation is possible, thereby terminating the iterative propagation.

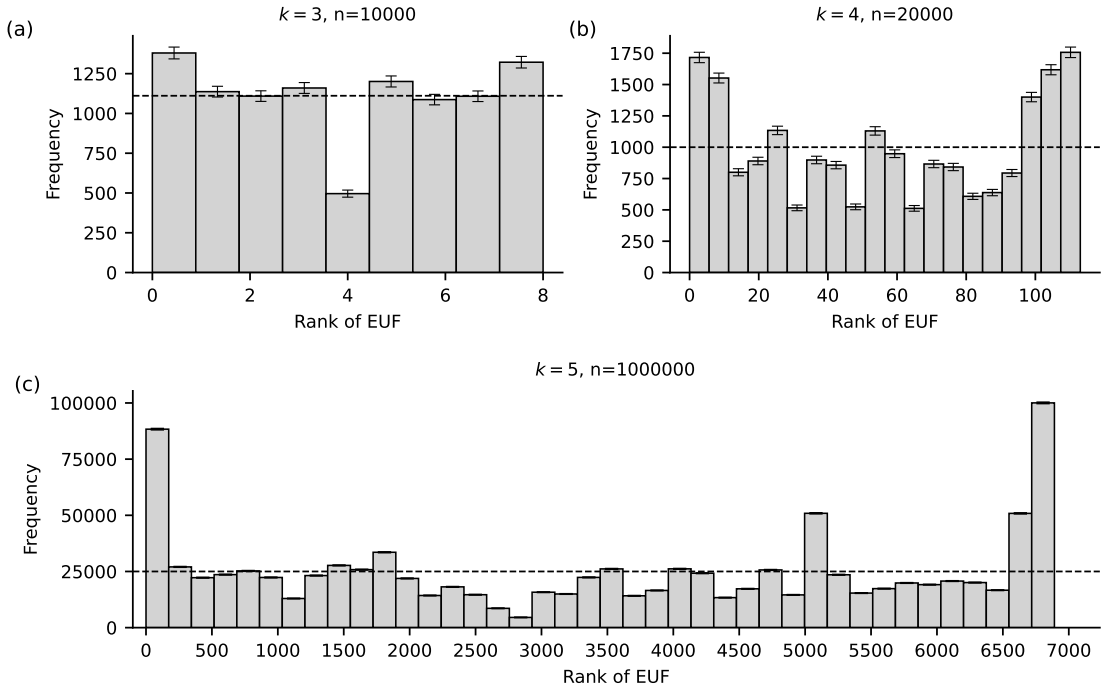
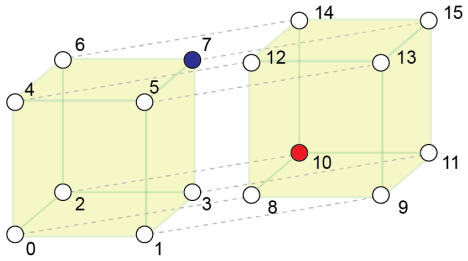
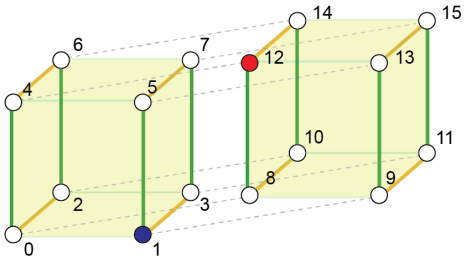


Figure 2.3: The distribution of positive monotone effective and unate functions (EUFs) sampled using the proposed EUF sampling algorithm for BFs with $k = 3$, $k = 4$ and $k = 5$. In each subplot, the x axis is the rank associated with a positive monotone EUF based on its integer encoding. Each EUF is ranked from 0 (highest rank) to the total number of distinct functions generated by the algorithm (the lowest rank). The ranking is based on the integer representation of the BF (see Section 2.4), with lower integers corresponding to higher ranks. The y axis in each subplot is the frequency of the positive monotone EUFs generated by our sampling algorithm. The subplots **(a)**, **(b)** and **(c)** shows the distribution of $n = 10000$, $n = 20000$ and $n = 1000000$ samples of $k = 3$, $k = 4$ and $k = 5$ input positive monotone EUF respectively, generated using our sampling algorithm (see Section 2.7). The horizontal dashed line represents the expected frequency value were the sampling be perfectly uniform. The error bars in all histograms are extremely tiny (sometimes too small to be visible). The distribution generated using the sampling algorithm is skewed towards the highest and lowest ranks, but is relatively close to the expected frequency value for other ranks.

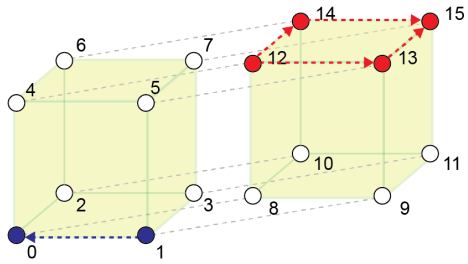
(a) BF with the fixed point constraints where the given sign combination is '+ - - +'



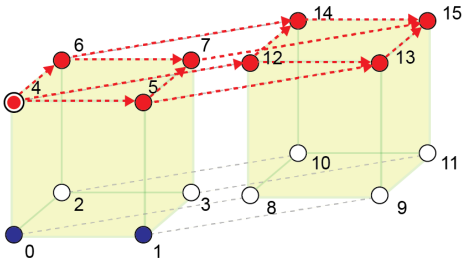
(b) Swap the hyperplanes $x_2 = 0$ with $x_2 = 1$, and $x_3 = 0$ with $x_3 = 1$ to obtain a positive monotone BF ('++++')



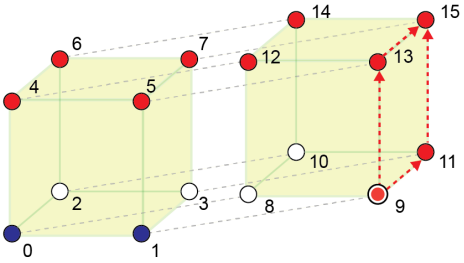
(c) Iteratively propagate the outputs of the colored vertices



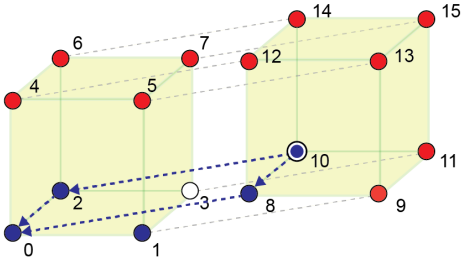
(d) Choose a random vertex (say 4) from the uncolored vertices, color it randomly (say 'red') and iteratively propagate its color



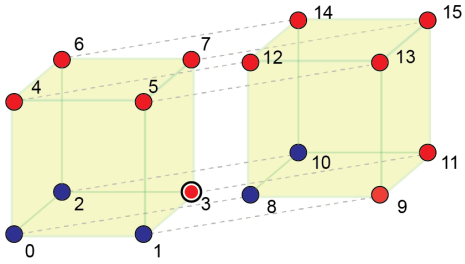
(e) Choose a random vertex (say 9) from the uncolored vertices, color it randomly (say 'red') and iteratively propagate its color



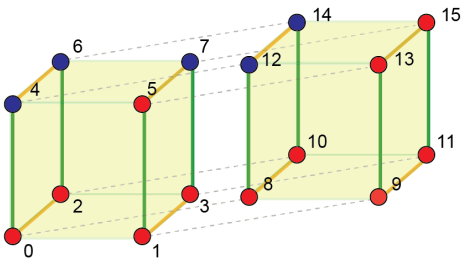
(f) Choose a random vertex (say 10) from the uncolored vertices, color it randomly (say 'blue') and iteratively propagate its color



(g) Color the remaining uncolored vertex (3) randomly (say 'red')



(h) Swap the hyperplanes $x_2 = 0$ with $x_2 = 1$, and $x_3 = 0$ with $x_3 = 1$ to get back the given sign combination ('+ - - +')



: 1

: unassigned vertex

: 0

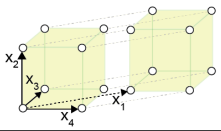
: randomly chosen vertex assigned output '1'

: randomly chosen vertex assigned output '0'

: Iteratively propagated fixed or randomly chosen 'red' colored vertices to vertices 'ahead' of it

: Iteratively propagated fixed or randomly chosen 'blue' colored vertices to vertices 'behind' it

Axes of the hypercube



Variable	Sign
x_1	+
x_2	-
x_3	-
x_4	+

Figure 2.4: Generating a random UF with a given sign combination that satisfies given fixed point constraints. We show via a 4-input BF how we generate random UFs that obey a given sign combination of the inputs and given fixed point constraints. We assume that for the given sign combination, the inputs x_1 and x_4 are activatory and inputs x_2 and x_3 are inhibitory. **(a) Initial configuration of the BF after imposing fixed point constraints.** Assume that the fixed points constrain the vertices 7 and 10 to have the output values 0 (blue) and 1 (red) respectively. **(b) Conversion of the BF to a positive monotone BF by negation of inhibitory inputs.** Inputs 2 and 3 are negated by swapping the hyperplanes $x_2 = 0$ and $x_3 = 0$ with the hyperplanes $x_2 = 1$ and $x_3 = 1$ respectively (the order does not matter). This is equivalent to exchanging the vertices along the green (or orange) colored edges that are perpendicular (or parallel) to the plane of the paper. This results in a BF where the fixed point constraints now satisfy positive monotonicity on all inputs. **(c) Iterative propagation of the vertex colors.** The color of the vertices 1 and 12 are iteratively propagated 'backward' and 'forward' respectively to vertices 'behind' and 'ahead' respectively as shown by the dashed blue and red arrows. **(d)-(g) Selection of a random uncolored vertex, coloring it randomly and iteratively propagating its color.** Since some vertices remain uncolored after step (c), a vertex is randomly chosen from the uncolored vertices (say 4) and colored randomly (say red) as shown in step (d). The color of this vertex is then iteratively propagated. At the end of step (d) since there are uncolored vertices, the above-mentioned procedure is repeated in steps (e), (f) and (g) till there are no vertices left to be colored on the hypercube. Sub-figure (g) shows the sampled positive monotone BF that satisfies the fixed point constraints. **(h) Conversion of the sampled positive monotone BF to the given sign combination.** The inputs 2 and 3 are negated again as was done in step (b) to attain the given sign combination ('+ - - +'). Following step (b), the hyperplanes $x_2 = 0$ and $x_3 = 0$ are swapped with the hyperplanes $x_2 = 1$ and $x_3 = 1$ respectively (corresponding to vertex swaps along the green (or orange) colored edges that are perpendicular (or parallel) to the plane of the paper). This BF is a random UF where the inputs obey the given sign combination and the BF satisfies the fixed point constraints.

2.7.3 Read-once functions (RoFs)

The logic here is very similar to the one for sampling EUFs. For nodes with up to eight inputs, it is possible to generate an exhaustive list of all the RoFs for which all inputs are activatory. We then step through this list of RoFs and for each function:

(a) We flip the signs of the inputs of the RoF so as to match the sign combination provided for that node in the original Boolean model.

(b) We check if the RoF obtained from (a) satisfies the biological fixed point constraints.

(c) If (b) is True, we store the resulting RoF. If (b) is False, we discard the RoF and go to the next iteration.

The obtained list of RoFs constitutes the allowed RoFs for that node.

2.7.4 Nested canalyzing functions (NCFs)

For nodes with up to nine inputs, it is possible to generate an exhaustive list of all the NCFs with all inputs being activatory. Following the logic for EUFs, we step through this list of NCFs and for each function:

(a) We flip the signs of the inputs of the NCF so as to match the sign combination provided for that node in the original Boolean model.

(b) We check if the NCF obtained from (a) satisfies the biological fixed point constraints.

(c) If (b) is True, we store the resulting NCF. If (b) is False, we discard the NCF and go to the next iteration.

The obtained list of NCFs constitutes the allowed NCFs for that node.

2.7.5 Link operator functions (LOFs)

Even for a node with ten inputs, it is computationally straightforward to generate all LOFs corresponding to a given sign combination of inputs. We iterate through this list of LOFs and, for each function:

(a) Check if the LOF satisfies the biological fixed point constraints.

(b) If (a) is True, we store the resulting LOF. If (a) is false, we discard the LOF and go to the next iteration.

The obtained list of LOFs constitutes the allowed LOFs for that node.

In summary, we developed algorithms to sample different classes of BFs, namely EFs, EUFs, RoFs, NCFs, and LOFs, subject to fixed point constraints. We exhaustively enumerate all EFs for up to four inputs and sample them for larger k. EUFs with a given sign combination are uniformly sampled for up to six inputs, while for larger input sizes we employ the iterative vertex color propagation algorithm. Whenever RoFs, NCFs, or LOFs are considered in a subsequent chapter, they are uniformly sampled with the specified sign combination. Note that, sampled EFs may or may not be consistent with the regulatory signs; those that are consistent correspond to EUFs.

If the analysis does not involve the imposition of biological attractor constraints (as in case of the study presented in Chapter 4), the sampling procedures remain unchanged, except that these constraints are omitted. In this case, for EUFs and their subclasses, we only require that the

sampled BFs at each node are consistent with the regulatory sign patterns associated with that node.

This chapter established the theoretical and methodological foundations for the studies presented in this thesis. We introduced the BN modeling framework for GRNs, reviewed random and reconstructed BN models, described biologically meaningful classes of BFs and their complexity measures, and outlined the sampling strategies used to generate ensembles of Boolean models under different biological and topological constraints. Building upon this foundation, the next chapter reviews the diverse stability measures used to characterize BN dynamics, which form the basis for the analyses presented in the subsequent chapters.

Chapter 3

Stability measures for Boolean network dynamics

Chapter overview. This chapter reviews several stability measures used to characterize the dynamics of Boolean networks (BNs). It begins with perturbation-based measures derived from damage spreading, including the Derrida coefficient, the final Hamming distance, and fragility. It then introduces stability measures based on attractor properties, followed by measures derived from polynomial approximations of Boolean functions (BFs) and the resulting notion of approximability of network dynamics. Finally, the chapter presents cellular automata-inspired measures that characterize the global organization of the state transition graph (STG), including bushiness and convergence. Together, these measures provide the methodological foundation for the analyses presented in the subsequent chapters.

Stability is a central concept in the study of Boolean network (BN) dynamics, particularly in the context of modeling biological regulatory networks, where robustness to perturbations and reliable long-term behavior are essential. However, stability in BNs is not a uniquely defined property. Instead, it is an inherently multifaceted concept whose interpretation depends on the dynamical lens through which the system is analyzed [120]. One may assess stability in terms of the response of the network to small perturbations of its state, the properties of attractors and their basins of attraction, the functional form of the underlying regulatory rules, or the global organization of the state transition graph (STG). Each of these viewpoints captures different aspects of dynamical

behavior and gives rise to distinct, yet complementary, stability measures.

Perturbation-based measures derived from damage spreading quantify whether small differences between network states tend to grow or decay over time. In particular, the Derrida coefficient (δ) [121] measures the average damage spreading after one time step and is widely used to distinguish between stable and chaotic dynamical regimes [60, 69, 109]. From a dynamical systems perspective, damage spreading can also be assessed at long times, leading to the notions of the final Hamming distance (h^{∞}) and fragility (ϕ) [59, 60]. We also consider stability measures based on attractors and their basins of attraction, such as the number of attractors, average attractor length, and the fraction of fixed point attractors. In addition, we discuss measures rooted in polynomial approximations of regulatory rules and the resulting notion of the approximability of network dynamics [76]. Finally, we introduce stability measures inspired by the theory of cellular automata, which characterize global properties of the state transition graph (STG), such as bushiness, and convergence [61–63].

3.1 Derrida's damage spreading and generalizations

A common way to test for stability of dynamics in BNs is to determine how two trajectories, initially separated by a single bit change, diverge with time. This is formalized via the so-called Derrida map [122] connecting the magnitude of the distance between the two trajectories (sometimes called *damage*) at time $t + 1$ to that at time t. If the initial damage tends not to spread, the system is considered to be *stable*, whereas the opposite case is taken as a signature of *chaotic* behavior. We consider three measures of stability [60] that are based on damage spreading: (a) the Derrida coefficient (δ), (b) the final Hamming distance (h^{∞}), and (c) the fragility (ϕ).

The Derrida coefficient, δ, is just the average damage size after a single step of the dynamics. If $\delta > 1$, the damage is likely to spread to a finite fraction of the network, while if $\delta < 1$, one expects the damage to die out, in which case the system may be considered to be in the ordered regime. Note that δ is equal to the network sensitivity ($\overline{s}$) of the system (see Chapter 2, Section 2.6) [60, 79], and hence, depends only on the BFs and not on the network architecture.

The measure "final Hamming distance", h^{∞}, quantifies the long-term average separation be-

tween trajectories. It is mathematically defined as

$$h^{\infty} = \left\langle \frac{1}{N} \sum_{i=1}^{N} \left\langle \|X(t) - X^{(\neg i)}(t)\|_1 \right\rangle_{t\to\infty} \right\rangle_{X\in\mathcal{T}} \tag{3.1}$$

Here, $X(t)$ is the network state at time t without the initial perturbation while $X^{(\neg i)}(t)$ is the state that is obtained by evolving the state that initially had the single bit (i^{th} bit) perturbation. $\mathcal{T}$ corresponds to the set of all trajectories. The symbol $\langle\cdot\rangle_{X\in\mathcal{T}}$ denotes the average taken over all possible initial network states and $\|X(t_f) - X^{(\neg i)}(t_f)\|_1$ represents the L1-norm, i.e., the Hamming distance, between the states $X(t_f)$ and $X^{(\neg i)}(t_f)$ at time t_f. Finally, $\left\langle \|X(t) - X^{(\neg i)}(t)\|_1 \right\rangle_{t\to\infty}$ denotes a time average taken in the limit of long times.

Given that h^{∞} is sensitive to phase shifts arising in cyclic attractors, Park *et al.* [60] proposed a modification to Eq. (3.1), leading to what they call "fragility", defined by

$$\phi = \left\langle \frac{1}{N} \sum_{i=1}^{N} \|\langle X(t)\rangle_{t\to\infty} - \langle X^{(\neg i)}(t)\rangle_{t\to\infty}\|_1 \right\rangle_{X\in\mathcal{T}} \tag{3.2}$$

The difference between ϕ and h^{∞} is that in the case of ϕ the time average is taken inside the norm, a choice that removes the sensitivity to the phase shift in cycles.

The Derrida coefficient is equivalent to the average network sensitivity [60]. Therefore it can be decomposed into characteristic of individual BFs and so for given values of k, its complexity remains linear in N. In contrast, the h^{∞} and ϕ cannot be simplified in this manner. Given that they depend on both the attractor lengths and the time to reach the attractors, they probably belong to the same computational complexity class as the determination of the attractors themselves (as we discuss in the next Section 3.2).

3.2 Stability measures based on properties of the attractors

Further, in studying dynamical stability, we considered several observables related to attractors and their basins of attraction, as outlined below:

(i) The total number of attractors: This is obtained by summing the number of fixed points and the number of cyclic attractors.

(ii) Fraction of fixed point attractors: This is the ratio of the number of fixed point attractors to the total number of attractors. If there are a total of n attractors, out of which m are fixed point attractors, then this quantity is given by m/n.

(iii) Fraction of states in the basins of fixed points: This is the ratio of the number of states converging to fixed point attractors to the total number of states, 2^N. If for an N-node BN, there are m fixed point attractors with the basin sizes (number of states in the basin of attraction) $W_1, W_2, \ldots, W_m$ respectively, then the quantity we consider is given by $\sum_{i=1}^{m} W_i/2^N$.

(iv) Average attractor length: The length of an attractor refers to the number of states that constitute the attractor. Fixed points are attractors of length 1, whereas cyclic attractors can have lengths ranging from 2 to 2^N in an N-node BN.

(v) Maximum attractor length: This quantity refers to the length of the largest attractor.

(vi) Weighted average attractor length: If for an N-node BN, there are n attractors, respectively with attractor lengths $L_1, L_2, \ldots L_n$ and basin sizes $W_1, W_2, \ldots W_n$ (summing up to 2^N), then the weighted average attractor length is given by

$$L_W = \frac{W_1}{2^N} L_1 + \frac{W_2}{2^N} L_2 + \ldots + \frac{W_n}{2^N} L_n = \frac{1}{2^N} \sum_{i=1}^{n} W_i L_i \tag{3.3}$$

(vii) Fraction of states on attractors: For an N-node BN, this quantity refers to the fraction of the network states belonging to attractors. If there are n attractors with attractor lengths $L_1, L_2, \ldots L_n$, then the quantity is given by $\sum_{i=1}^{n} L_i/2^N$.

The stability measures discussed in this section involve identifying all the attractors and/or the sizes of their basins of attraction. To achieve this, we utilize the R library “BoolNet" [123], which allows an exhaustive identification of the attractors and their corresponding basins for a given model. It is known that finding a fixed point attractor is NP-hard and counting the total number of fixed points is #P-complete [124]. Consequently, most observables based on attractor properties are likely to belong to those same complexity classes. Due to excessive computational demands, analyzing significantly larger networks for large ensembles of models remains prohibitive.

3.3 From polynomial approximations of regulatory rules to the "approximability" of network dynamics

In continuous dynamical systems, major insights follow from linearizing the equations of motion. In the case of BF, the only way to have a linear update rule is for it to depend on a single variable. Therefore, BFs of multiple variables are intrinsically nonlinear. Nevertheless, some BFs may be more nonlinear than others. A question thus immediately comes to mind: do the different classes of BFs (e.g., EFs, EUFs, RoFs, NCFs, LOFs) have significantly different levels of nonlinearities? To address this question, we follow the framework of Manicka *et al.*, [76] who in effect use the Glauber formula [125] to write a BF, B as a sum of polynomial terms:

$$B(x_1, x_2, \ldots, x_k) = \sum_{t_1} \sum_{t_2} \cdots \sum_{t_k} \prod_i P(x_i, t_i) B(t_1, t_2, \ldots, t_k) \tag{3.4}$$

where

$$P(x_i, t_i) = \frac{1 + 4(x_i - 1/2)(t_i - 1/2)}{2} \tag{3.5}$$

is a Kronecker delta function on Boolean variables, but written so that it can also be applied to real variables (note that the product of these Kronecker delta functions is analogous to the use of Lagrange interpolation polynomials). Eq. (3.4) then directly provides a polynomial representation of the BF to which one can apply the standard tools of continuous dynamical systems. To proceed, the polynomial of Eq. (3.4) is Taylor expanded about the mid-point of the hypercube (all $x_i = 1/2$) and then truncated at the desired order. The order 0 case leads to the constant approximation, the order 1 case to the linear approximation, etc. As a result, the degree of nonlinearity of a BF can be quantified by the error introduced when using the degree 1 approximation. More generally, for any degree used in the truncation, one can define an error via the square of the difference between the value of the BF and that of the truncated polynomial when averaging over all 2^k possible input values. We call this the mean error of truncation and it depends only on the BF and the truncation degree. The mean error of truncation of a BF when the truncation degree is 1 will be used as a measure of nonlinearity of that BF, thereby allowing us to compare the nonlinearities across the different classes of BFs.

Manicka *et al.* [76] were interested in how well a truncated dynamics could approximate the true Boolean dynamics at the level of the whole network. Rather than considering the approximation for a single time step, they focused on the long-term behavior, imposing that the error be evaluated after 500 time steps. They call this the MAE (for mean approximation error) where the mean is obtained by averaging over as many random initial conditions as possible and over all the variables of the network. For the MAE-associated results reported in Chapter 4 of this thesis, we followed this prescription exactly when comparing the behavior of different classes of BFs.

3.4 Cellular automata inspired stability measures for Boolean networks

A cellular automaton (CA) is a self-contained discrete dynamical system. It consists of cells (or sites) arranged on a regular lattice and that are assigned values from a finite alphabet. These values are synchronously updated at each time step according to a function that takes as inputs the values of cells in a fixed neighborhood of each cell.

Wuensche considered the simple case of 1-dimensional (1D) CA consisting of a 1D ring of N sites with each site carrying a value '0' or '1' [62]. The values C_i at each site i get updated synchronously in discrete time steps according to a Boolean rule f whose inputs are the values of i's neighboring sites but can include the site i itself.

If we consider a 1D lattice with N cells and a size K neighborhood, then the time evolution of the i^{th} cell can be represented by [62]

$$C_i^{(t+1)} = f\left(C_{i-r_{left}}^{(t)}, \ldots, C_{i-1}^{(t)}, C_i^{(t)}, C_{i+1}^{(t)}, \ldots, C_{i+r_{right}}^{(t)}\right)$$

where, to satisfy the periodic boundary conditions, for $x < 1$, $C_x = C_{N+x}$ while for $x > N$, $C_x = C_{x-N}$. Here, r_{left} and r_{right} are the neighborhood radii (number of neighbors) to the left and right respectively, and thus $K = r_{left} + 1 + r_{right}$.

A 1D CA architecture can be viewed as a network in which the cells (or sites) correspond to the nodes and the interacting K-neighbors as the input edges to the nodes. At any instant, each cell (or site) holds a value either '0' or '1'. The arrangement of these 0s and 1s across the entire

network forms the network's current configuration or *state*, *i.e.*, the list of values C_i at all sites, which can be represented by a binary string. At each time step, the synchronous update transforms the current state to the subsequent one, a process that can be iterated indefinitely. A time-ordered sequence of states generated in this manner is referred to as a *trajectory*. Having a finite state space of size 2^N, a trajectory will either converge to a fixed point or to a limit cycle, both of which are termed *attractors*.

Alike BNs, the dynamics of a CA can be represented by a state transition graph (STG), in which each state has exactly one *successor* but may have a variable number of immediate *pre-images* or *predecessors*, ranging from 0 to 2^N (see Chapter 2, Section 2.1). The number of pre-images of a state is called its in-degree.

Since the STGs arising from this CA architecture and from synchronous BN dynamics share the same structural properties, we characterize BN dynamics using mathematical measures that originate from the analysis of STGs in CA. In particular, we build on the work of Wuensche [61–63] and define the global measures of *bushiness* and *convergence* below. Wuensche also introduced another quantity based on the Boolean update rule acting at each site of the 1D CA lattice, known as the Z-parameter [62, 78, 126]. Here, we present an intuitive definition of Z-parameter in the context of CA. Its adaptation to BNs is deferred to Chapter 5.

3.4.1 Global measures of bushiness and convergence of the STG

We quantify 'bushiness' of the STG by defining two quantities: (a) G-density and (b) average in-degree of the *non-GoE* states. ***G*-density** is defined as the fraction of garden-of-Eden (*GoE*) states (see Chapter 2, Section 2.1). This quantity captures the degree of contraction or irreversibility of the dynamics, measuring the fraction of states removed from phase space after a single update. The average in-degree of *non-GoE* states is defined as the average of the number of predecessors (in-degree) of *non-GoE* states, a quantity we denote by $\langle d_{in} \rangle_{non-GoE}$.

Consider a GRN with N nodes (genes) and associated update rules. Then, the associated STG will have 2^N states. Suppose, the number of *GoE* states is G. Then the G-density associated with the STG is given by

$$G\text{-density} = \frac{G}{2^N}$$

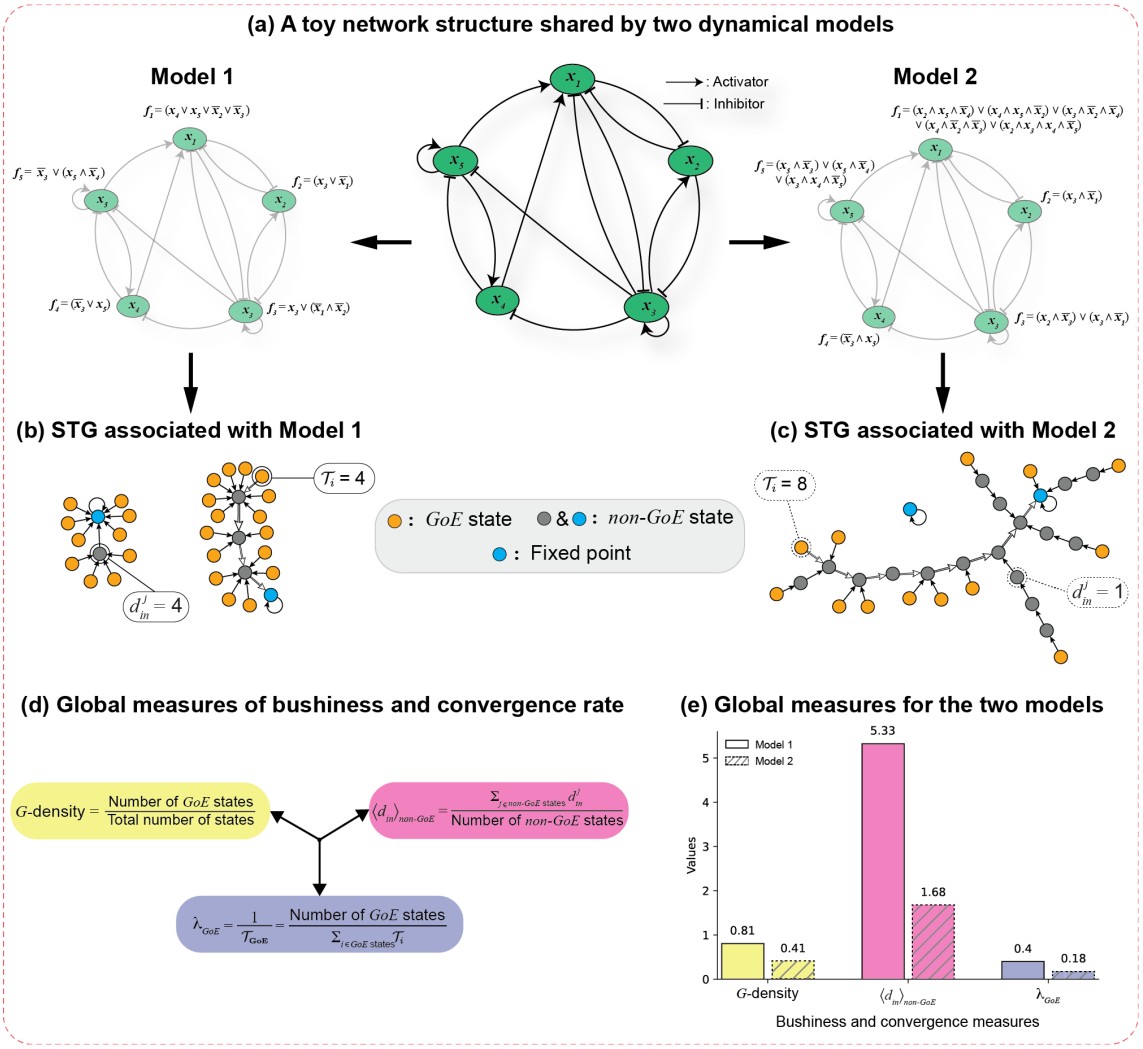


Figure 3.1: Measures of 'bushiness' and 'convergence' of STGs. **(a)** In the center is a toy network structure with 5 nodes and 14 edges. On its two sides are Boolean models (Model 1 and Model 2) with identical network structures but differing regulatory logic rules (as shown at each node). **(b)** and **(c)** display the complete STGs containing 32 states for both Model 1 and Model 2. Visual inspection shows that the STG for Model 1 appears significantly more bushy than the STG for Model 2. The transient path from a chosen *GoE* (encircled) to the attractor is traced via the unfilled arrow heads. $\mathcal{T}_i$ denotes the transient lengths (steps to reach the attractor) for the i^{th} *GoE* state (encircled *GoE* states in the subplot) in the STG. d_{in}^j denotes the in-degree of the j^{th} *non-GoE* state (encircled grey *non-GoE* states in the subplot) in the STG. **(d)** The global measures that quantify the 'bushiness' of the STG are *G*-density and average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$), whereas the measure that quantifies the 'convergence' is the average convergence rate of trajectories originating from *GoE* states (λ_{GoE}). **(e)** Barplots representing the values of the bushiness and convergence measures described in (d) for Model 1 and Model 2. The x and y axes correspond to the three measures and their values respectively. The bars for the two models for the same quantities have been shown together to illustrate quantitatively the visible differences between (b) and (c).

In a STG, each state has exactly one successor and so the number of edges in the STG is equal to 2^N. Since, the number of *GoE* states is G, number of *non-GoE* states must be $2^N - G$. Hence, the average in-degree of the *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) is given by the total number of edges divided by the total number of *non-GoE* states:

$$\begin{aligned} \langle d_{in} \rangle_{non-GoE} &= \frac{2^N}{2^N - G} \\ &= \frac{1}{1 - \frac{G}{2^N}} \\ &= \frac{1}{1 - G\text{-density}} \end{aligned}$$

So, $\langle d_{in} \rangle_{non-GoE}$ can be expressed as a function of the G-density as follows:

$$\langle d_{in} \rangle_{non-GoE} = \frac{1}{1 - G\text{-density}} \tag{3.6}$$

We quantify the 'convergence' of the STGs by defining the *average convergence rate* of trajectories originating at *GoE* states ($\boldsymbol{\lambda_{GoE}}$) as

$$\lambda_{GoE} = 1/\mathcal{T}_{GoE} \tag{3.7}$$

where $\mathcal{T}_{GoE}$ is the average of the lengths of transients originating at *GoE* states. All states that are not members of the attractor are called 'transient' states.

To illustrate how these measures can be computed, we take simple cases shown in Figure 3.1(a) that are based on toy models (model 1 and model 2) which share the same network structure but differ in the BFs used. Their associated STGs (see Figures 3.1(b)-(c)) show contrasted levels of bushiness and convergence. Clearly, the three measures computed on the STG (see Figure 3.1(d)) of the toy model 1 are higher than that for the STG of model 2 as provided in the barplots (see Figure 3.1(e)). From these, the reader may expect the bushiness and convergence to go hand in hand but in fact the situation is more subtle as we will show in Chapter 5.

As a variant to λ_{GoE}, we also consider the reciprocal of the average of the transient lengths of trajectories originating at *all* states of the STG, which we denote as $\lambda_{all} (= 1/\mathcal{T}_{all})$. This quantity has the advantage that it can be estimated numerically via sampling, and we denote the associated

(stochastic) estimator as $\lambda_{random}(= 1/\mathcal{T}_{random})$.

3.4.2 The Z-parameter in Cellular Automata

Various quantities have been introduced in CA based on properties of the rule tables, in particular to characterize the global nature of CA dynamics [88,127,128]. One such parameter is Wuensche's Z-parameter which has been shown to be a useful predictor for the bushiness of the STG [62] in the context of CA. Wuensche proposed the Z-parameter as an indicator of irreversibility of the discrete dynamics [62, 63, 78, 91]. The starting point considers an arbitrary CA state and asks which could be its pre-images. Wuensche tries to determine these pre-images by exploiting the 1D architecture of the CA, attempting to infer the values of consecutive cells by shifting a window from left to right. For a partially known pre-image, three different scenarios may arise based on the rule applied while determining the next unknown cell in the pre-image [126]:

(a) determined (the value of the unknown cell is uniquely determined, leading to the continuation to the next unknown cell)

(b) ambiguous (both '0' and '1' are possible for the unknown cell, leading to the continuation to the next unknown cell for both possibilities) and

(c) forbidden (neither '0' nor '1' satisfies the condition, resulting in the dismissal of the partial pre-image).

Wuensche introduced the parameter Z_{left} as the probability that the value of the next unknown cell to the right of a given partial pre-image is uniquely determined, corresponding to locally reversible dynamics. Similarly, the parameter Z_{right} is defined as the probability that the value of the next unknown cell to the left of the partial pre-image is uniquely determined. The Z-parameter is then defined as the maximum of these two quantities [62]. The probabilities are calculated from the CA rule table by determining the fraction of configurations where the uniqueness condition holds. In the case of BNs, there is no inherent spatial ordering of the inputs to a node, and the BFs associated with different nodes need not be identical. Consequently, the definition of the Z-parameter must be extended to apply to BNs. In Chapter 5, we present our adaptation of Z-parameter to BNs.

This chapter reviewed the diverse stability measures used to characterize BN dynamics from complementary perspectives, including perturbation-based measures, attractor properties, approximability, and cellular automata-inspired descriptors. Together, these measures provide a compre-

hensive framework for quantifying the stability and dynamical behavior of BNs. These measures form the methodological basis for the systematic analyses of random and reconstructed BNs presented in the subsequent chapters.

Chapter 4

Choices of regulatory logic class modulate the dynamical regime in random Boolean networks

This chapter is based on the published manuscript [77].

Chapter overview. This chapter systematically investigates how different classes of Boolean functions (BFs) influence the stability of random Boolean networks (RBNs). We compare five biologically relevant classes of BFs across multiple network topologies, average connectivities, and network sizes using the stability measures introduced in Chapter 3. The chapter first describes the construction of the RBN ensembles and the statistical framework used for comparing stability measures. It then examines how the choice of regulatory logic affects network stability, investigates the influence of network connectivity and size on the observed stability hierarchy, and concludes with a discussion of the implications of these findings for modeling biological regulatory networks.

Since the advent of RBNs to model gene regulatory networks (GRNs) [19, 24–27], much work has focused on the dynamics of RBNs to determine which characteristics of such systems ensure stability or robustness [41–43, 129]. In particular, Kauffman studied the dynamics of RBNs. He provided some insights on the number and length of attractors as a function of the number (N)

of nodes in the RBN [19]. Each node in his N-node random network is influenced by k other random nodes in the network via a randomly chosen logic rule from the set of all BFs with k inputs. Kauffman suggested that the networks with fixed in-degree (k) of 2 were critical networks operating at the *edge of chaos* and hypothesized that the number of attractors in a RBN increases as $\sqrt{N}$. It was also claimed that in RBNs, the typical length of an attractor grows asymptotically as a power of N or slower. However, both these conjectures were eventually found to be inaccurate. Specifically, it was shown that both the number of attractors and the mean attractor length grow [94–96] faster than any power law of N. Further, Kauffman also found that for $k > 2$, the behavior of the system was chaotic [5].

Subsequently, there has been considerable research on the stability of the dynamics in RBNs when varying network topology as well as the choice of logic rules or BFs [41–43, 129]. For instance, Kauffman *et al.* [41] found that networks with nested canalyzing rules are stable in terms of the total sensitivity of BFs in Boolean networks (BNs) with power law and exponential in-degree distributions. Further, Murrugarra and Laubenbacher [42] analyzed two ensembles of random networks–one employing random rules and the other employing nested canalyzing rules. Their findings revealed that networks using nested canalyzing functions (NCFs) exhibit more robust dynamics, are characterized by fewer attractors whose lengths are also shorter compared to what happens in networks with random rules. Moreover, Subbaroyan *et al.* [51] have shown that NCFs are highly preponderant in the reconstructed biological BNs and NCFs exhibit the minimum average sensitivity for any number of inputs. A meta-analysis of reconstructed BNs for diverse biological systems revealed that those models exhibit more canalyzation and stable dynamics than expected [51, 53, 56]. Furthermore, previous studies [33, 60] have also claimed that cells operate in the ordered regime, remaining far from the edge of chaos.

In studying the impact of BFs on stability of RBNs, significant attention has been given on canalyzing functions (CFs) [41, 43, 129]. Some previous efforts have also touched on the properties when restricting to its NCF subclass [41, 42, 51]. In this chapter, we present a detailed and systematic investigation comparing the effect of five different classes of BFs, namely *Effective Functions* (EFs) [81], *Effective Unate Functions* (EUFs) [55], *Read-once Functions* (RoFs) [108], *Nested Canalyzing Functions* (NCFs) [7, 68, 69] and *Link Operator Functions* (LOFs) [110–112] on the stability of RBNs. We also investigate whether the observations change with topological

variations, be it through modifications of the average number of inputs, the in-degree and out-degree distributions, or the network size. Figure 4.1 provides an overview of the systematic study presented in this chapter. Specifically, the figure illustrates the different network topologies considered, the range of network sizes and average in-degrees chosen. It also highlights the classes of BFs and the stability measures evaluated in this study for all these different classes.

4.1 Ensembles of RBNs: constraining network topologies and restricting Boolean functions to a class

Past studies on RBNs used random BFs at each node but made more specific choices for the network architecture, i.e., who drives whom, generically referred to as the topology of the network. The two most common choices are as follows. In the first, one assigns (oriented) interactions between randomly selected nodes, subject to a given average number λ of edges per node. This procedure corresponds to the Erdös–Rényi [130] ensemble (but with oriented edges), hereafter denoted P-P, as when the number of nodes becomes large, the distribution of in-degrees (k_{in}) and out-degrees (k_{out}) becomes Poissonian with parameter λ. In the second, one imposes that both the in- and the out-degrees be fixed, corresponding to what is called the k-regular graph ensemble, hereafter denoted R-R. That choice was motivated [131] by the desire to ensure that the nodes had comparable update rules since two rules with different numbers of inputs are difficult to compare. In addition to those two ensembles of network architectures, in this chapter, we consider a mixed case where one fixes the in-degree to one particular value [27], leaving free the out-degree so it stays Poissonian. We will call it the R-P ensemble and it will allow us to test the relevance of fluctuations in the out-degree on the network dynamics.

In addition to controlling network architecture, we also control the type of BFs used at the nodes of the network. Specifically, within each of the 3 topological ensembles introduced (P-P, R-R and R-P), we do not consider random logical update rules but instead rules that have particular properties such as unateness or are canalyzing. Thus, we impose that each node in the network is assigned a rule at random within the considered class, namely, EF, EUF, RoF, NCF, and LoF (see Chapter 2, Section 2.7). Note that for all practical purposes, the EF class essentially corresponds to random BFs, the only constraint being that those BFs be effective; when the number of inputs

is 3 or more, this constraint is nearly always satisfied for a random BF [51].

As a technical point, the assignment of BFs to nodes requires that we be able to randomly draw BFs with a uniform measure within each of the five classes considered. Unfortunately, this becomes computationally infeasible for certain classes such as EUF when the in-degree becomes larger than six. Thus, in the case of the P-P ensemble, we can be confronted with situations where the in-degree is too large to guarantee a uniform drawing. If so, we resort to a sampling that only approximates the uniform case but that has nevertheless been shown to deviate little from uniformity (see Chapter 2, Section 2.7).

For our study, we generate the ensembles of RBNs as follows:

(i) We use 1000 networks in each case.

(ii) We consider networks with number of nodes $N = 12, 16$ and 20 and analyse different properties associated with the attractors and MAE of the network.

(iii) δ, h^{∞} and ϕ are computed for networks with $N = 12$.

(iv) The in-degrees (or mean in-degrees in case of Poisson in-degree distribution) considered are 2, 3, 4 and 5. Nevertheless, in the case of Poisson degree distribution, the in-degree is kept within the range 1 and 7. The value 1 is to ensure that the node has a BF attached to it (this corresponds to the so called zero-truncated Poisson distribution) while 7 is to ensure that we can uniformly (or almost uniformly) sample BFs in each class as explained above.

(v) The classes of functions considered are EF, EUF, RoF, NCF and LOF.

(vi) We generate 1000 models each for every class per network to study different network attractor properties.

(vii) We generate 10 models each for every class per network topology for the computation of δ, h^{∞}, ϕ and MAE.

4.2 The Jensen–Shannon distance and statistical tests for comparing distributions

In this chapter, we use an ensemble approach, and thus a stability observable can be thought of as a random variable. We seek to characterize the associated distributions for different classes of BFs and compare them. A standard way to compare two distributions is to implement a binning

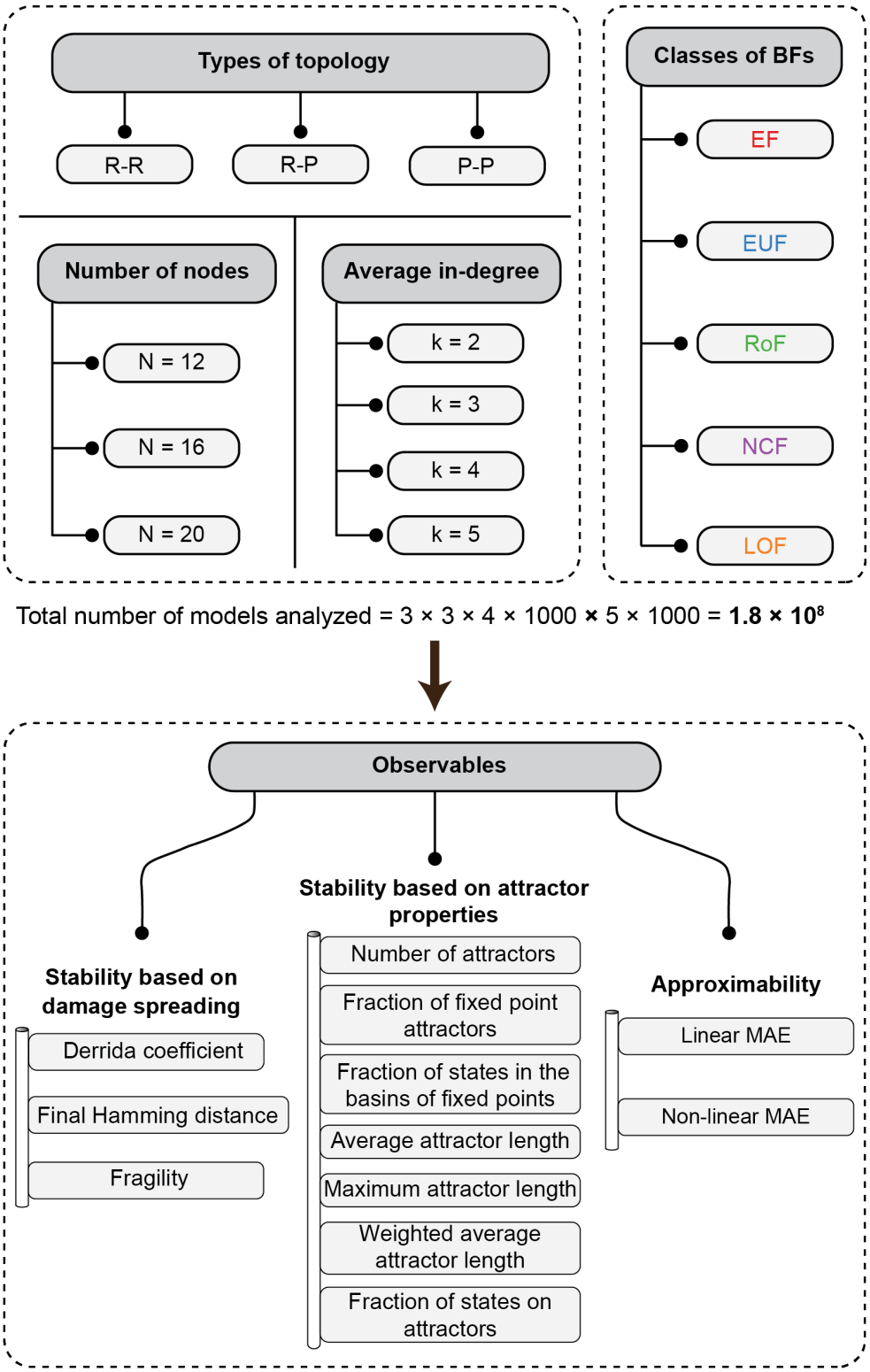


Figure 4.1: Summary of the design of the study described in Chapter 4. This figure provides an overview of the study conducted in Chapter 4, outlining the key components. The left upper panel illustrates the different network characteristics considered, including variations in network size and average in-degree. The right upper panel highlights the classes of Boolean functions (described in Chapter 2) chosen for the analysis. The lower panel presents different observables and stability measures (described in Chapter 3) examined in the analysis across different network structures and logic rules.

and measure the Kullback–Leibler (KL) [132] divergence of the resulting histograms. The KL divergence, denoted by $D_{KL}(H_1||H_2)$ is a type of statistical distance, that quantifies how much a model distribution H_2 deviates from the reference distribution H_1. Mathematically, it is defined as

$$D_{KL}(H_1||H_2) = \sum_i p_i^{(1)} \log_2(p_i^{(1)}/p_i^{(2)}) \tag{4.1}$$

where $p_i^{(1)}$ and $p_i^{(2)}$ are the respective probabilities of falling in bin i according to the histograms H_1 and H_2. $D_{KL}(H_1||H_2)$ vanishes when the two histograms are identical and it rises as the difference between the two histograms increases. Note that in general $D_{KL}(H_1||H_2) \neq D_{KL}(H_2||H_1)$, i.e., this divergence is not a distance and not even symmetric. Thus, to follow common practice, we use in all our comparisons a symmetric quantity, the so-called Jensen–Shannon (JS) distance [133, 134]:

$$D_{JS}(H_1, H_2) = \sqrt{\frac{D_{KL}(H_1||(H_1+H_2)/2) + D_{KL}(H_2||(H_1+H_2)/2)}{2}} \tag{4.2}$$

This has the remarkable property of really providing a distance, that is, the order of the arguments does not matter and this measure satisfies the triangle inequality. Eq. (4.2) allows us to quantify how different two distributions are via a *distance* that conveniently lies in the range [0, 1].

Given any two empirical distributions, i.e., finite samples each drawn from their own distribution, one can test whether they might have been drawn from the same distribution. The Python function scipy.stats.ks_2samp provides a p-value for that hypothesis using the standard Kolmogorov–Smirnov (K-S) test. What is more interesting is the prospect of testing the possibility that the first distribution dominates the second one, i.e., its cumulative (or *integrated*) distribution is everywhere higher than that of the other. A p-value for that hypothesis is provided by the scipy.stats.ks_2samp function by requiring the test to be one-sided.

4.3 Stability is enhanced along our hierarchy of Boolean function classes

A key aim of the study presented in this chapter was to systematically analyze the effect of different classes of logical rules on the dynamics of RBNs. To this end, we considered a suite of stability measures described in Chapter 3 for our analysis.

The measure final Hamming distance (h^{∞}; see Section 3.1 in Chapter 3) [60] quantifies the long-term average separation between trajectories with an initial Hamming distance of 1. In Figure 4.2, each point is associated with a graph chosen at random within the R-R topology ensemble ($k_{in} = k_{out} = 4$ and $N = 12$). The value on each axis is the average of h^{∞} over 10 different realizations of the BF rules when taken at random from the corresponding labeled class. Broadly, we observe $h^{\infty}_{EF} > h^{\infty}_{EUF} > (h^{\infty}_{RoF}$ or h^{∞}_{NCF} or $h^{\infty}_{LOF})$, with the last three classes RoF, NCF and LOF not following any particular ordering. This hierarchy of h^{∞} indicates that progressing from EF to EUF to any of the three remaining classes (RoF, NCF or LOF) steers the dynamics towards the ordered regime (or away from the chaotic regime).

As shown in Figure 4.3, such a hierarchy is also evident when using other measures considered to probe stability, whether they are based on damage spreading (see Chapter 3, Section 3.1), attractor properties (see Chapter 3, Section 3.2), or linear and nonlinear mean approximation errors (MAEs; see Chapter 3, Section 3.3). Indeed, these observables show a monotonic trend along the hierarchy. For instance the number of attractors decreases along the hierarchy while the fraction of fixed point attractors increases, both observables behaving as expected when dynamical systems become more stable. In all cases we see that the classes RoF, NCF and LOF have very similar distributions while the EF and EUF stand apart. Given the values of h^{∞} for EF and EUF classes, it is evident their associated dynamics is significantly more chaotic than the ones arising when using the other classes of BFs (see $h^{\infty} = 1$ boundary separating chaotic from stable dynamics in Figure 4.3(i)).

In practice, computing either h^{∞} or ϕ (fragility) requires iterating the dynamics a very large number of times, and so it is reassuring that other easier-to-compute measures provide similar results. A common proxy for h^{∞} is the Derrida coefficient δ which measures the damage spreading after a single time step (with $\delta = 1$ separating the chaotic and stable dynamical behaviors). In Figure 4.3(h), we see that δ has distributions with particularly low overlaps when comparing our five classes of BFs. To probe this property, we show in Figure A.1 in Appendix A, the analog of Figure 4.2 upon using δ. There we see that, for δ, one has the more extended hierarchy $\delta_{EF} > \delta_{EUF} > \delta_{RoF} > (\delta_{NCF}$ or $\delta_{LOF})$ but where there is no systematic ordering between the NCF and LOF classes. The low overlap of distributions is also visible in Figure A.2(a) in Appendix A, which displays the joint values of h^{∞} and δ, and moreover, this observation is also evident from

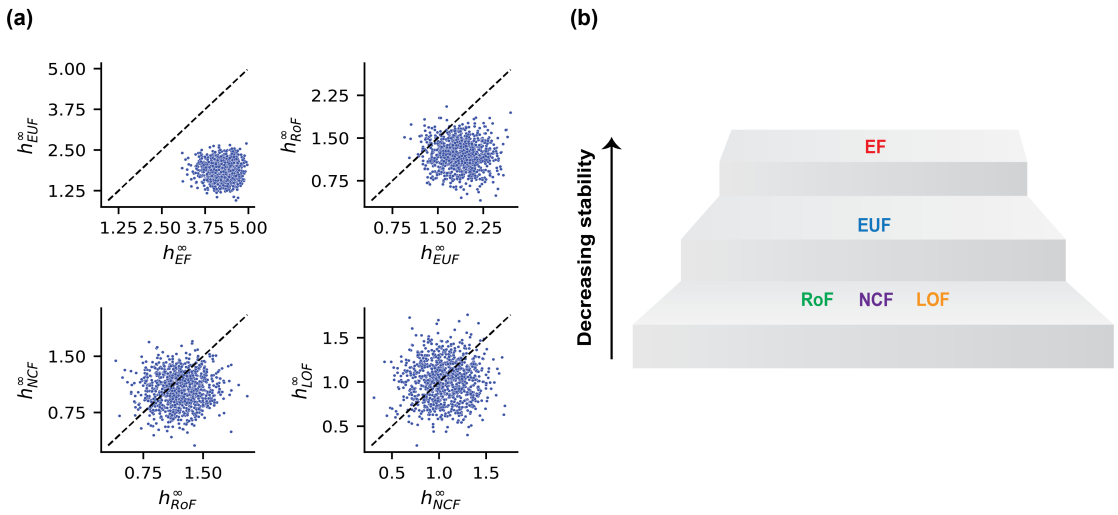


Figure 4.2: Comparison of mean values of h^{∞} when using different classes of BFs on identical networks. For each of the 1000 different networks with k-regular (R-R) topology ($k_{in} = k_{out} = 4$, $N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The values of h^{∞} were then averaged over the 10 models and plotted. (a) Each sub-figure provides comparisons across two successive classes of the putative hierarchy. The dashed line is the X=Y boundary. (b) The right panel shows a visual illustration of stability hierarchy when one uses the value of h^{∞}.

a comparison of the marginal distributions displayed as box plots. Unexpectedly, the correlation between h^{∞} and δ is relatively weak. Furthermore, while h^{∞} is directly tied to the long-term dynamics of the system and depends on the network connectivity, δ depends only on the BFs. In fact, δ is identical to the so-called *average network sensitivity* which does not capture information about the network connectivity [60, 79]. Based on our observations, using δ as a proxy for h^{∞} is not satisfactory. Interestingly, it may be preferable to use for instance the weighted average attractor length L_W (see Chapter 3, Section 3.2, Eq. (3.3)) that has a higher correlation with h^{∞} and provides similar hierarchy prediction (see Appendix A, Figure A.2(b)). Figures A.3-A.8 in Appendix A reveal that the monotonic trend for δ, h^{∞} and ϕ as observed in Figure 4.3(h)-(j) are consistent irrespective of the value of k and the type of topology. For the seven observables associated to attractors, we see a monotonic trend as shown in Figure 4.3(a)-(g). Depending on the observable considered, the trend is either monotone increasing or decreasing. For the two observables "Fraction of fixed point attractors" and "Fraction of states in the basins of fixed points", the trend is monotone increasing, whereas for the other five observables the trend is decreasing. But, in all cases, progressing from EF to EUF to other classes results in increasing stability, i.e., one

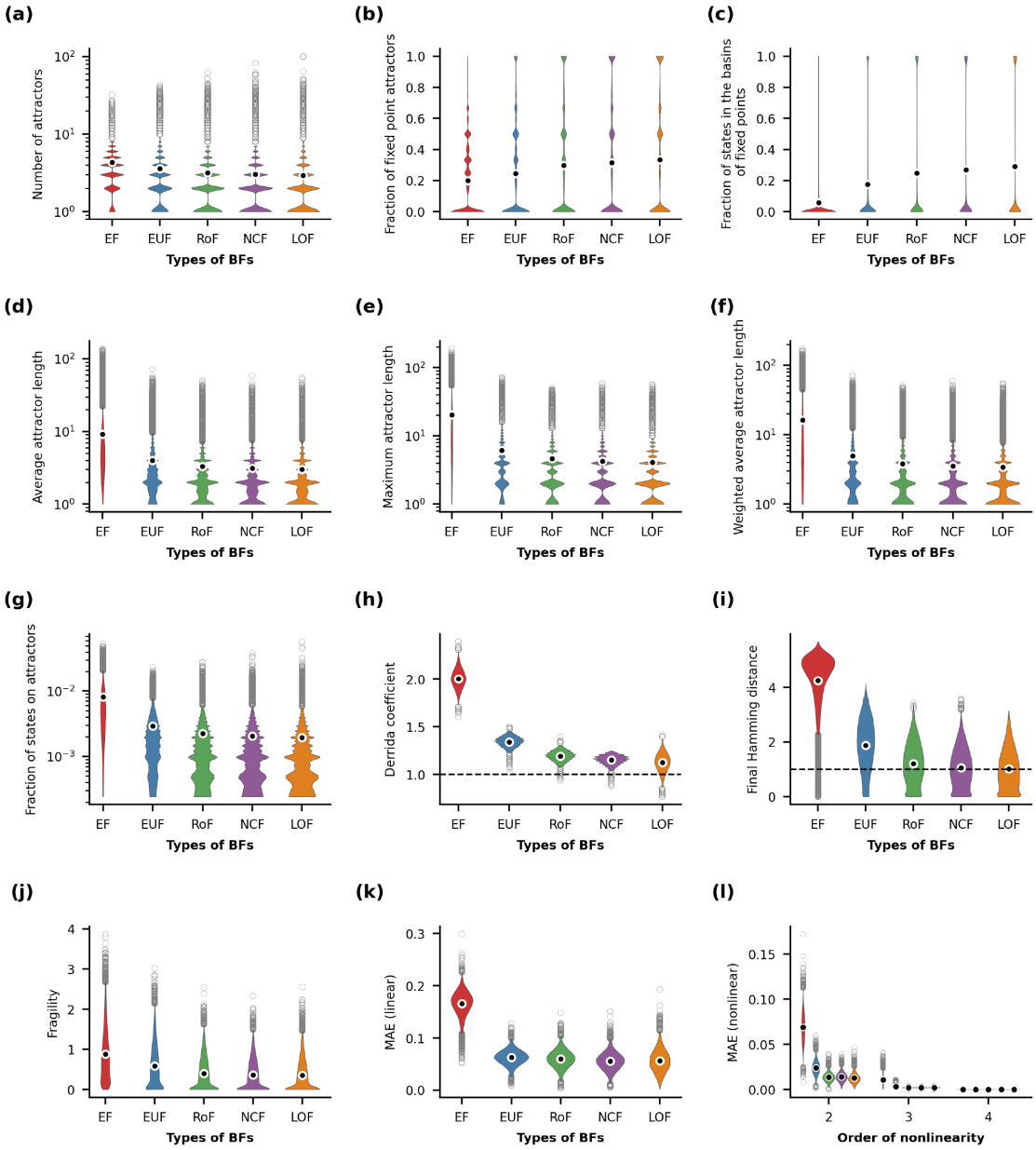


Figure 4.3: Distributions of stability observables when using different classes of BFs. Within each sub-figure, the violins represent the distributions of the corresponding observable according to the EF, EUF, RoF, NCF and LOF class used for the BFs. For each of the 1000 different networks with R-R topology ($k_{in} = k_{out} = 4$, $N = 12$), 1000 models were generated for the observables in (a)-(g) and 10 models were generated for the observables in (h)-(l). The distributions of the observables were then plotted. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

satisfies the stability hierarchy EF < EUF < RoF < NCF < LOF.

Figures A.9-A.49 in Appendix A illustrate the distributions of various observables associated to the attractors, the JS distances between different distribution pairs, and p-values corresponding to the one-sided K-S test. From these plots, it is evident that for all seven observables associated to attractors, the monotonic trend holds consistently across all values of N, k and topologies considered. In each case, for every pair of distributions, we have computed the JS distances and conducted the one-sided K-S test. Notably, the extremely low p-values from the K-S test in each case strongly support the stability hierarchy. When considering the JS distances between the distributions for any of the seven observables associated to attractors, we obtain smaller values for any pair among the RoF, NCF and LOF classes compared to those involving EF or EUF classes. Further, both the linear and nonlinear MAEs exhibit a consistent hierarchy across varying values of N, k and types of topology (see Appendix A, Figures A.50-A.52). For the EF class, we obtain consistently the highest MAE, indicating that this class of logic rules is the least approximable. In contrast, the average of both linear or nonlinear MAEs for other BF classes is found to be lower compared to the EF class across varying N, k and topology. Still, we find a slight difference in the obtained hierarchy for linear and nonlinear MAEs. Specifically, for linear MAE the obtained hierarchy is EF > EUF > RoF > LOF > NCF, while for quadratic and higher-order MAE the obtained hierarchy is EF > EUF > RoF > NCF > LOF, irrespective of N, k and topology (see Appendix A, Figures A.50-A.52).

Previous studies have demonstrated that the extent of *canalyzation* of logic rules or BFs can influence the the stability of dynamics [41–43, 51, 129]. Specifically, canalyzation can steer the dynamics of the system towards a more ordered regime. In this chapter, we have considered five classes of functions: EF, EUF, RoF, NCF and LOF. The relationship among these classes is as follows: EUF is a subclass of EF, RoF is a subclass of EUF, and both NCF and LOF are subclasses of RoF. Notably, LOF is a subclass of NCF only for $k = 2$ and 3. For other values of k, NCF and LOF are only overlapping. For each of the five considered classes, there is a non-null intersection with the class of canalyzing functions (CFs) [5]. However, apart from NCFs, no other class is a proper subset of CF. Our results reveal that imposition of the unateness condition itself on BFs significantly drives the dynamics of BNs away from the chaotic regime. Furthermore, applying the rules from the RoF, NCF and LOF classes leads to much greater stability in the dynamics

of RBNs. Although, the values of the stability measures for RoF, NCF and LOF classes are quite similar in most cases, the one-sided K-S test reveals that there is a consistent trend, namely we find the stability ordering RoF < NCF < LOF. It is important to note that both NCF and LOF are subclasses of RoF, but NCF is a special subclass of CF too. In contrast, there are two types of LOFs (AND-pairs and OR-pairs) that do not always belong to the CF class. Specifically, for $k = 4$ and 5, 25% and 36.36% respectively of the total LOFs are not CFs, rather they are *collectively canalyzing* functions [135]. A k-input BF is *collectively canalyzing* if the output of the function becomes fully determined by fixing a specific subset of i inputs ($1 < i < k$); however, the output remains undetermined when fewer than i inputs are fixed. Furthermore, while we were exploring the *approximability* measure proposed in [76] across different classes of BFs, an independent study [136] appeared which also showed that increased canalyzation in RBNs can explain their increased approximability. On similar lines, we show here that NCFs indeed result in minimum linear MAE, i.e., maximum approximability. However, for quadratic MAE, we find that the minimum value for approximability is achieved by LOFs.

4.4 The dependence of stability on network average degree is specific to the different classes of functions

From the earliest works on damage spreading [5, 121], it was apparent that a network's average degree has a strong influence on the system's dynamics. Specifically, the higher the average degree, the greater the damage spreading in the network. In Figure 4.4, each sub-figure displays the mean values of an observable when using the different BF classes, as a function of k (R-R topology with $N = 12$). For random choice of functions ($\sim$ EF class), in almost all cases the stability decreases as k increases. These trends are as expected: if one increases the number of inputs to a generic BF, there are more possibilities of changing the input, and thus, of spreading any damage. Consequently, the dynamical behavior of a network is pushed towards the chaotic regime as k increases. More importantly, for all k values in Figure 4.4, the stability hierarchy EF < EUF < (RoF or NCF or LOF) is maintained. At larger values of N, the same pattern continues to hold (see Figures A.53 and A.54, Appendix A). Consider for instance the weighted average attractor length (L_W) observable that was found to be well correlated with h^∞. Figure 4.5 shows that the stability

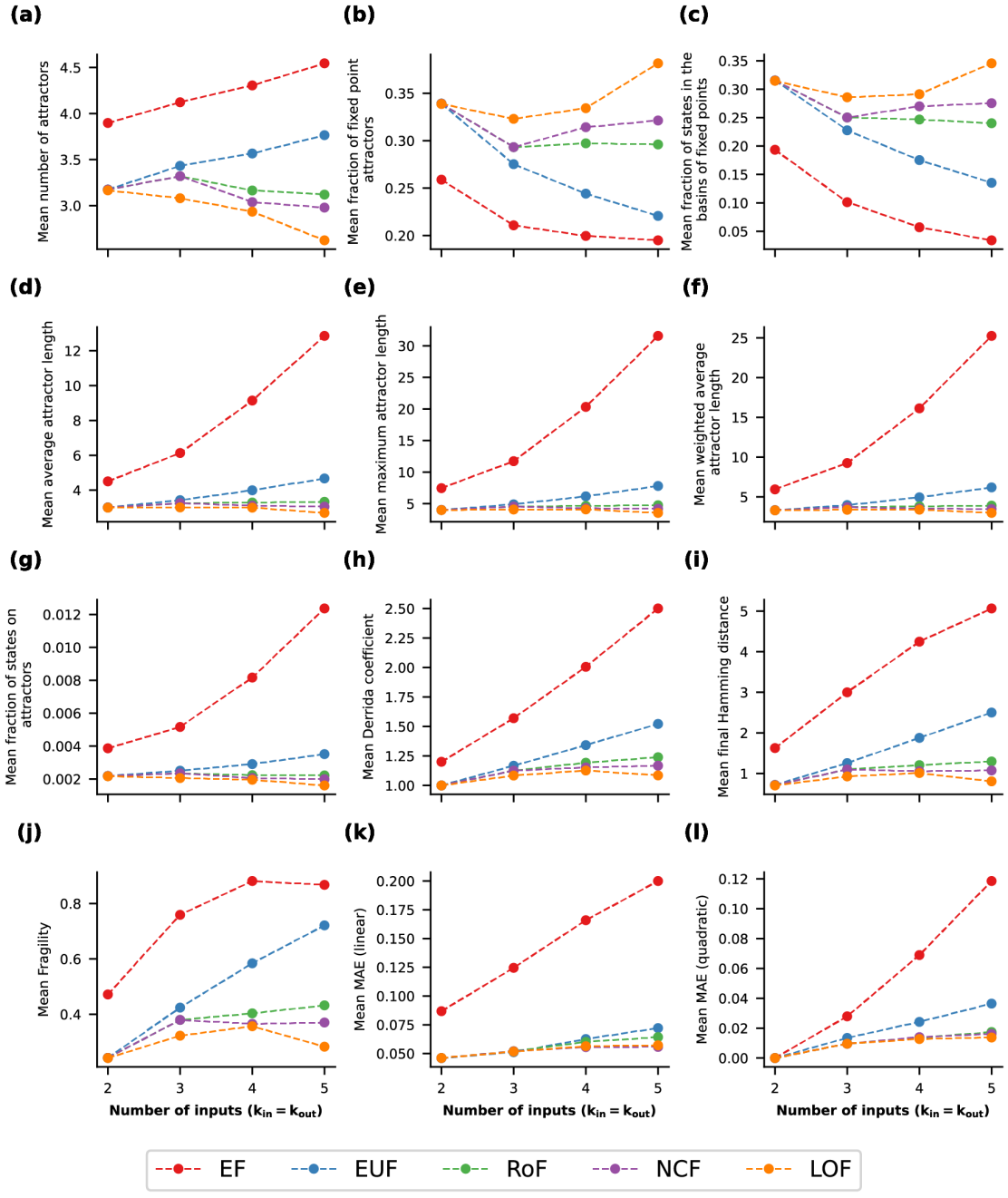


Figure 4.4: Mean values of 12 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 12$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for the R-R network topology.

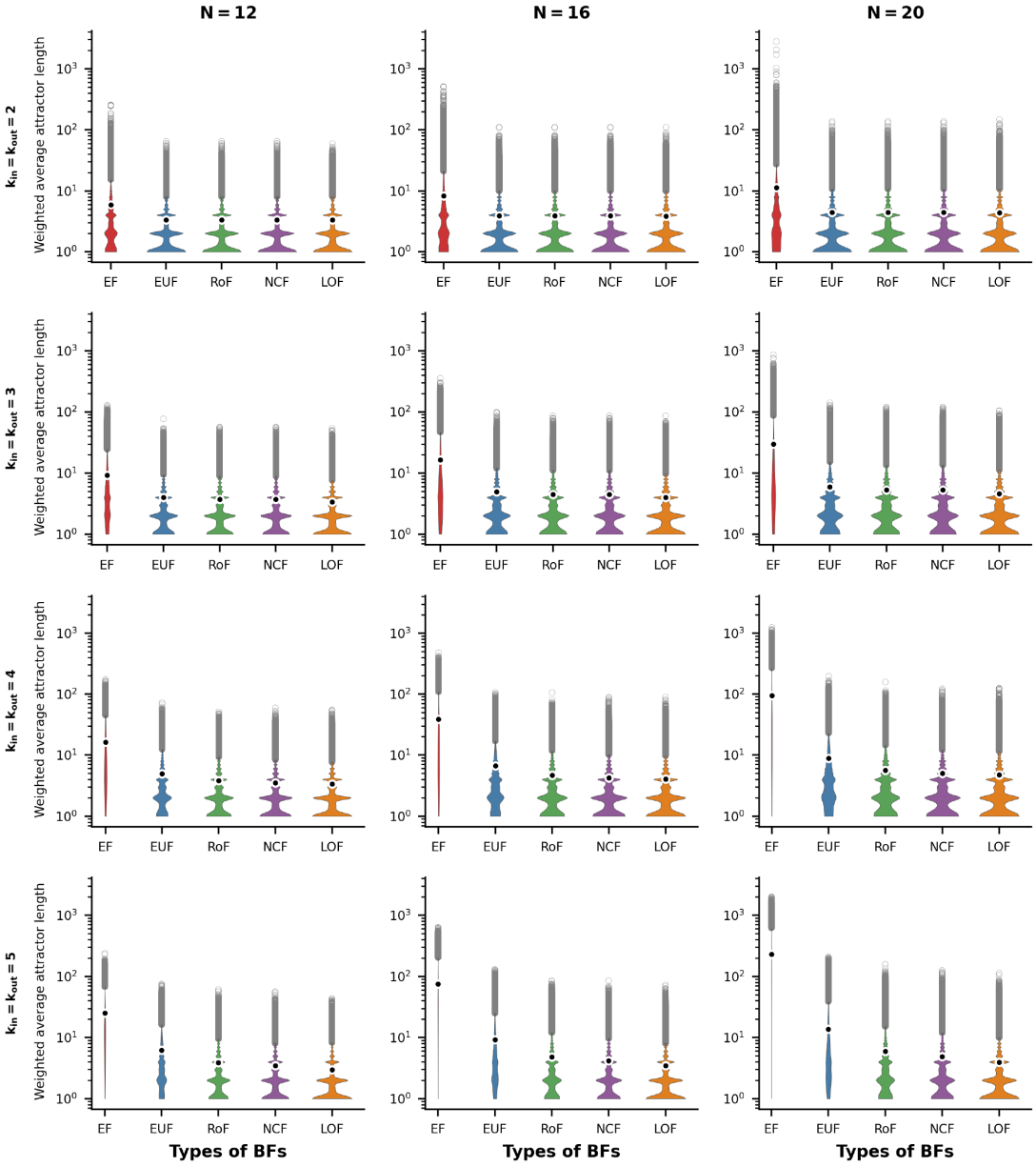


Figure 4.5: Distributions of L_W (the weighted average attractor length) for R-R networks. The violin plots display the distribution of the weighted average attractor length for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the weighted average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

hierarchy EF < EUF < (RoF or NCF or LOF) is consistently observed across different values of N and k investigated here. The sub-figures depicting the mean values of L_W (see Figure 4.4(f) and Appendix A, Figures A.53(f) and A.54(f)) indicate a general trend of increasing values of L_W with increasing k (R-R topology) for the EF class across different values of N (12, 16 and 20). However, for the classes EUF, RoF, NCF and LOF, the variation in the mean L_W with increasing k is less pronounced. Now, higher L_W is associated with more chaotic dynamics. Thus, increasing k pushes the dynamics more towards the chaotic regime in case of EF class (random rules). But, for logic rules in the classes such as RoF, NCF or LOF, increasing k has only a small effect on the dynamics. This difference also arises for other observables like average attractor length, maximum attractor length, or fraction of states on attractors. Indeed, for each of these observables, there is a significant change in the mean value with increasing k for the EF ensemble but not for the other ensembles (EUF, RoF, NCF and LOF).

For the observables linear MAE and quadratic MAE, there is a visible trend of increasing mean value with increasing k for every ensemble (see Figure 4.4 and Appendix A, Figures A.53 and A.54). However, the rate of increase is much higher in the EF ensemble than in the other ensembles. We observe a more interesting trend for the observables number of attractors, fraction of fixed point attractors and fraction of states in the basins of fixed point attractors (see Figure 4.4 and Appendix A, Figures A.53 and A.54). For these observables, the EF and EUF classes show an expected trend: the mean values of the number of attractors increases with increasing k while the mean values of the fraction of fixed point attractors and the fraction of states in the basins of fixed point attractors decreases with increasing k. However, one sees the opposite trend in the mean values of theses observables with increasing k for the LOF and NCF classes. Even though for the LOF and NCF classes the trend is not completely monotonic, it is evident that these classes have a tendency to push the dynamics more towards the ordered regime for higher k (at least in terms of these three observables). Figures A.55-A.60 in Appendix A indicate that the above-mentioned observations hold true in the other network topologies we have considered.

From these results, it is evident that the stability of RBNs is significantly influenced by the choice of logic rules as one increases the average degree or connectivity of the network. When random rules are employed, increasing k drives the dynamics further towards the chaotic regime. However, by restricting to biologically meaningful classes of BFs namely, EUF, RoF, NCF and

LOF, we find that there is no significant change in the values of the stability measures with increasing k. In other words, our results reinforce the beneficial dynamical consequences of restricting to biologically meaningful classes of BFs like RoFs, NCFs and LOFs when reconstructing and modeling GRNs corresponding to biological processes using the Boolean framework.

4.5 The stability hierarchy is robustly maintained across network size

Similarly to the investigation reported in Figure 4.4, the mean values of stability observables can be analyzed as a function of network size N. Figure 4.6 displays the mean values of the stability observables for different BF classes as a function of N (R-R topology with $k_{in}(= k_{out}) = 4$). The nine sub-figures in Figure 4.6 display nine different observables; however h^{∞} and ϕ are omitted from this analysis due the associated computational demand for larger N. Additionally, the observable δ is excluded as it is solely dependent on the BFs and is independent of network sizes. For the four observables, namely number of attractors, fraction of fixed point attractors, fraction of states in the basins of fixed point attractors and fraction of states on attractors, the stability decreases with increasing N regardless of the classes of BFs. In particular, the mean values of the number of attractors increases while the mean values of the fraction of fixed point attractors, the fraction of states in the basins of fixed point attractors and the fraction of states on attractors decreases as N increases. Interestingly, we observe a distinct trend for the observables linked to the attractor lengths. For the EF class, the mean values of the three observables namely, average attractor length, maximum attractor length and weighted average attractor length, exhibit an increasing trend as N increases. However, we do not observe any significant change in the mean values of these observables for the BF classes EUF, RoF, NCF and LOF. On the other hand, we observe that the mean values of MAE (linear or quadratic) remain almost invariant with increasing N irrespective of the choice of BF class. Note that, for any given value of N, the stability hierarchy EF < EUF < (RoF or NCF or LOF) is consistently maintained as discussed in Section 4.3. Moreover, for R-R network topology, the above-mentioned observations hold true for other values of k as well (see Appendix A, Figures A.61-A.63). Lastly, the above-mentioned observations also hold true even if one analyzes other network topologies (see Appendix A, Figures A.64-A.71).

Previous research on RBNs has investigated how observables such as the number of attractors or the typical length of attractors are influenced by the increase in BN size [19,94–96]. To the best of our knowledge, this is the first systematic study on five classes of BFs to show that the stability hierarchy among different classes of functions (EF < EUF < (RoF or NCF or LOF)) is consistent across different BN sizes. However, with growing BN size, there is a significant influence of the choice of logic rules on the stability of BNs. Furthermore, the observed effect on the stability can vary depending on which observable one considers. For example, with increasing BN size, the observable "number of attractors" shows a very similar trend across different classes of BFs. In contrast, the observables associated with the length of attractors show a significantly different trend for the biologically meaningful functions (EUF, RoF, NCF and LOF) in comparison to the case of random functions (EF). With EF rules, the lengths of the attractors (whether average, weighted average, or maximum) grow very fast with increasing BN size and push the dynamics towards the chaotic regime. In contrast, the use of biologically meaningful functions prevents the typical cycle lengths from going very high. Therefore, these results provide additional support for restricting update rules to the biologically meaningful classes RoFs, NCFs and LOFs when reconstructing and modeling GRNs corresponding to biological processes in the Boolean framework.

4.6 Discussion

Our aim in this chapter was to comprehensively and holistically analyze the effects of both network structure and logical rules on the dynamics of RBNs. We considered both aspects using many different stability measures ranging from damage spreading (see Chapter 3, Section 3.1) to various attractor properties (see see Chapter 3, Section 3.2). We have also utilized the recently introduced measure of approximability (MAE) as described in Section 3.3 in Chapter 3.

We find that all stability measures follow the expected hierarchy: EF < EUF < (RoF or NCF or LOF), in the order of least to most stable. This hierarchy is particularly robust, holding irrespective of network size (N), connectivity (k), and even the topological ensemble considered. Unexpectedly, the Derrida Coefficient (δ), which is commonly considered as a proxy for the long-term damage spreading measure (h^{∞}), is found to be quite weakly correlated with h^{∞} and thus may not be a satisfactory substitute to infer long-term damage spreading dynamics. In contrast, we find that the weighted average attractor length (L_W) is quite highly correlated with

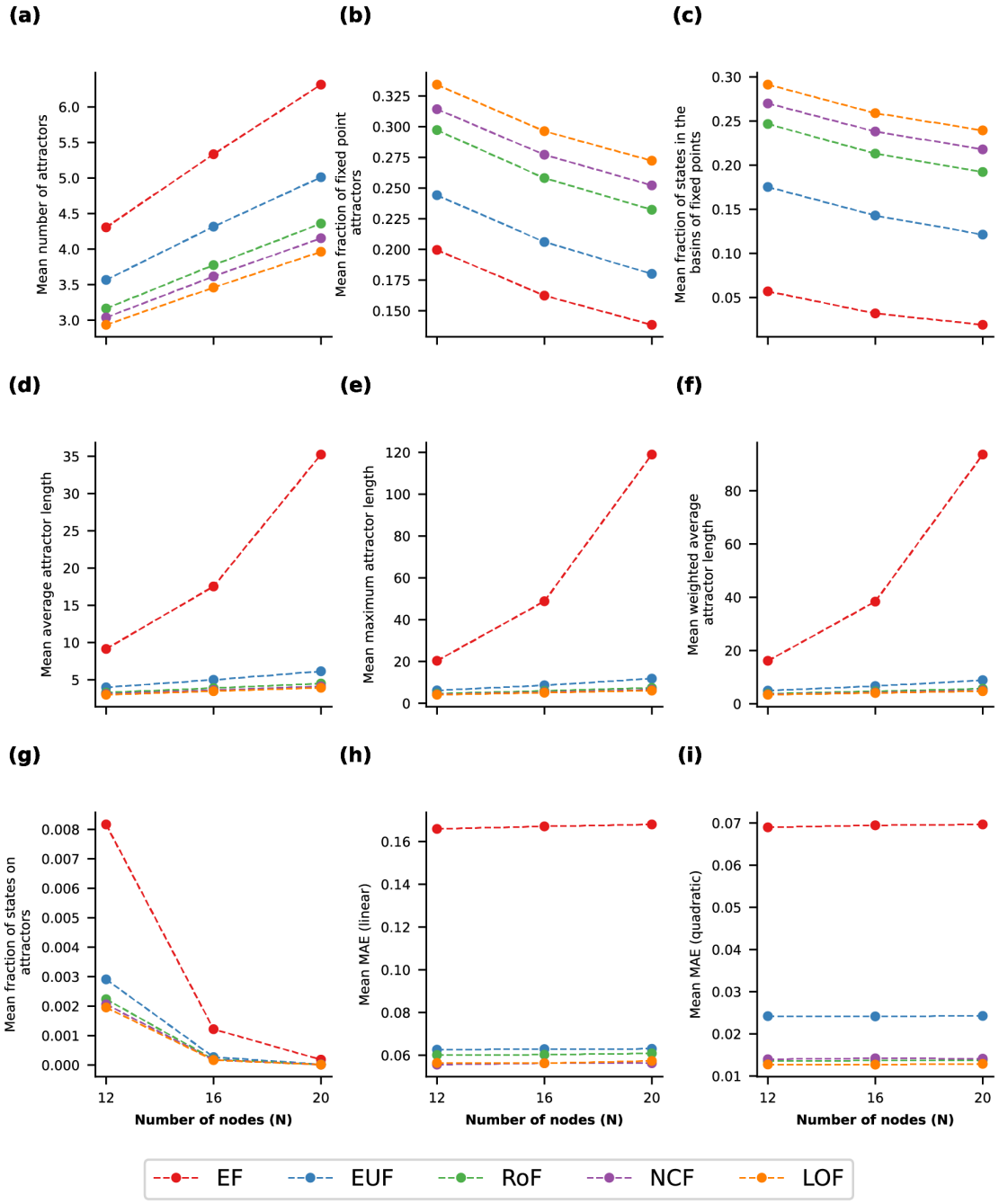


Figure 4.6: Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 4$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-R network topology with $k_{in} = k_{out} = 4$

h^{∞}, so it can be used when necessary as a substitute for h^{∞}.

Next we focused on the average connectivity (k) of the network. As expected, the system typically gets pushed towards chaotic dynamics as one increases k, however the effects are quite variable across the classes of BFs. Specifically, the random class of functions (EFs) seem to be far more affected by the change in k compared to the other classes EUF, RoF, NCF and LOF which have been shown to be more biologically meaningful [51]. For some attractor properties (namely the fraction of fixed point attractors and the fraction of states in the basins of fixed point attractors), we find that the overall trend of decrease in stability is not shared by the classes LOF and NCF at high k. Nevertheless, the patterns observed are robust, being insensitive to network topology and network sizes.

Lastly, we examined the effects on the dynamics of varying network size (N) when considering the network attractor properties and the MAE. There is significant variation depending upon the observable considered. For example, the average number of attractors shows an increasing trend upon increasing the size of the network across all BF classes. In contrast, the observable "average length of attractors" only shows an increasing trend with network size for the EF rules, whilst not showing significant difference for the other classes of BFs. Such class-specific behaviour is similar to that observed when examining the repercussions of varying the average connectivity of the network.

An important limitation of the study presented in this chapter is its restriction to rather modest values of N; indeed many of the observables we considered become computationally intractable at large N. Thus, some trends such as the dependence on k seem out of reach if one wants to determine the large N limit; indeed, large k values require large N values if finite size effects are to be small. However even at $N = 20$ we have been unable to obtain long-term damage spreading measures such as h^{∞} and ϕ due to constraints on our computational resources. Hence an immediate improvement would be to find a way to provide reliable measurements or better proxies for larger sized networks and to verify if the trends remain the same. One could also try to explore the effects of performing our analyses under asynchronous [29] dynamics and see if one can still reach similar conclusions.

Chapter 5

Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph

This chapter is based on the published manuscript [79].

Chapter overview. This chapter investigates how biologically meaningful classes of Boolean functions (BFs) shape the global dynamics of reconstructed Boolean gene regulatory networks (GRNs) through the structural properties of their state transition graphs (STGs). It first describes the generation of biologically plausible ensembles of Boolean models from ten published regulatory networks. The chapter then examines how different classes of regulatory logic influence the bushiness and convergence of STGs. Next, it introduces descriptors of local dynamics, including new adaptations of Wuensche's Z-parameter for Boolean networks (BNs) and network sensitivity, and examines their distributions across BF classes. Finally, the chapter assesses these local descriptors as computationally efficient proxies for bushiness before concluding with a discussion of the biological implications of the findings.

Boolean networks (BNs) have been extensively used to model the dynamics of gene regula-

tory networks (GRNs) to explain a wide range of biological processes including differentiation, metabolism, apoptosis and proliferation, among many others [7, 32–34, 36–38, 107, 110]. The full dynamics of a BN can be described by a state transition graph (STG) in which a fixed point attractor typically corresponds to a biological steady state expression pattern specific to a cell type, while its basin of attraction corresponds to all state expression patterns that will converge to that steady state under the BN dynamics [19, 137, 138]. Such dynamics must exhibit robustness to perturbations if cells are to maintain homeostasis [33], making it imperative to investigate features of the STG associated with robustness and how regulatory logic rules can impact those features.

In the framework of elementary cellular automata (CA), Stephen Wolfram systematically classified all CA rules into four classes based on the dynamical behavior they exhibited [139]. Prior to our work, an analogous framework that categorized global dynamical properties of BNs based on types of Boolean function (BF) had not been proposed, though there had been extensive studies on how network structure and BFs can affect the dynamics of BNs [140–144], dynamical regimes in random BNs [7, 121, 145–147] and in reconstructed BNs [51, 53]. In this chapter, we therefore ask how different classes of regulatory logical rules in BNs can affect systemic dynamics. For that, we provide mathematical measures that build on previous work studying STGs in CA, especially by Andrew Wuensche [61–63].

The primary objectives of the study presented in this chapter are to delineate how different classes of logic rules, namely effective functions (EFs), effective and unate functions (EUFs), read-once functions (RoFs) and nested canalyzing functions (NCFs), shape the structural features of the STG of a Boolean GRN, and search whether there exist good proxies for those features. To do so, we first select ten published reconstructed Boolean GRNs and for each one generate four ensembles of models, having the same network structure and the same set of biological fixed points but differing in the type of logic rule employed. We consider different *global* measures to characterize two features of the structure of STGs: (a) bushiness and (b) rate of convergence. We determine how these measures vary across the four ensembles for each of the ten GRNs, particularly how distinct the effects of using biologically meaningful BFs (EUFs, RoFs and NCFs) are compared to using EFs (which are random BFs for all practical purposes). Next we explore the correlation between bushiness and convergence. We then address the challenge of adapting the Z-parameter, originally proposed by Wuensche in the context of CA to characterize the identifiability

of state pre-images under a given update rule [78], to BFs in BNs. To this end, we propose a scheme based on permutations of the inputs to a BF. Subsequently we posit four variants of the Z-parameter at the level of the BF and their respective counterparts at the level of the network which we collectively refer to as the network Z-parameters. One last quantity we consider is the so-called network sensitivity or average sensitivity of a network [53, 58]. We then study the distributions of these *local* descriptors of dynamics, that is, the four network Z-parameters and network sensitivity, in our ensembles and evaluate correlations between those descriptors. Finally, we explore whether these local descriptors can serve as 'proxies' for the bushiness and convergence. It is noteworthy to mention that while the descriptor 'network sensitivity' has been linked to robustness of BNs, its connection to the structural features of the STG has not been investigated prior to this study as per our knowledge.

5.1 Generating ensembles of biologically plausible models of ten published Boolean GRNs for living systems

This section describes our procedure to generate ensembles of biologically plausible Boolean models that we utilized for the subsequent analyses presented in this chapter. We first choose ten reconstructed Boolean GRNs that are published in the literature. Their abbreviations and references are provided in Table 5.1. In Appendix B, we provide the models in BoolNet format in Tables B.1-B.10, their biological fixed points in Tables B.11-B.20, and the types of BFs used in Tables B.21-B.30. For each GRN, we generate four biologically plausible ensembles of models, all of which share the same network structure of the GRN and satisfy the same biological fixed points [64, 148], but differ only in the types of BFs used, namely, EFs, EUFs, RoFs and NCFs (see Chapter 2, Section 2.5 for description of the types of BFs).

Our methodology to generate such ensembles using these types of BFs is detailed in Section 2.7 in Chapter 2. [64, 148]. We shall refer to these ensembles (sampled or exhaustively enumerated) of models as EF-ensemble, EUF-ensemble, RoF-ensemble and NCF-ensemble (see Appendix B, Tables B.31-B.40). Details of the ten reconstructed GRNs of living systems are provided in Section B.1 in Appendix B. As it was computationally infeasible to generate all the allowed UFs for nodes with more than six inputs, we proposed an algorithm [79] to sample UFs (and EUFs).

Table 5.1: Ten published reconstructed Boolean GRN models. 'Names of models' column provides the names of each of the Boolean GRN network structures that are used in this work. The 'Abbreviation' column provides the abbreviated form we designate for each of the ten models. The 'Nodes' and 'Edges' columns provide the number of nodes and edges respectively for each of the ten models.

S. No.	Names of models	Abbreviation	Nodes	Edges	Publication year	PMID/ Others	Reference
1	Root Stem Cell Niche GRN	RSCN-GRN	9	19	2010	20920363	[149]
2	Epithelial-Mesenchymal Transition GRN	EMT-GRN	12	40	2023	36968076	[150]
3	NeoCortex Developmental GRN	NCD-GRN	10	19	2010	20862356	[151]
4	Early Heart Development GRN	EHD-GRN	15	39	2012	23056457	[152]
5	Myeloid Differentiation GRN	MD-GRN	11	30	2011	21853041	[37]
6	T-Helper Cell differentiation GRN	THC-GRN	23	34	2006	16542429	[110]
7	EGFR signalling pathway	EGFR-GRN	13	21	2019	30582550	[153]
8	Epithelial-Mesenchymal Transition (with Senescence) GRN	EMT-Senescence-GRN	9	30	2023	Conference Proceeding	[154]
9	Flower Organ Specification GRN	FOS-GRN	13	42	2010	20303988	[155]
10	Gonadal Sex Determination GRN	GSD-GRN	19	75	2015	26573569	[156]

Our algorithm is based on an iterative color propagation scheme on the vertices of the Boolean hypercube (see Chapter 2, Section 2.7).

After applying the above procedure, for each of the ten published models, we determined the number of allowed functions at each node. These numbers are provided in Tables B.31-B.40 in Appendix B. Multiplying the number of allowed BFs over all nodes provides the total number of Boolean models that satisfy the expected constraints. Tables B.31-B.40 in Appendix B show that in spite of the multiple constraints applied, the space of biologically plausible models is still extremely large. Hence we sampled 10^6 models from those ensembles in which the number of models exceeded 10^6 (except for FOS-GRN and GSD-GRN where we sampled 10^5 models), and used the exhaustive set of models otherwise. We generated the STGs for all 21032377 models across all ten GRNs and all ensembles using the R package BoolNet [123], and moreover, computed the global measures described in Section 3.4.1 in Chapter 3: (a) G-density, (b) average in-degree of the *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$), and (c) average convergence rate of the trajectories originating at *GoE* states or (λ_{GoE}). Note the use of large ensembles of models ensures the statistical reliability of our results.

In the ensuing sections of this chapter we study the robustness of different BNs via their degree of contraction and convergence (see Chapter 3, Section 3.4.1) for models in ensembles that are obtained from ten reconstructed GRNs, and identify which properties of the underlying rules drive high bushiness (contraction) and convergence. Furthermore, we provide a connection between the values of the above-mentioned 'global' measures of robustness using the STG and those of descriptors of 'local' dynamics. Also, in what follows and unless stated otherwise, we use the term 'distribution' of a quantity (global or local measures) to mean the distribution of that quantity in any one of the four ensembles.

5.2 Biologically meaningful functions lead to highly bushy and convergent STGs

In this chapter, we concentrate on characterizing the overall global dynamics based on two features of the STG: the rate of contraction – hereafter referred to as 'bushiness' – and the rate of 'convergence'. The term 'bushiness' serves as a visual metaphor associated with the average number of

edges feeding into nodes of the STG. To illustrate that, consider two Boolean models (model A and model B) derived from a myeloid differentiation GRN (MD-GRN) [37] that differ in the type of regulatory logic rule used but recover the same biological attractors (see Figure 5.1(a)). Figures 5.1(b) and 5.1(c) display the complete STGs for model A and model B respectively.

Clearly, the STG of model A appears highly 'bushy' compared to that of model B. We quantify such bushiness using two measures. The first is given by the reduction of the space of states when applying the update dynamics. In continuous-time dynamical systems, the analog of this reduction is the contraction rate specifying how phase space volume diminishes in time; in reversible dynamics such as in Hamiltonian systems, phase space volume is conserved, while in dissipative systems phase space gets contracted. In our discrete-time case, we quantify this reduction by comparing the number of states produced from one update of the BN to the initial number of states. It is easy to see that the reduction factor is just the fraction of states that have in-degree equal to 0 in the STG, i.e., the G-density (see Chapter 3, Section 3.4.1). The second is based on the in-degree of nodes, taking the average over all nodes that have incoming edges. Both of these measures quantify the *irreversibility* of the BN dynamics and they are closely related (see Chapter 3, Section 3.4.1).

On the other hand, we quantify *convergence* based on the length of trajectories (or, transient lengths) originating at *GoE* states because these give the maximal trajectory lengths (see Chapter 3, Section 3.4.1). The three quantities: (a) G-density, (b) average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) and (c) average convergence rate of the trajectories originating at *GoE* states (λ_{GoE}) will be referred to as *global* measures throughout this chapter (see Chapter 3, Section 3.4.1 for the definitions and Figure 3.1 for visual illustration). In CA, Wuensche observed that highly bushy STGs with shorter transients have more ordered dynamics while less bushy STGs with longer transients have more chaotic dynamics [157]. Prior to our work, studies of such structural features of STGs have been primarily restricted to elementary CA and random BNs [61, 91, 158, 159], with no such investigation in reconstructed Boolean GRNs. In our work, we address this challenge and in doing so we exhibit the strong association between bushiness, rate of convergence and dynamical *robustness* which is an essential feature of biological systems. This last characteristic is exemplified via a Derrida plot (see Figure 5.2) [7, 122] showing the system's response to a perturbation after one time step for the two models A and B. The behaviour at long time can also

be considered (see Figure 5.1(f)) [60]. Thus, investigating the structural characteristics of STG will enable selection of suitable models [64] among multiple possibilities, helping us identify models that exhibit greater robustness.

In the pedagogical example of MD-GRN (see Figure 5.1), we observe the implication of choice of models on the global dynamics. Next, we would like to unveil the fact that, the functions in model A were chosen from a biologically meaningful class (specifically, NCF), whereas the functions in model B were random EFs. While both of the models yield the same biological attractors, model A demonstrates more bushy and convergent STG (a signature of robust or ordered dynamics) compared to model B. To explore the implications of various function classes on the global dynamics more broadly, we examine the distributions of the global measures across different ensembles (see Section 5.1) for the ten published Boolean GRNs (see Table 5.1). We first obtain the distribution of the G-density as shown by the box plots in Figure 5.3(a) and Figure B.1 in Appendix B. We observe two striking and consistent patterns across all ten GRNs. First, STGs in ensembles employing biologically meaningful BFs (EUF-ensemble, RoF-ensemble and NCF-ensemble) are more bushy than STGs in the EF-ensemble, an indication of more robust dynamics. Second, the STGs in the NCF-ensemble are typically more bushy than those in the EUF-ensemble or RoF-ensemble. Given this trend for robustness, if one performs model selection during Boolean GRN reconstruction, it is appropriate to adhere to biologically meaningful classes [64]. Moreover, to enforce the robustness constraint systematically, one can focus on the models with top G-density values within those classes, significantly narrowing down the search space. Similar results follow for the distribution of the average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) since it is a non-linear monotone increasing function of G-density (see Chapter 3, Eq. (3.6); Appendix B, Figures B.2 and B.3).

Next, we consider the distribution of the *average convergence rate* of trajectories originating at *GoE* states (λ_{GoE}). From Figure 5.3(b) and Figure B.4 in Appendix B, we observe that on average, the models in the EF-ensemble have lower convergence (smaller values of λ_{GoE}) in comparison to ensembles that employ biologically meaningful BFs. We also find that the ensembles constrained with biologically meaningful BFs have similar values of λ_{GoE}. These trends are consistent across the ten GRNs as can be seen in Figure 5.3(b) and Figure B.4 in Appendix B.

Lastly, the Spearman correlation coefficients between G-density and λ_{GoE} for ensembles

across the 10 GRNs (see Appendix B, Figures B.5 and B.6) reveal a positive correlation but that is surprisingly small, indicating that the two measures capture different structural characteristics of the STG.

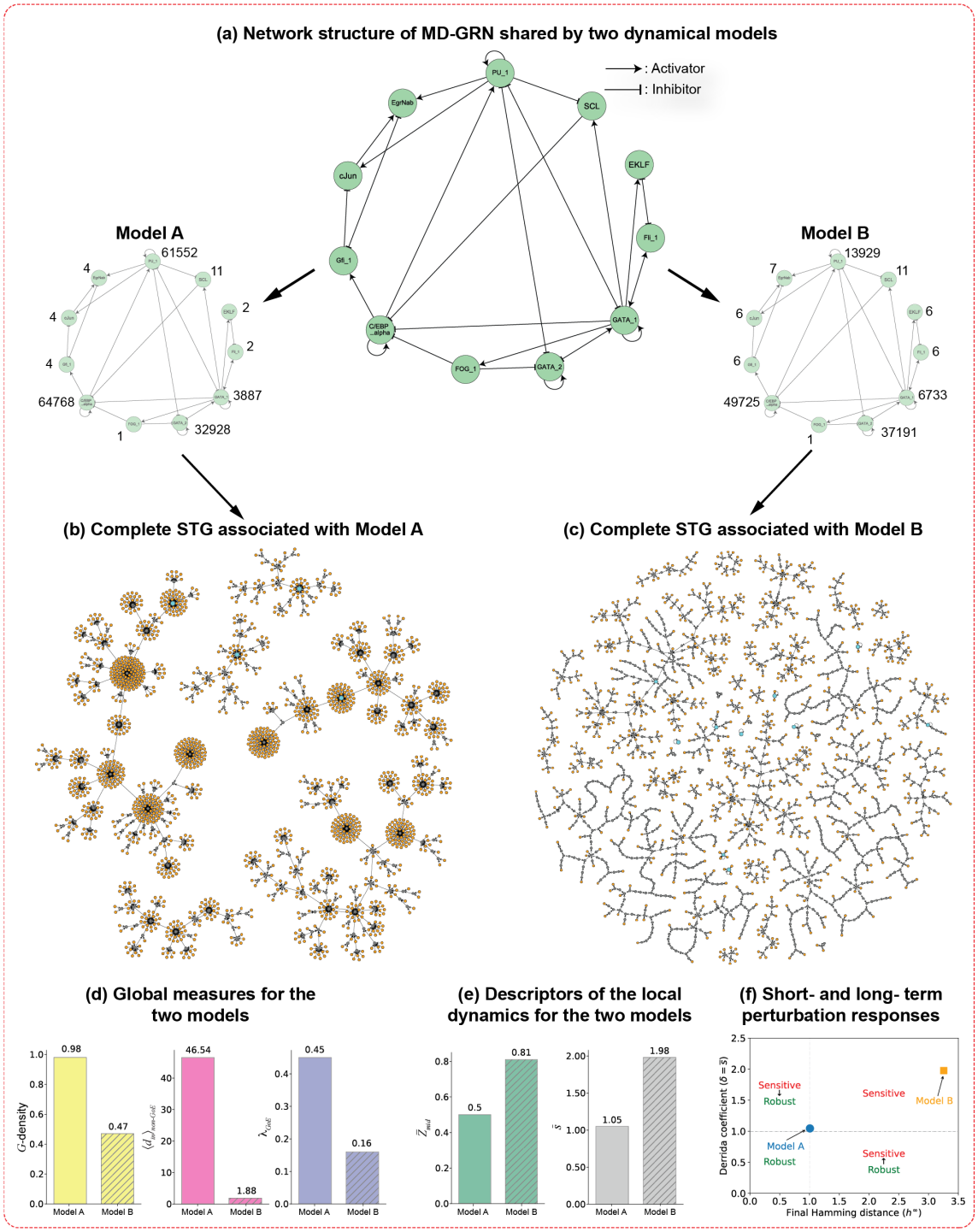


Figure 5.1: STGs for two different choices of Boolean rules for the biological network MD-GRN with the values of their global measures of bushiness and convergence, descriptors of local dynamics and short-and long-term perturbation responses. **(a)** The network in the centre of the top panel is the Myeloid differentiation GRN (MD-GRN). On its two sides are Boolean models (Model A and Model B) with identical network structure but different regulatory logic rules. For these two models, the genes are arranged in the same manner as depicted in the network structure in the centre. The BFs at each node are encoded as integers as described in Section 2.4 in Chapter 2. In the truth table, the (left to right) order of the inputs for each of the genes in the network is as follows: GATA_2: [GATA_2, GATA_1, FOG_1, PU_1], GATA_1: [GATA_2, GATA_1, Fli_1, PU_1], FOG_1: [GATA_1], EKLF: [GATA_1, Fli_1], Fli1: [GATA_1, EKLF], SCL: [GATA_1, PU_1], C/EBP_alpha: [GATA_1, FOG_1, SCL, C/EBP_alpha], PU_1: [GATA_2, GATA_1, C/EBP_alpha, PU_1], cJun: [Gfi_1, PU_1], EgrNab: [cJun, Gfi_1, PU_1], Gfi_1:[EgrNab, C/EBP_alpha]. **(b)** and **(c)** display the complete STGs containing 2048 states obtained for Model A and Model B respectively. The 'orange', 'blue' and 'grey' nodes indicate the *GoE* states, fixed points and *non-GoE* states that are not fixed points respectively. Visual inspection shows that the STG for Model A appears significantly more bushy than the STG for Model B. **(d)** Barplots representing the bushiness (G-density, $\langle d_{in} \rangle_{non-GoE}$) and convergence ($\lambda_{GoE}$) measures for Model A and Model B. Each measure is shown as a separate bar plot that quantifies the visible difference between (b) and (c). **(e)** Barplots representing the values of the descriptors of local dynamics ($\overline{Z}_{mid}$ and $\overline{s}$) for Model A and Model B. **(f)** Short (δ)- and long (h^{∞})- term perturbation responses for the two models (see Chapter 3, Section 3.1). In the 'Robust' regime (lower left quadrant), both short-term and long-term responses [60] remain below 1 whereas, in the 'Sensitive' regime (upper right quadrant) both the responses are above 1; these regimes correspond to 'ordered' and 'chaotic' dynamics respectively. The two off-diagonal quadrants indicate disagreement in the short- and long-term responses, an unlikely situation, with the small arrows indicating the change when going from short to long times. Model A lies close to the critical boundary but, Model B lies completely in the Sensitive or disordered regime. Note that δ is identical to the network sensitivity $\overline{s}$.

5.3 Descriptors of local dynamics

Unfortunately, obtaining the STG and its features can be computationally expensive. In particular, all three global measures studied in the previous section, require listing the *GoE* states, a task that becomes computationally intractable with increasing network size [160,161]. It is thus appropriate to search for proxies of those three global measures that remain tractable even for large networks. Although this was not done explicitly by Wuensche, he had introduced the Z-parameter (in his studies of CA), a quantity that can be computed directly from the rule table which measures the probability that the next unknown cell in a partial pre-image is uniquely determined (see Chapter 3, Section 3.4.2) [62,78,126]. He showed that the Z-parameter is a useful predictor for the bushiness of the STG [62]. In extending the Z-parameter to BNs, one is confronted with the fact that in contrast to the situation in elementary CA, each node in a BN is allowed to have a different 'rule'

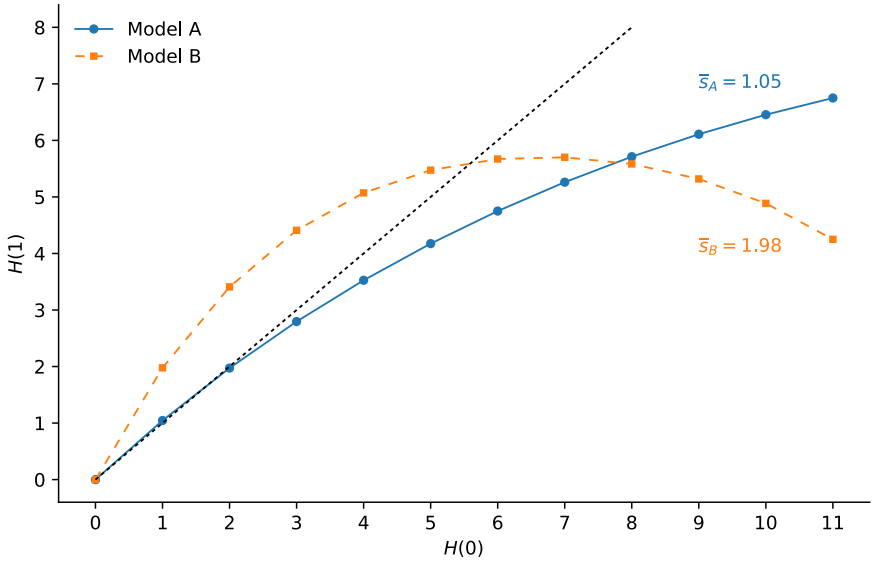


Figure 5.2: Derrida plot showing the progression of various Hamming distances over one time step for Models A and B shown in Figure 5.1. This figure shows the evolution of the Hamming distance between a pair of network states over a single time step, under the synchronous update scheme for the two example models, Models A and B, in Figure 5.1. Models A and B have 11 nodes, therefore the Hamming distance between a pair of network states ranges from 0 to 11. All possible pairs of network states have been considered for each initial Hamming distance $H(0)$. Then the mean of the Hamming distance $H(1)$ of all pairs of states, obtained after evolving over a single time step, is plotted against the initial distance $H(0)$. The blue and orange curves correspond to the Models A and B respectively. The black dashed straight line represents $H(0) = H(1)$. The network sensitivities for the Models A and B are denoted by $\overline{s}_A$ and $\overline{s}_B$ respectively.

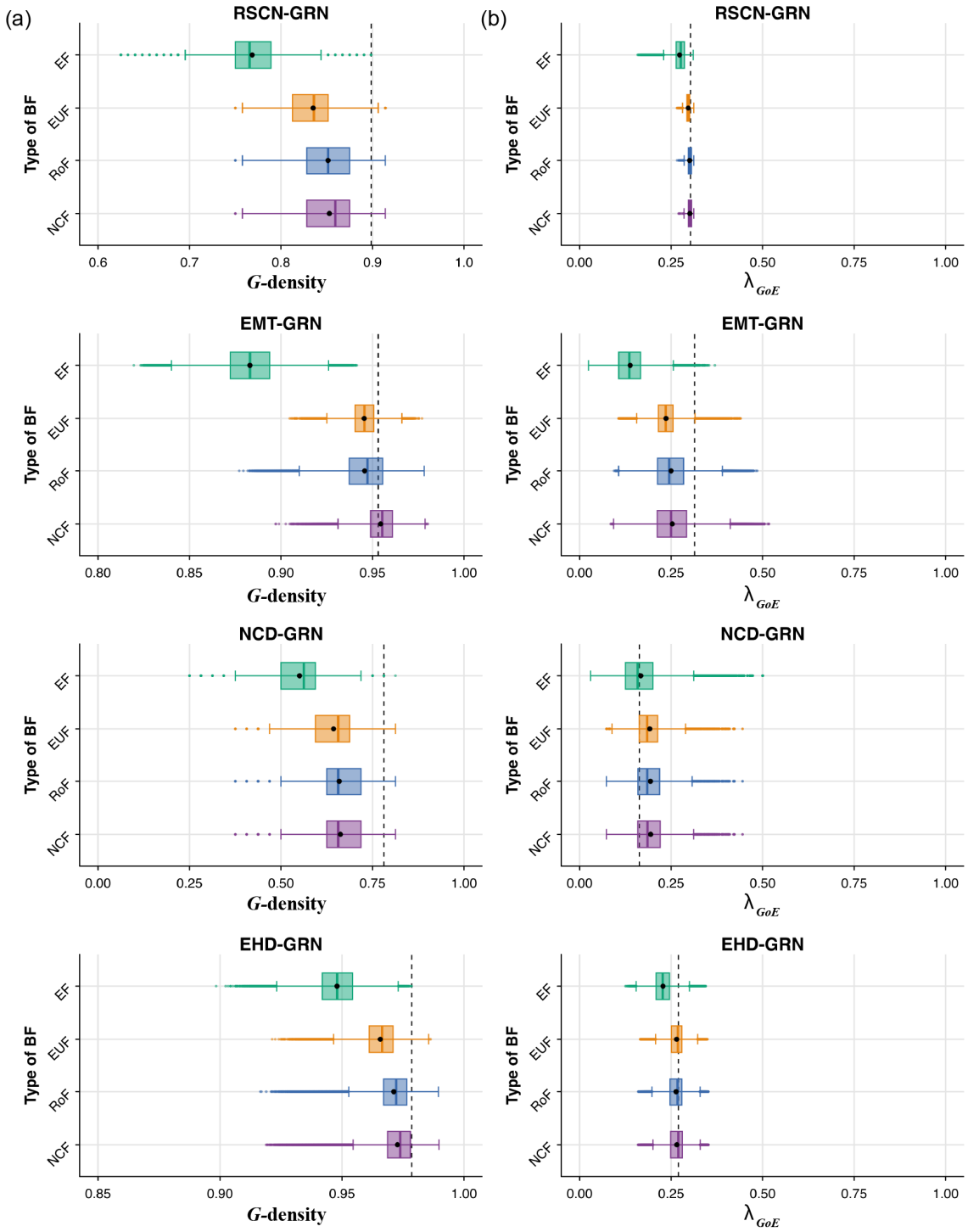


Figure 5.3: Distribution of global measure values for different ensembles generated using network structures from 4 published Boolean GRNs. The box plots in **(a)** and **(b)** display the distributions of G-density and λ_{GoE} respectively in the 4 ensembles that each use a given type of regulatory logic. The mean and median of the distributions are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in (a) and (b) correspond to the G-density and λ_{GoE} respectively for the Boolean model provided by the modelers in the published article.

and there is no natural ordering of its inputs, making it non-trivial to adapt the Z-parameter to BNs. Prior to our work, this challenge of finding rule-based measures in BNs had not been addressed, and we fill this gap.

5.3.1 Adaptation of Z-parameter in cellular automata to Boolean networks

In CA, the Z-parameter as defined by Wuensche reflects the degree to which pre-images of states in the state space are identifiable for a given automaton update rule [78] (see Chapter 3, Section 3.4.2). A low Z value indicates a lower probability to have a unique pre-image, that is, a lower degree of reversibility of the dynamics [62, 63, 78, 91]. A Z-parameter value of 1 corresponds to perfectly reversible dynamics. We may thus expect a lower Z value to be associated with a higher in-degree of states in the STG and thereby a higher G-density.

For any one-dimensional CA rule, Wuensche introduced a computational procedure to define Z_{left} and Z_{right}, associated with reading the spatial arrangement of inputs from left to right and from right to left respectively, and then he defined the Z-parameter as their maximum [62] (see Chapter 3, Section 3.4.2). Since there is no notion of 'left' or 'right' of a node for BFs in BNs, the above definition needs to be extended.

5.3.1.1 2^i-blocks of a Boolean function

To adapt the Z-parameter to BFs, we first introduce a combinatorial structure on the truth table of a BF that underlies Wuensche's construction. Consider a BF f with k inputs $(x_1, x_2, \ldots, x_k)$ and its truth table written in standard lexicographic order. Then, a 2^i-block is defined as a set of 2^i consecutive rows in the truth table that share fixed values of the first $(k - i)$ inputs and span all possible values of the remaining i inputs. To understand it better, suppose in the truth table of f there are pairs of rows (R, R') such that:

$$R = (x_1 = a_1, x_2 = a_2, \ldots, x_{k-1} = a_{k-1}, x_k = 0) \text{ and } R' = (x_1 = a_1, x_2 = a_2, \ldots, x_{k-1} = a_{k-1}, x_k = 1)$$

In the canonical ordering of truth tables, these rows are consecutive and we refer to them as a 2^1-block.

We can similarly define 2^2-blocks by considering the rows

$$(x_1 = a_1, x_2 = a_2, \ldots, x_{k-2} = a_{k-2}, x_{k-1} = 0, x_k = *) \text{ and}$$
$$(x_1 = a_1, x_2 = a_2, \ldots, x_{k-2} = a_{k-2}, x_{k-1} = 1, x_k = *)$$

where $* \in \{0, 1\}$ means that any values are allowed. Such blocks are now formed by 2^2 consecutive rows. More generally, we will call a set of rows a 2^i-block if it corresponds to combining two adjacent 2^{i-1}-blocks corresponding to:

$$(x_1 = a_1, \ldots, x_{k-i} = a_{k-i}, x_{k-i+1} = 0, x_{k-i+2} = *, \ldots, x_k = *) \text{ and}$$
$$(x_1 = a_1, \ldots, x_{k-i} = a_{k-i}, x_{k-i+1} = 1, x_{k-i+2} = *, \ldots, x_k = *)$$

In this way, one can go up to a 2^k-block that in fact corresponds to all rows in the truth table. Note that, these blocks form a hierarchical partition of the truth table: at a given level i, all 2^i-blocks are disjoint and together they cover the entire truth table.

5.3.1.2 Calculation of Z_{left}

For a given ordering of the inputs, we can follow Wuensche's prescription to compute the value of Z_{left}. For a 2^i-block, Wuensche determines whether it satisfies a switch-like property:

$$f(x_1, x_2, \ldots, x_{k-i}, 0, * \ldots, *, *) = T \text{ and } f(x_1, x_2, \ldots, x_{k-i}, 1, *, \ldots, *, *) = \overline{T} \tag{5.1}$$

where $T \in \{0, 1\}$ and $\overline{T}$ is its complement, while the $*$s mean that the two equalities are to be satisfied regardless of the corresponding input values of $x_{k-i+2}, \ldots, x_k$ (For $i = k$, all but the first of the arguments are '*' in Eq. (5.1)). If a 2^i-block satisfies Eq. (5.1), we say it is "complement canalyzing" because, in the context $(x_1, x_2, \ldots, x_{k-i})$, the values 0 and 1 for x_{k-i+1} are both canalyzing and lead to complementary outputs. To be concise, a complement canalyzing 2^i-block will be abbreviated as a CC-2^i-block. It is worthwhile to note that all complement canalyzing blocks are disjoint, so that a CC-2^i-block and a CC-2^j-block will have no rows in common. We then consider the vector $\mathbf{n} = (n_1, n_2, \ldots, n_{k-1}, n_k)$ where n_i is the sum of the cardinalities of the CC-2^{k-i+1}-blocks, that is the total number of rows in the truth table belonging to such blocks. Given

the vector **n**, Wuensche defines the fractions $R_1, R_2, \ldots, R_k$:

$$R_i = \frac{n_i}{2^k} \tag{5.2}$$

Since the complement canalyzing blocks are all disjoint, the sum of the R_i is at most 1. Finally, the parameter Z_{left} for this BF f (with the given ordering of inputs) is given by:

$$\begin{aligned} Z_{left} &= R_k + R_{k-1}(1 - R_k) + R_{k-2}(1 - R_{k-1})(1 - R_k) + \ldots & (5.3) \\ &= R_k + \sum_{i=1}^{k-1} R_{k-i}\left(\Pi_{j=k-i+1}^{k}(1 - R_j)\right) & (5.4) \end{aligned}$$

Figure 5.4(a) shows Wuensche's approach to compute the elements of **n** by iterating over all candidate 2^i-blocks, beginning with $i = 1$. Figure 5.4(b) illustrates our alternate approach that is based on recursively visiting $(k-i+1)$ dimensional hyperplanes (these have a direct correspondence with the 2^{k-i+1}-blocks), going from the largest blocks to the smallest ones and checking at each step of the recursion whether the hyperplanes contribute to **n**. Having obtained **n**, Figure 5.4(c) shows how to compute Z_{left}. The pseudocode for this algorithm is provided in Algorithm 1. Some interesting properties of Z_{left} for any BF are: (a) Z_{left} of f is invariant under the negation of any of its inputs (see Appendix B, Section B.2, Property B.1). (b) Z_{left} of f is invariant under complementation (see Appendix B, Section B.2, Property B.2).

In a similar fashion, one can compute the value of Z_{right} too, which is essentially computing the Z_{left} with the reversed order of the inputs to the BF. In the context of CA, Wuensche then defines his Z-parameter as the maximum of Z_{left} and Z_{right}.

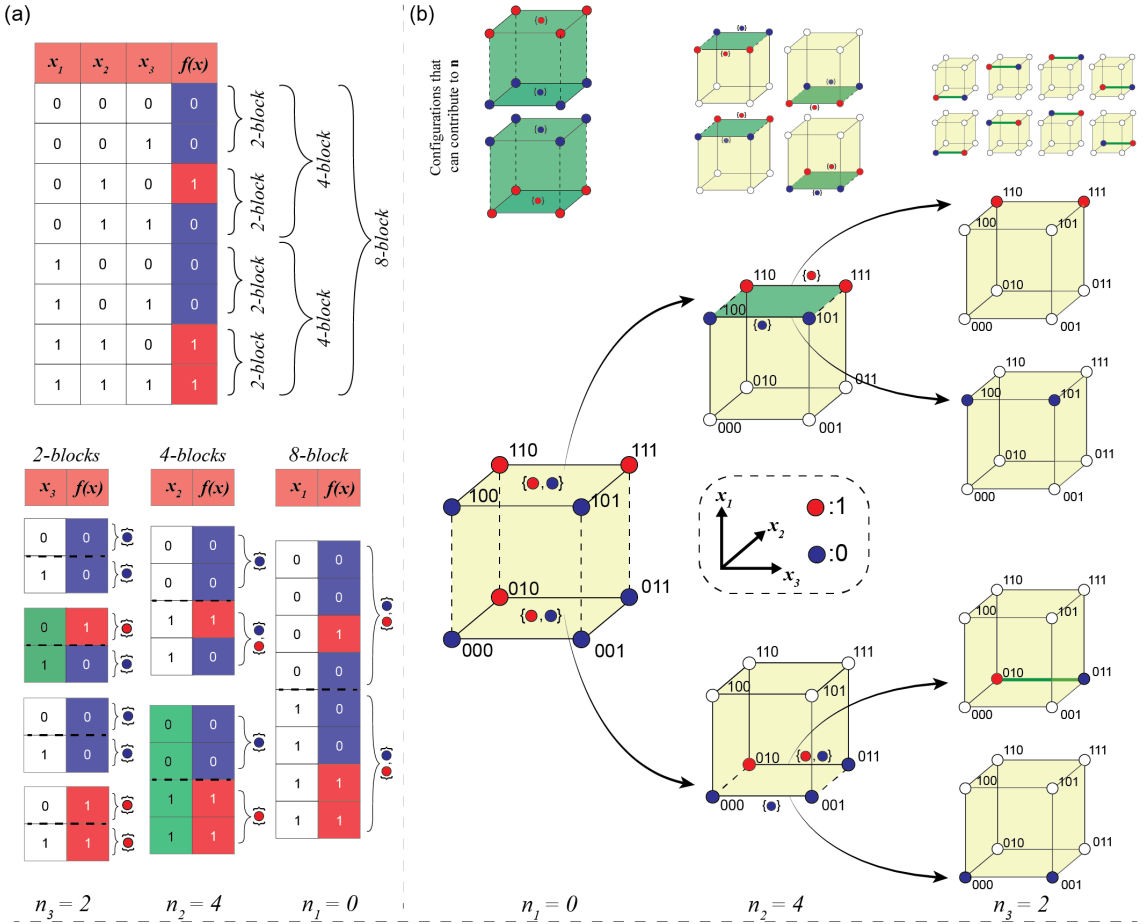

(a)
(b)
(c)
(d)
Configurations that can contribute to n
2-blocks
4-blocks
8-block
$n_3 = 2$
$n_2 = 4$
$n_1 = 0$
$n_1 = 0$
$n_2 = 4$
$n_3 = 2$
$n_1 = 0 \rightarrow R_1 = n_1/2^3 = 0$
$n_2 = 4 \rightarrow R_2 = n_2/2^3 = 0.5$
$n_3 = 2 \rightarrow R_3 = n_3/2^3 = 0.25$
$Z_{left} = R_k + R_{k-1}(1 - R_k) + R_{k-2}(1 - R_{k-1})(1 - R_k) + \ldots$
$Z_{left} = R_3 + R_2(1 - R_3) + R_1(1 - R_2)(1 - R_3)$
$Z_{left} = 0.25 + 0.5(1 - 0.25) + 0(1 - 0.5)(1 - 0.25) = 0.625$
$Z_{max} = 0.75$
$Z_{min} = 0.25$
$Z_{ave} = \frac{0.75 + 0.25 + \ldots + 0.75}{6} = 0.542$
$Z_{mid} = \frac{Z_{max} + Z_{min}}{2} = 0.5$

Figure 5.4: Calculation of Z-parameters of BFs in BN models. **(a)** Calculation of the vector **n** from the truth table as proposed by Wuensche. Top panel: The truth table of a particular BF for which the Z-parameters will be computed. The colors red and blue correspond to the output values 1 and 0 respectively. The list of 2^i-blocks are shown via braces, where $i \in \{1, 2, 3\}$. Bottom panel: For every 2^i-block, get the values spanned by the output in its top (respectively bottom) half (these halves are separated by the dashed lines). The corresponding values are shown to the right of the braces. If there is a single value for both top and bottom braces and if these are different, then we have a CC-2^i-block and it contributes 2^i to **n**. The corresponding cells in the column of the $(k - i + 1)^{th}$ variable are colored in green. n_1, n_2 and n_3 are the contributions from the CC-2^3-, CC-2^2- and CC-2^1-blocks (of values 0, 4 and 2 here) defining the vector **n**. **(b)** An equivalent formalism for calculating **n** on the Boolean hypercube. This sub-figure illustrates how elements of **n** can be computed by recursively splitting the Boolean hypercube into two hyperplanes, obtaining at each step an element of **n** starting from n_1. Top panel, from left to right: the cubes show which configurations (vertex colorings) can contribute to n_1, n_2 and n_3 respectively. For instance, a 2^3-block contributes to n_1 if and only if all the vertices in the plane $x_1 = 0$ are colored red (or dark blue) *and* all those in plane $x_1 = 1$ are colored dark blue (or red), as shown at the top left of this sub-figure. Bottom panel: By comparing the actual node coloring to the templates contributing in the top panel, we see that there is no contribution to **n** from the 2^3-block, so $n_1 = 0$. The next step of the recursion (just to the right) concerns contributions of 2^2-blocks (associated with two faces of the cube). Here, the upper face (colored dark green) contributes to n_2 whereas the lower face does not (based again on comparing to the templates just above them). Finally, still on the right of that, we have the case of the 2^1-blocks that may contribute to n_3. We have colored in green the ones that do contribute. Reaching the 2^1-blocks marks the end of the recursion. Note that if a grouping (cube, face or in higher dimensions any hyperplane) contributes to **n**, then none of its sub-parts will contribute to **n**. **(c)** This sub-figure summarizes the computation of Z_{left} of a BF once the vector **n** is obtained. **(d)** Computation of the four Z-parameter variants: Z_{max}, Z_{min}, Z_{ave} and Z_{mid}. Left panel: The Boolean expression of the BF in the truth table in sub-figure (a) is $x_2 \wedge (\overline{x}_3 \vee x_1)$, as displayed in the center of the hexagon. We then apply the $k!$ permutations (cf. arrows) to produce the (distinct) permuted BFs as shown at the end of these arrows. The values shown in the outer hexagon are the Z_{left} values for each of the permuted Boolean expressions. Right panel: a hexagon is produced for each of the 4 variants of the Z-parameter: Z_{max} is the maximum Z_{left} value over all permutations. Z_{min} is the minimum Z_{left} value over all permutations. Z_{ave} is the average of the Z_{left} over all permutations. Z_{mid} is the average of Z_{max} and Z_{min}.

Being a probability, Z varies between 0 and 1. Ideally a value of Z close to 1 should indicate that there is a unique pre-image for many – if not most – states. Then many or most of the STG's nodes should have in-degrees equal to 1 (corresponding to reversible dynamics) in which case the STG will not be bushy at all. In contrast, a low Z value should indicate strongly irreversible dynamics in which case the STG will have many states with a high in-degree. Because the average in-degree of the STG is mathematically constrained to be 1, this situation will also lead to having many *GoE* states, so the STG will be bushy.

Algorithm 1 Recursive algorithm to determine the vector **n** arising in the computation of Z_{left}.

Input 1: f is the truth table of the BF represented as a string of bits.
Input 2: k_{var} is the number of inputs of f.
Input 3: k_{fixed} is the number of inputs of the BF for which the Z_{left} is to be computed.
Output: The array of elements of the vector **n**

function ZLEFTVECTOR(f, k_{var}, k_{fixed})
  if $k_{var} == 1$ **then** ▷ If the contribution of 2^1-blocks have to be computed
    if (f ==‘01’) or (f ==‘10’) **then**
      return $[2, 0, \ldots, 0]$ ▷ The **n** vector of length k_{fixed} with element $n_k = 2$ and other elements as 0
    else
      return $[0, 0, \ldots, 0]$ ▷ The **n** vector of length k_{fixed} with all elements as 0
    end if
  else
    Split f into upper (f_{upper}) and lower (f_{lower}) halves.
    $Unique f_{upper} \leftarrow$ Values arising in f_{upper}
    $Unique f_{lower} \leftarrow$ Values arising in f_{lower}
    if (Length of $Unique f_{upper}$ is 1) **&** (Length of $Unique f_{lower}$ is 1) **&** ($Unique f_{upper} \neq Unique f_{lower}$) **then** ▷ These conditions guarantee that there will be a contribution to **n** vector
      $CountArray \leftarrow [0, 0, \ldots, 0]$ ▷ Count array is the **n** vector
      $CountArray[k_{inp} - 1] \leftarrow 2^{k_{inp}-1}$
      return $CountArray$
    else
      return ZLEFTVECTOR($f_{upper}, k_{var} - 1, k_{fixed}$) + ZLEFTVECTOR($f_{lower}, k_{var} - 1, k_{fixed}$)
    end if
  end if
end function

5.3.1.3 Z-parameters for Boolean networks

In this section we propose a number of extensions to Wuensche's Z-parameter for BFs. Since there is no natural spatial order or arrangement of the inputs to a node in a BN, we define our Z-parameters in a manner that accounts for all permutations on its inputs (see Chapter 2, Section 2.4). Given a k-input BF, we compute the Z_{left} for all $k!$ permutations on its inputs, from which we define four Z-parameter variants: $\boldsymbol{Z_{max}}$, $\boldsymbol{Z_{min}}$ and $\boldsymbol{Z_{ave}}$ are the maximum, minimum and average respectively, and $\boldsymbol{Z_{mid}}$ is the average of Z_{max} and Z_{min} (see Fig. 5.4(d)).

The analytical formulas to compute Z_{max}, Z_{min} and Z_{mid} for k-input NCFs with bias P are (see Chapter 2, Section 2.1 for the definition of bias and Appendix B, Section B.2 for the proofs):

$$Z_{max}^{NCF} = \begin{cases} \dfrac{P}{2^{k-1}} \text{ if } P < 2^{k-1}, \\ 2 - \dfrac{P}{2^{k-1}} \text{ if } P > 2^{k-1} \end{cases} \tag{5.5}$$

$$Z_{min}^{NCF} = \frac{1}{2^{k-1}} \tag{5.6}$$

$$Z_{mid}^{NCF} = \begin{cases} \dfrac{1+P}{2^k} & \text{if } P < 2^{k-1} \\ 1 - \dfrac{P-1}{2^k} & \text{if } P > 2^{k-1} \end{cases} \tag{5.7}$$

More explicitly, Z_{min} is dependent only on the number of inputs (k), while Z_{max} and Z_{mid} vary linearly with the bias P for $0 < P < 2^{k-1}$ and $2^{k-1} < P < 2^k$ for a fixed k, making their computations efficient and scalable for NCFs.

Regarding Z_{ave}, we found that its computation for any k-input BF as mentioned above is equivalent to computing the average of Z_{left} values only for the *non-equivalent* permutations of that BF (see Chapter 2, Section 2.4.2), leading us to two conjectures, both of which we prove in Section B.3 in Appendix B: (a) The number of non-equivalent permutations divides $k!$. (b) Each of the m *non-equivalent* permutations occurs an equal number of times. Furthermore, the fraction of non-equivalent permutations of BFs decreases with increasing number of inputs for BFs randomly drawn from NCFs and RoFs, and for BFs in reconstructed BNs, whereas it increases for BFs randomly drawn from EFs or EUFs (see Figure 5.5). In sum, these results enable us to significantly reduce the computation time of Z_{ave} in practice.

We define the *network* Z-parameter as the average of a considered Z-parameter over all the

nodes of a BN and denote it by $\overline{Z}_x$ where $x \in \{max,\ min,\ ave,\ mid\}$.

$$\overline{Z}_x = \frac{1}{N}\sum_{i=1}^{N} Z_x^i \tag{5.8}$$

Here, $i \in \{1, 2, 3, \ldots, N\}$ is the index of a node in the BN with N nodes, and Z_x^i is the Z-parameter of the BF at node i.

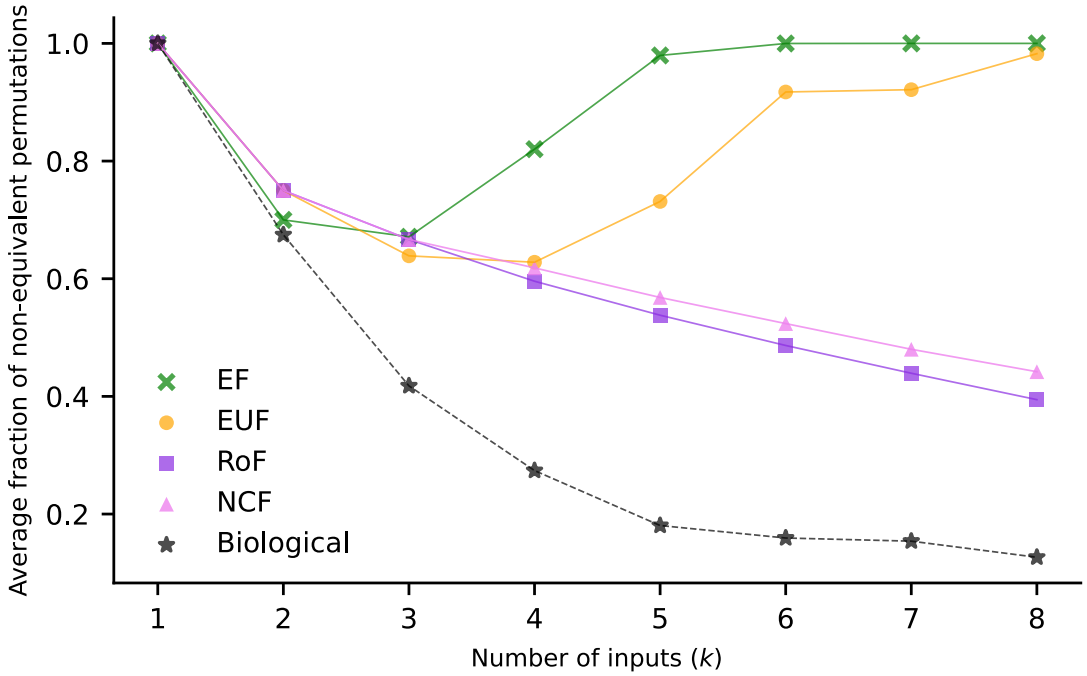


Figure 5.5: The fraction of non-equivalent permutations for different types of BFs. The x-axis is the number of inputs to the BF. The y-axis is the average of the fraction of non-equivalent permutations over a set of k-input BFs. Each data point on the plot represents the average of the fractions of non-equivalent permutations over the set of all (or sampled) k-input BFs belonging to a particular type of BF (as indicated by the shape of the associated point). For inputs $k = 1, k = 2, k = 3, k = 4, k = 5$, the exhaustive set of BFs was used in the computation for all types of BFs (except for 5-input EFs for which 100000 BFs were sampled), whereas for inputs $k = 6, k = 7, k = 8$, 100000 sampled BFs were generated for all types of BFs. The data points corresponding to the 'biological' function type were obtained from a reference biological dataset of 2687 BFs derived from 88 published Boolean models [51]. It is seen that the 'biological' case generally consists of BFs with a low fraction of non-equivalent permutations, justifying the usefulness of our approach for computing Z_{ave}.

5.3.2 Network sensitivity as a descriptor of local dynamics

In addition to the four variants of Z-parameter, we also consider network sensitivity (see Chapter 2, Section 2.6) [53, 58] as a descriptor of local dynamics. Although the measure is known to relate to robustness in BNs, its association with STG structure has not been addressed in prior work, to our knowledge. Being defined as the mean of the average sensitivities of the BFs in the BN, it is also a BF-based measure, similar in spirit to the Z-parameters. Together, the four network Z-parameters $\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ along with the network sensitivity $\overline{s}$, constitute the set of descriptors of local dynamics. In the following sections, we examine the distributions of these descriptors across different ensembles, relate them to global properties of the STG such as bushiness and convergence, and assess their effectiveness as computationally tractable proxies.

5.4 Biologically meaningful BFs have lower local dynamics descriptor values

To bypass the computational challenges encountered when computing global measures for larger networks, we explore five descriptors of local dynamics (see Section 5.3) as potential proxies: $\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ and $\overline{s}$. Consequently, we begin by examining their distributions across different ensembles (see Section 5.1). The distributions of $\overline{Z}_{max}$ (see Appendix B, Figures B.7 and B.8), $\overline{Z}_{min}$ (see Appendix B, Figures B.9 and B.10), $\overline{Z}_{ave}$ (see Appendix B, Figures B.11 and B.12), and $\overline{Z}_{mid}$ (see Figure 5.6(a) and Appendix B, Figure B.13) reveal that ensembles generated with biologically meaningful BFs have lower values for all four Z-parameters compared to the ensemble generated with EFs. This is consistent across all ten reconstructed GRNs (see Table 5.1). Restricting to the ensembles generated with biologically meaningful BFs, we find that $\overline{Z}_{max}$ and $\overline{Z}_{ave}$ do not show any particular trend across the ten GRNs, whereas $\overline{Z}_{min}$ and $\overline{Z}_{mid}$ consistently achieve the lowest values in the NCF-ensemble, across all ten GRNs. Note that the $\overline{Z}_{min}$ for the NCF-ensembles always have zero variance since Z_{min}^{NCF} depends only on k (see Eq. (5.6) and Appendix B, Section B.2 for a proof). The distributions of our last descriptor of local dynamics, network sensitivity denoted by $\overline{s}$, are shown in Figure 5.6(b) and Figure B.14 in Appendix B, indicating that NCFs lead to the lowest values compared to all other BF classes (in agreement with previous observations [51]).

Lastly, the Spearman correlation coefficient among the five descriptors of local dynamics, as given via heatmaps in Figures B.15 and B.16 in Appendix B, shows that most pairs of descriptors show moderate to very strong positive correlation across various GRNs. This is illustrated in particular via scatter plots between $\overline{Z}_{ave}$ and $\overline{Z}_{mid}$ (see Appendix B, Figures B.17 and B.18), and $\overline{s}$ and $\overline{Z}_{mid}$ (see Appendix B, Figures B.19 and B.20). In the following section, we explore which of these descriptors can serve as good proxies of bushiness or convergence.

5.5 Descriptors of local dynamics can serve as good proxies of bushiness of STGs

Wuensche showed that, in CA, the Z-parameter exhibits a 'marked' correlation with the G-density [62] of the STG. This observation suggests that qualitative information about the bushiness of an STG in CA may be inferred directly from properties of the local update rules, without explicitly constructing the full STG. Extending this to the BN framework, we ask whether descriptors of local dynamics (in fact, characteristics of the rule tables for each node) can serve as good proxies of bushiness. In what follows, we examine the relationship between the descriptors of local dynamics (see Section 5.3) and the G-density of the STG, and assess their potential as predictors of the bushiness of STG.

5.5.1 Correlation between the descriptors of local dynamics and G-density

To quantify the relationship between the descriptors of local dynamics and bushiness of STG, we compute the Spearman correlation coefficients of the G-density with the five descriptors of local dynamics: $\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ and $\overline{s}$. The resulting correlations are summarized in the heatmaps shown in Figure 5.7 and Figure B.21 in Appendix B. From these figures it is evident that for most GRNs, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ and $\overline{s}$ show a moderate to strong negative correlation with the G-density and are more strongly correlated with it than are $\overline{Z}_{max}$ and $\overline{Z}_{min}$. This suggests that those three descriptors may serve as useful proxies of the bushiness. Note that since $\langle d_{in} \rangle_{non-GoE}$ is a monotone increasing function of the G-density (see Chapter 3, Eq. (3.6)), the Spearman correlation coefficients of $\langle d_{in} \rangle_{non-GoE}$ with the five descriptors will be the same as that of G-density and hence are not

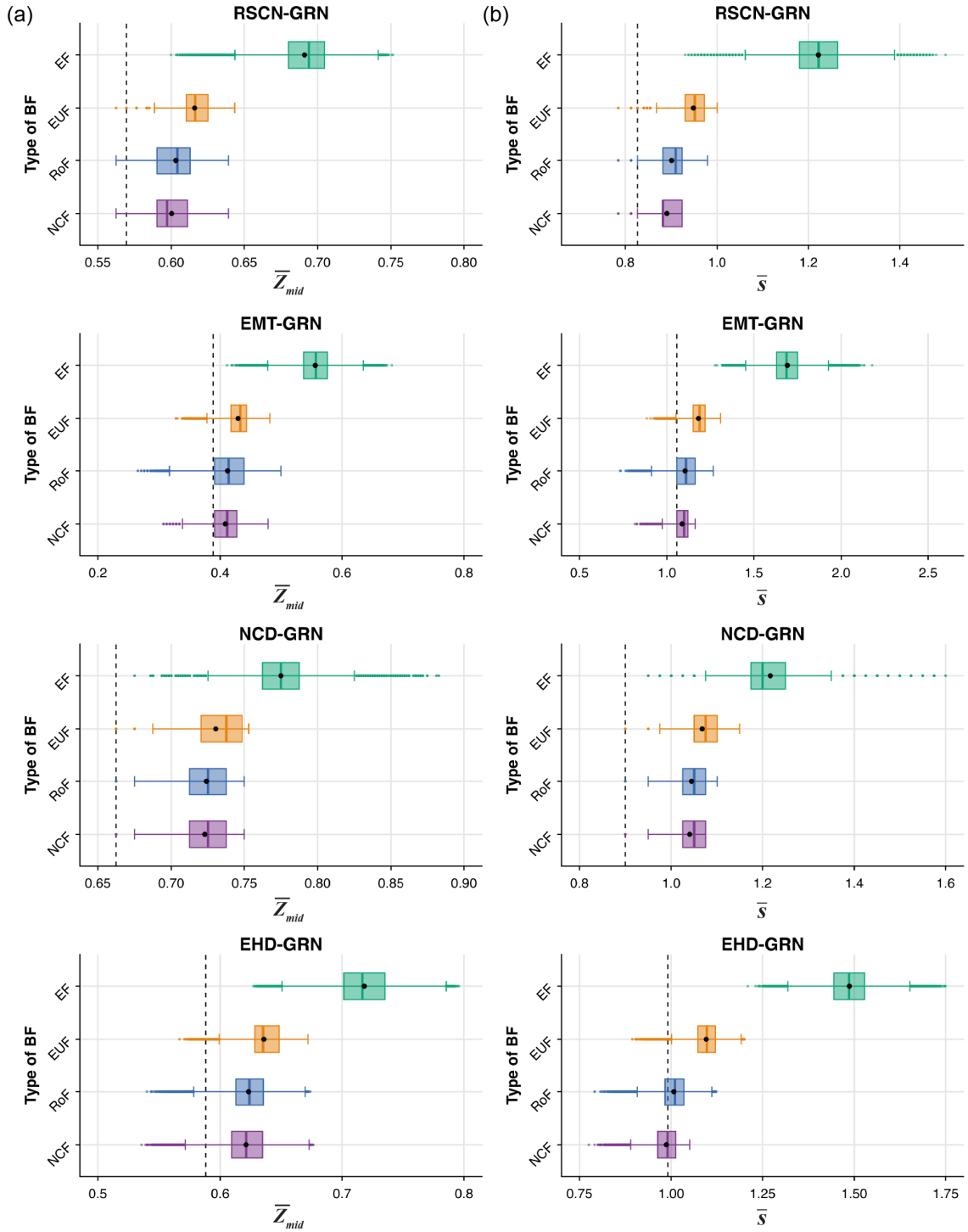


Figure 5.6: Distribution of the values of the descriptors of local dynamics for different ensembles generated using network structures from four published Boolean GRNs. The box plots in **(a)** and **(b)** display the distributions of $\overline{Z}_{mid}$ and $\overline{s}$ respectively in the four ensembles that each use a given type of regulatory logic. The mean and median of the distributions are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in (a) and (b) correspond to the $\overline{Z}_{mid}$ and $\overline{s}$ respectively for the Boolean model provided by the modelers in the published article.

shown in the heat maps. The scatter plots of G-density with each of the five descriptors of local dynamics are provided in Appendix B: $\overline{Z}_{max}$ (see Figures B.22 and B.23), $\overline{Z}_{min}$ (see Figures B.24 and B.25), $\overline{Z}_{ave}$ (see Figures B.26 and B.27), $\overline{Z}_{mid}$ (see Figures B.28 and B.29) and $\overline{s}$ (see Figures B.30 and B.31).

5.5.2 Performance evaluation using AUROC and AUPRC

The observed correlations suggest that certain local descriptors, such as $\overline{Z}_{mid}$ and $\overline{s}$ may be used as practical proxies for the G-density in situations where generating the complete STG is computationally infeasible. To quantitatively assess the predictive value of these descriptors in a model selection context, we evaluate their performance using standard classification metrics: the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). In this setting, the objective is to select models with high G-density based solely on a local descriptor. In a computational modeling framework, one typically will have many (too many) models that are compatible with the constraints imposed, such as presence of given steady states. To shrink the size of this set of plausible models, we impose in addition that the models be robust. Since we do not have these models' G-density, we instead compute their $\overline{Z}_{mid}$ values. Then model selection corresponds to keeping only the models for which $\overline{Z}_{mid}$ is say in some X percentile, with the hope that doing so will select models whose G-density is on average at a high percentile.

Implementing AUROC and AUPRC metrics requires ground truths. For that, we assign truth labels using increasing percentile thresholds on the G-density of the models. Specifically, we iterate through the thresholds ranging from 0.1% to 99.9%, labeling models having G-density percentile above the threshold as 'true', and the others as 'false'. For each set of ground truths we compute the AUROC and AUPRC scores of the proxy $\overline{Z}_{mid}$, and then these are averaged over all threshold levels, providing a comprehensive evaluation of the proxy's performance.

For instance, if one selects the models having average relative rank of 0.05 for $\overline{Z}_{mid}$, the corresponding average relative rank for the G-density is typically 0.2, very far from the value of 0.5 which would arise if $\overline{Z}_{mid}$ had no predictive power (see Appendix B, Figures B.32 and B.33). To assess more globally the relevance of using $\overline{Z}_{mid}$ as a proxy for G-density, we computed the corresponding AUROC and AUPRC measures (see Appendix B, Table B.41). From the table, we

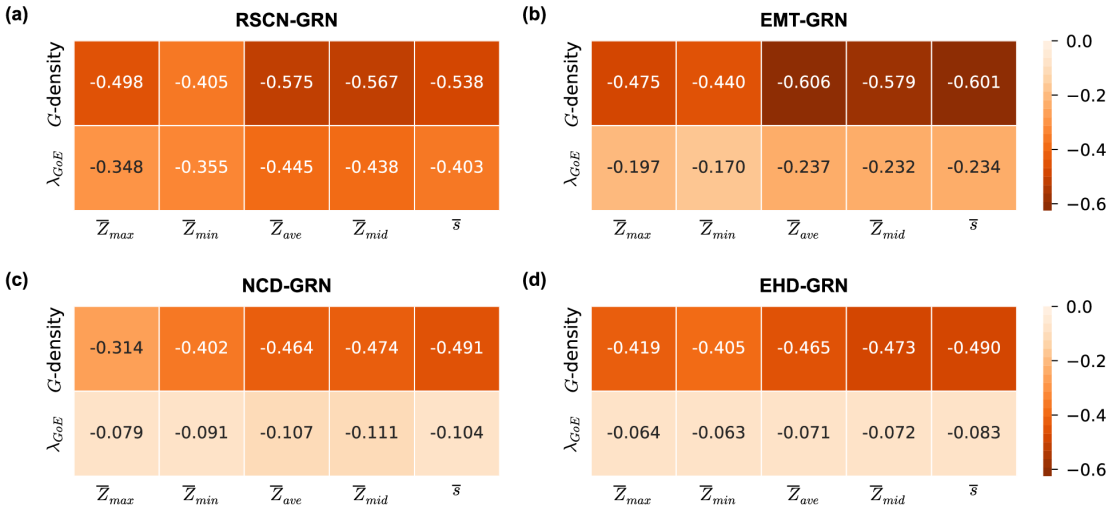


Figure 5.7: Correlation heat map between descriptors of local dynamics and global measures (bushiness and convergence rate) for ensembles generated using network structures from four published Boolean GRNs. The EF-ensemble for each network structure was used to generate the heat maps. The rows of each sub-figure correspond to quantities computed on the STG, namely G-density and λ_{GoE}. The columns of each sub-figure correspond to descriptors of local dynamics, namely, the network Z-parameters ($\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$) and network sensitivity ($\overline{s}$). The heat maps show the pair-wise Spearman correlations between descriptors of local dynamics and global measures of bushiness and convergence rate defined on the STG. The descriptors of local dynamics show a moderate to strong negative correlations with G-density and very weak to moderate negative correlations with λ_{GoE}.

observe for instance that the average AUROC value consistently exceeds 0.72 for majority of the GRNs, indicating favorable performance in this context.

While this quantification of performance is limited to networks of moderate size, for which the computation of the G-density is tractable, it nonetheless offers insight into the expected behavior in larger networks.

5.6 The average convergence rate of trajectories originating at random states is a good proxy for the average convergence rate of trajectories originating at *GoE* states

Analogous to the case of G-density, we compute the correlation of λ_{GoE} with the descriptors of local dynamics (see Figure 5.7 and Appendix B, Figure B.21) and find that they are negative as expected, but quite weak for most ensembles. Therefore, it is inappropriate to use any of these descriptors as proxies for λ_{GoE}. Note, nevertheless, that for most GRNs, the $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ and $\overline{s}$ show stronger correlations than $\overline{Z}_{max}$ and $\overline{Z}_{min}$. The scatter plots between the λ_{GoE} and each of the five descriptors of local dynamics are provided in Appendix B: $\overline{Z}_{max}$ (see Figures B.34 and B.35), $\overline{Z}_{min}$ (see Figures B.36 and B.37), $\overline{Z}_{ave}$ (see Figures B.38 and B.39), $\overline{Z}_{mid}$ (see Figures B.40 and B.41) and $\overline{s}$ (see Figures B.42 and B.43).

As an alternative to these descriptors, we propose λ_{random} (see Chapter 3, Section 3.4.1) as a possible proxy for λ_{GoE}. The advantage of this approach is that, computing λ_{random} does not require construction of the full STG. Instead, it can be evaluated by sampling a set of random initial states from the state space and computing the corresponding trajectories. Our grounds for this proposition is that λ_{GoE} is very strongly correlated with λ_{all} (see Chapter 3, section 3.4.1; Appendix B, Figures B.44 and B.45), which provides an impetus to compute λ_{random} because if so, one no longer has to compute the STG. Figures B.46 and B.47 in Appendix B reveal that λ_{random} can indeed serve as an excellent predictor of λ_{GoE}.

5.7 Discussion

In the context of CA, Wuensche showed that the signature of highly ordered or convergent dynamics is the presence of short, highly branching transient trees, with a high proportion of leaves in the STG, whereas the signature of chaotic dynamics is long transients, low branching in the STG and a low proportion of leaves therein [62, 78, 157]. Such characterizations are absent in generalized automata that mimic biological systems, namely, BNs – which generally lack a regular network architecture and employ heterogeneous rules across their nodes. **In this chapter, we presented a systematic analysis of the bushiness and convergence of STGs for different reconstructed GRNs using the measures G-density and λ_{GoE}, showing that biologically meaningful BFs give rise to more bushy and more convergent STGs than random EFs.** Notably, the use of ten different GRNs makes our results very general and minimizes any bias that may be associated with particular GRNs. Note, biologically meaningful BFs occupy a tiny fraction in the space of all BFs and consequently can severely restrict the space of biologically plausible Boolean models (see Appendix B, Tables B.31-B.40) and enable model selection therein [64].

Contrary to our expectation that higher bushiness (i.e., higher contraction rate) results in higher convergence, we found that the G-density and λ_{GoE} measures are only weakly correlated. A posteriori, it is easy to convince oneself that the two measures indeed capture different features. For example, consider a rooted tree of constant connectivity K for which all the leaves are at the same distance H from the root. The bushiness, that is G-density, is close to $(K-1)/K$ while the transient times are all equal to H, leading to a convergence rate of $1/H$. Thus bushiness and convergence rate can be varied essentially independently.

In this chapter, we were able to exhaustively enumerate the *GoE* states of all the models. However, identifying the *GoE* states of a BN is, in general, a #**P**-complete problem, implying that computing the G-density of an arbitrary BN is at least as hard as solving an NP-complete problem [161]. This computational barrier motivates us to introduce proxies that can bypass this computational complexity and provide information about the bushiness of the STG without requiring explicit enumeration of its *GoE* states. Among the network Z-parameters we proposed and explored in this chapter, the choice $\overline{Z}_{mid}$ showed the highest correlation with G-density, and for the class of NCFs it is highly scalable. We found that the network sensitivity ($\overline{s}$) is also a

good proxy of G-density. This observation is supported by the fact that NCFs, which are known to produce more ordered dynamics compared to EUFs and EFs [51, 53], achieve the minimum average sensitivity among all BFs [51], leading to lower values of $\overline{s}$. Such dynamics are directly linked to the robustness of the system to perturbations, and therefore, a highly bushy STG is a signature of robust dynamics. Note that the descriptors defined at the level of individual nodes are 'unaware' of the network architecture (which is not an issue in CA). Z-parameters other than the four discussed in this chapter may be defined of course. For instance, one could consider the median of all Z_{left} values of a BF. But we find that this measure is computationally more demanding compared to Z_{mid}, and exhibits weaker correlation with G-density at the network level than $\overline{Z}_{mid}$ or $\overline{Z}_{ave}$. Finally, since the average convergence rate correlates very weakly with the descriptors of local dynamics, we tested an alternative approach and found that the average convergence rate of trajectories originating from randomly sampled states (λ_{random}) serves as an excellent proxy for λ_{GoE}. **Overall, a key methodological contribution of this chapter is the development of computationally efficient proxy measures that avoid the need for computationally expensive state-space analyses while providing reliable estimates of the corresponding global dynamical properties.**

Data and code availability statement

All data and codes needed to reproduce the results in this chapter are deposited in GitHub and are available at: https://github.com/asamallab/BushySTG.

Chapter 6

A systematic evaluation of simple threshold-based Boolean rules for modeling gene regulatory networks

This chapter is based on the published manuscript [80].

Chapter overview. This chapter presents a systematic evaluation of simple threshold-based Boolean rules for modeling gene regulatory networks (GRNs). It begins by introducing threshold functions and two commonly used threshold majority rule (TMR) variants, namely Ising majority rules (IMRs) and Bit-based majority rules (BMRs). The chapter then examines the functional properties, intrinsic logical complexity, and empirical enrichment of TMRs through comparisons with nested canalyzing functions (NCFs). Next, it evaluates the ability of TMR-based models to recover biological attractors and their associated basin size distributions. Finally, it compares the dynamical properties of TMRs and NCFs and discusses the implications of these findings for selecting appropriate regulatory logic in Boolean models of GRNs.

In the post-genomic era, large-scale data-driven analyses have demonstrated that the architecture of biological networks exhibits significant deviations from randomness [8, 44, 45], while quantitative modeling indicates that the regulatory logic is not random either [7, 33, 34]. However, determining the precise regulatory logic (encapsulating the combinatorial control) at each node of

a gene regulatory network (GRN) remains a formidable task. As a result, modelers often rely on certain assumptions that can provide reasonable outcomes tailored to the specific context of their research.

The use of threshold functions (ThFs) in the context of modeling Boolean GRNs has drawn significant attention. The concept dates to the seminal work of McCulloch and Pitts, who modeled neuronal dynamics [70], and has since been widely adopted in the design of digital circuits and studies of artificial neural networks [162–165]. Surprisingly, the use of ThFs in the context of GRN modeling was initiated only in 1994 by Wagner, who focused on transcriptional networks [73, 74]. More recent work has shown that ThFs can be successfully applied to model the dynamical behavior of many different regulatory systems [75, 166–168]. However, a realistic application of ThFs hinges on two types of parameters: the 'weights' or 'strengths' associated with the different regulators and the activation thresholds specific to each target node. Obtaining reliable values for these parameters is a major obstacle. For example, in their model on *Arabidopsis thaliana* flower morphogenesis, Mendoza and Alvarez-Buylla treated both thresholds and interaction strengths as integers to simplify computations in the absence of detailed quantitative information [75]. Following a different strategy, Sevim and Rikvold [167] drew the interaction weights from a Gaussian distribution [167], while Li *et al.* [168] recently employed a regularized logistic regression to infer both the weights and the thresholds.

To circumvent the difficulty of parameter estimation, many modelers adopt a coefficient-free scheme where all regulatory interactions are assigned a strength of ± 1 and activation thresholds are set to zero. Under these assumptions, the target node's next state is solely determined by its own state and the weighted sum of its regulators' current states. Two main variants of this approach have become standard in the literature. In the first variant, each node takes the values $+1$ (to indicate 'on') or -1 (to indicate 'off'). In this chapter, we refer to this variant as IMR (Ising majority rule). This variant has been employed in the studies of epithelial-mesenchymal transition dynamics [65, 66, 169]. It has also been used in comparative analyses of frustration in biological versus random networks [170]. In the second variant, nodes take on binary values in the Boolean sense, adopting 1 for 'on' and 0 for 'off'. We refer to this variant as BMR (Bit-based majority rule). This variant has been employed in the cell-cycle models of *S. cerevisiae* [67] and *S. pombe* [34]. BMRs have also found applications outside the realm of GRNs, for instance in

modeling of the interactions of gut microbiome species [171, 172]. These simple variants (IMRs and BMRs) of ThFs, which we collectively refer to as threshold majority rules (TMRs), provide an elegantly simple and uniform framework for modeling GRNs. Nonetheless, it is important to examine if this minimalism overlooks key aspects of biological behavior. To this end, we present a systematic analyses of TMRs to assess their appropriateness as a go-to modeling strategy for GRNs.

6.1 Threshold function and its subtypes

Threshold functions (ThFs) play a very prominent role in modeling regulation in both artificial neural networks and gene regulatory networks (GRNs) [70–75, 164, 173]. In such a framework, the state of a given node (e.g., a gene) is determined by the collective effect of its activatory and inhibitory regulators [70, 174]. Let x_i denote the target node, and assume that it receives signals from k regulators $x_1, x_2, \ldots, x_k$. A ThF assigned to the node x_i can be expressed as

$$x_i(t+1) = \Theta\left(\sum_{j=1}^{k} w_{ij} x_j(t) + \theta_i\right) \tag{6.1}$$

where $w_{ij} \in \mathbb{R}$ is the weight (i.e., regulatory strength) of the directed edge from x_j to x_i, and $\theta_i \in \mathbb{R}$ is the threshold for node x_i. The function $\Theta(s)$ is an indicator function, that returns 'on' (1) if $s > 0$, 'off' (0 or -1) if $s < 0$, and a predefined value when $s = 0$.

The class of functions defined by Eq. (6.1) is also known as linearly separable Boolean functions (LSBFs) [175, 176]. A k-input BF can be represented as a k-dimensional hypercube whose 2^k vertices correspond to input combinations and are labeled as 'on' or 'off'. LSBFs (i.e., ThFs) form a specialized subclass of BFs whose hypercube vertices can be separated by a single hyperplane with all 'On' vertices on one side, and all 'Off' vertices on the other. Moreover, every ThF belongs to the family of unate functions (UFs) (see Chapter 2, Section 2.5 for the definition of UF) [177, 178]. These two classes coincide for $k \leq 3$ but ThFs form a proper subset of UFs for $k \geq 4$.

The exact numbers of ThFs are known for $k \leq 9$ [179]. However, we are really only interested in *effective* [81] ThFs (EThFs). Effective functions (EFs) are those for which all inputs are relevant, i.e., for each input, there exists a combination of remaining input values such that toggling the

input ('on' to 'off' or 'off' to 'on') changes the output (see Chapter 2, Section 2.5 for a more formal definition). Otherwise, the BF is called an ineffective function (IEF), and the inputs that do not influence the output are referred to as ineffective inputs. The exact numbers of EThFs have been published only for $k \leq 8$ [162]. However, the counts of EThFs can be easily computed using a recursion for any k up to 9. Let $|\mathrm{ThF}_k|$ denote the total number of ThFs on k inputs, and let $|\mathrm{ThF}_k^{eff}|$ denote the number of EThFs of k inputs. Every ThF in ThF_k is either an EThF or obtained by extending an EThF on $i < k$ inputs to a k-input BF by adding $(k - i)$ dummy inputs. Given that there are exactly two constant BFs for any k (constant BFs are also ThFs), and that each EF on i inputs can be embedded into a k-input BF in $\binom{k}{i}$ ways, we have

$$|\mathrm{ThF}_k^{eff}| + 2 + \sum_{i=1}^{k-1} \binom{k}{i} |\mathrm{ThF}_i^{eff}| = |\mathrm{ThF}_k|$$

Rearranging,

$$|\mathrm{ThF}_k^{eff}| = |\mathrm{ThF}_k| - 2 - \sum_{i=1}^{k-1} \binom{k}{i} |\mathrm{ThF}_i^{eff}|,$$

where $|\mathrm{ThF}_1^{eff}| = |\mathrm{ThF}_1| - 2 = 4 - 2 = 2$ as each of four BFs with 1 input is a ThF. The two EThFs with 1 input are $f(x) = x$ and $f(x) = \overline{x}$. Now, this recursion allows us to obtain the number of EThFs for any k up to 9; those values are provided in Table C.1 in Appendix C.

A particularly important subclass of ThFs is formed by nested canalyzing functions (NCFs; see Chapter 2, Section 2.5) [7, 68, 69]. Interestingly, the NCF class is the intersection of the read-once functions (RoFs) (see Chapter 2, Section 2.5 for the definition of RoFs) [108] and ThFs [180]. Being a subset of ThF, a suitable choice of w_{ij} and θ_i in Eq. (6.1) can yield any NCF. Note that we used recursion to exhaustively enumerate all NCFs up to $k = 9$ [64, 148].

In most experimental studies only the signs of the interactions (i.e., activation or inhibition) are reliably measured, while the precise strength of that regulation remains uncertain. While modeling the dynamics of GRNs, modelers often adopt the simplest possible 'unit-weight' assumption: each activatory or inhibitory edge from node x_j to x_i is assigned as $w_{ij} = +1$ or -1, respectively while $w_{ij} = 0$ if no edge exists. Thresholds are likewise set to 0 ($\theta_i = 0$), so that the update of node x_i depends only on the algebraic sum of its incoming signals and (potentially) the value of x_i itself. If that sum is positive, the node switches 'on'; if it is negative, it switches 'off'; and if it is exactly zero, the node retains its previous state. To date, modelers have typically employed one of

the two binary conventions for the node states when using these simple threshold-based rules:

(a) Each x_i takes values in $\{-1, +1\}$, which we refer to as *Ising majority rule* or IMR [65, 66].

(b) Each x_i takes values in $\{0, 1\}$, which we refer to as *Bit-based majority rule* or BMR [34, 67].

More formally:

- **IMR update:** Suppose in a network with N nodes, the node x_i is regulated under IMR by k regulators x_j for $j = 1, 2, \ldots, k$. Then its state at time $t + 1$ is given by:

$$x_i(t+1) = \begin{cases} +1, & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) > 0, \\ -1, & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) < 0, \\ x_i(t), & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) = 0. \end{cases} \tag{6.2}$$

- **BMR update:** Suppose in a network with N nodes, the node x_i is regulated under BMR by k regulators x_j for $j = 1, 2, \ldots, k$. Then its state at time $t + 1$ is given by:

$$x_i(t+1) = \begin{cases} 1, & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) > 0, \\ 0, & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) < 0, \\ x_i(t), & \text{if } \sum_{j=1}^{k} \sigma_{ij} x_j(t) = 0. \end{cases} \tag{6.3}$$

In IMR $x_i \in \{-1, +1\}$ and in BMR $x_i \in \{0, 1\}$ for each $i \in \{1, 2, \ldots, N\}$ and in both formalisms one has

$$\sigma_{ij} = \begin{cases} +1, & \text{if } x_j \text{ activates } x_i, \\ -1, & \text{if } x_j \text{ inhibits } x_i, \\ 0, & \text{if } x_j \text{ does not regulate } x_i. \end{cases} \tag{6.4}$$

Note that, in the case of BMR, if an input is off, it does not affect the output, irrespective of whether the input is activatory or inhibitory. However, this is not true in case of an IMR. We refer to both of these variants of simple threshold-based rules collectively as *Threshold Majority Rules* or TMRs.

TMRs may appear attractive since they bypass introducing regulatory weights and provide a simple and uniform way to model network dynamics. However, it is important to ask whether this simplicity comes at the cost of biological realism and more specifically, whether TMRs are

appropriate for modeling biological GRNs. To examine this, we undertake a systematic analysis of both variants of TMRs, namely IMRs and BMRs. We also compare them with the NCFs, which are highly enriched in empirical datasets of BFs [51]. First we address function-level properties by identifying BFs equivalent to TMRs, assessing whether they exhibit features such as odd bias and canalyzation as commonly observed in empirically curated BFs (e.g., NCFs) [51], examining their intrinsic regulatory complexity, and evaluating their frequency in empirical datasets relative to theoretical expectations. We then turn to network dynamics, investigating whether TMR-based models recover biological attractors and their basin-size distributions, and whether they exhibit dynamical stability or chaos. Taken together, through these analyses we assess whether the simplicity of TMRs justifies their continued use in Boolean modeling or warrants a more critical stance.

6.2 Representation of threshold majority rules as Boolean functions and their comparison to nested canalyzing functions

To make a meaningful comparison and to understand different properties of TMRs, it is essential to interpret TMRs in a common framework with NCFs or any other BF classes. Specifically, we want to explore the full range of input-output behavior, i.e., identify the BFs that correspond to IMRs and BMRs under different sign combinations of the regulators and their properties (see Figure 6.1).

We begin with IMRs. Consider a node x_i governed by an IMR and regulated by k *external* regulators x_j (i.e., regulators external to x_i, so that there is no self-regulation, i.e., no "self-loop" or auto-regulation or self-interaction), with predefined signs σ_{ij}, indicating activation or inhibition, where $1 \leq j \leq k$. For odd k, the signed sum $S = \sum_{j=1}^{k} \sigma_{ij} x_j$ cannot be zero for $x_j \in \{-1, +1\}$, so the output depends only on the regulators and not on x_i itself. The corresponding BF therefore lies in the space of all k-input effective functions (EFs) (see Figure 6.1) [81]. When k is even, however, S may be zero, in which case the output depends on the current state of x_i. IMRs thus introduce an implicit self-activation, and the corresponding BF must be interpreted as a $(k+1)$-input EF (see

Figure 6.1).

Now, consider the case in which the k regulators of x_i includes x_i. In principle, such a system corresponds to a k-input BF, but not all such BFs are effective. When k is even and the self-regulation is inhibitory, the BF becomes ineffective because the self-input does not affect the output, whereas the remaining $(k-1)$ inputs do (see Figure 6.1). For $k = 2$, even when the self-regulation is activatory, the BF is not an EF: if x_2 is regulated by both x_1 and x_2, the four sign combinations $(x_1, x_2) = ++, +-, -+$ and $--$ yield the single-input EFs x_2, x_1, x_2 and $\overline{x}_1$, respectively. Hence, no two-input EF corresponds to an IMR. In contrast, when k is odd, the signed sum cannot be zero, so any IMR, even with self-regulation, corresponds to a k-input EF, just as in the case without self-regulation.

We can further characterize the bias (the number of input combinations yielding an 'on' output) of an EF f which is equal to an IMR with k regulators: (a) for odd k, the bias is 2^{k-1}; (b) for $k = 2$, the bias is 1; (c) for even $k \geq 4$, an external set of regulators produces a bias of 2^k; (d) for even $k \geq 4$, a set of regulators that includes an activatory self-regulation produces a bias of 2^{k-1}; and (e) for even $k \geq 4$, a set of regulators that includes an inhibitory self-regulation produces a bias of 2^{k-2} (see Appendix C, Properties C.1 and C.2). This pattern implies that any EF which is equal to an IMR with more than two regulators (with or without a self-regulation) and any two-regulator IMR without a self-regulation will always have an even bias. In an empirical survey of BFs extracted from published Boolean models, Subbaroyan *et al.* [51] previously showed the predominance of odd-biased BFs. Moreover, no BF equal to an IMR with more than two regulators can be canalyzing (see Appendix C, Property C.3), standing at odds with both Kauffman's theoretical framework of canalyzation [5] and the prevalence of canalyzing functions (CFs; see Chapter 2, Section 2.5 for definition) in empirical datasets [51].

Compared to IMRs, BMRs operate on a fundamentally different principle: the sign of a regulatory interaction matters only when the corresponding input is active. In a voting analogy, in BMRs inactive inputs cast blank votes, whereas in IMRs an interaction's sign can affect the vote. Clearly, the output of a BMR is determined solely by comparing the total regulatory strength of active activators and active inhibitors. Consider a target node x_i with k external regulators x_j. There is always at least one input configuration (all $x_j = 0$) for which the signed sum vanishes; in these cases, the output depends on x_i. Thus, a BMR with k external inputs corresponds to a

$(k+1)$-input EF, for both even and odd k (see Figure 6.1).

The situation becomes more nuanced when self-regulation is present (see Appendix C, Property C.4). For $k > 1$, if the self-regulation is inhibitory and not all other regulators are inhibitory, it becomes ineffective and the BMR corresponds to a $(k-1)$-input EF. If all k (for $k > 0$) regulators (including self-regulation) are inhibitory, the rule collapses to a constant-zero BF with no effective inputs (see Figure 6.1). By contrast, if the self-regulation is activatory, the BMR corresponds to a k-input EF for all $k > 2$ (see Figure 6.1), while for $k = 1, 2$ that is the case only when all regulators are activatory.

Unlike IMRs, the bias of BMRs depend on the regulatory sign pattern. For a BMR with k external regulators, a of which are activators, the bias of the equivalent BF is (see Appendix C, Property C.5)

$$P^k_{BMR} = 2 \cdot \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k}{a}$$

yielding odd- or even-biased EFs depending on the values of a and k. Odd bias occurs precisely when $a \wedge \bar{k} = 0$ ('$\wedge$' denoting the bitwise AND operation on the binary representations of a and $\bar{k}$; see Appendix C, Property C.5, Corollary C.5.1). Notably, when k is a Mersenne number (one less than a power of two), all corresponding BMRs are odd-biased (see Appendix C, Property C.5, Corollary C.5.1). When self-regulation is present, the bias becomes (see Appendix C, Property C.6)

$$P^k_{BMR} = \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k-1}{a-\delta}$$

where $\delta = 0$ for activatory and $\delta = 1$ for inhibitory self-regulation. If the self-regulation is inhibitory, or if $k = 2$ and the regulators are not both activatory, the BF is ineffective. The bias of the reduced equivalent EF is therefore $\frac{1}{2}P^k_{BMR}$.

Unlike IMRs, BMRs can be canalyzing if several conditions are met. A BMR over k external regulators is canalyzing if and only if either all regulators are activators or all are inhibitors (see Appendix C, Property C.7). Moreover, such BFs are not only canalyzing but also NCFs (see Appendix C, Property C.5, Corollary C.5.1), specifically, they correspond to OR and AND type rules, respectively. When the set of regulators also include the target node, then the BF is canalyzing if at most one regulator is activatory or at most one is inhibitory (see Appendix C, Property C.8).

Despite their many contrasting features, IMR and BMR can occasionally correspond to the

same BF. To make this comparison, we treat BFs as equal if their truth table outputs match when we identify the values '-1' (in IMR) and '0' (in BMR) as the same logical 'off' state and '+1' as the logical 'on' state. Under this correspondence, let f_1 and f_2 denote the BFs equal to an IMR and a BMR respectively, each with k regulators and the same sign configuration of the regulators. Then f_1 and f_2 are equal only if: (a) $k = 1$ and there is a self-activation, or (b) k is even and the number of activators equals the number of inhibitors (see Appendix C, Property C.9).

These properties reveal that BFs corresponding to TMRs often lack critical logic features found in NCFs and consequently lack features typically observed in biological GRNs [51]. For instance, IMRs almost always produce even-biased, non-canalyzing rules, while BMRs permit more variation but still provide very limited scope for canalyzation. Yet the most striking shortcoming of TMRs is their mishandling of self-inhibition (i.e., self-regulation with negative weights). IMRs ignore self-inhibitions when the number of regulators is even, and BMRs always do so regardless of the parity of the number of regulators. This behavior conflicts with biological reality, where negative auto-regulation often plays a central role. For example, studies of the *Escherichia coli* transcriptional regulatory network indicate that nearly 70% of transcription factors are self-regulating with about two-thirds of those being inhibitory [181, 182].

By contrast, NCFs exhibit distinct characteristics. They are odd-biased and form a subset of canalyzing functions (CFs) and effective functions (EFs), ensuring all inputs have functional relevance. In particular, when a sign-conforming NCF (scNCF) is assigned to a node, its interactions are specified by the incoming edges, whereas as we explained above, that is not always true in the case of IMRs or BMRs. Unlike TMRs, NCFs neither require artificial self-regulations nor render any existing regulatory interactions ineffective.

6.3 Threshold majority rules exhibit higher averaged Boolean complexity and sensitivity than nested canalyzing functions

In prior work [51], the low average sensitivity of NCF and the minimum Boolean complexity of its superset RoF were proposed to explain their enrichment in empirical datasets. The definitions

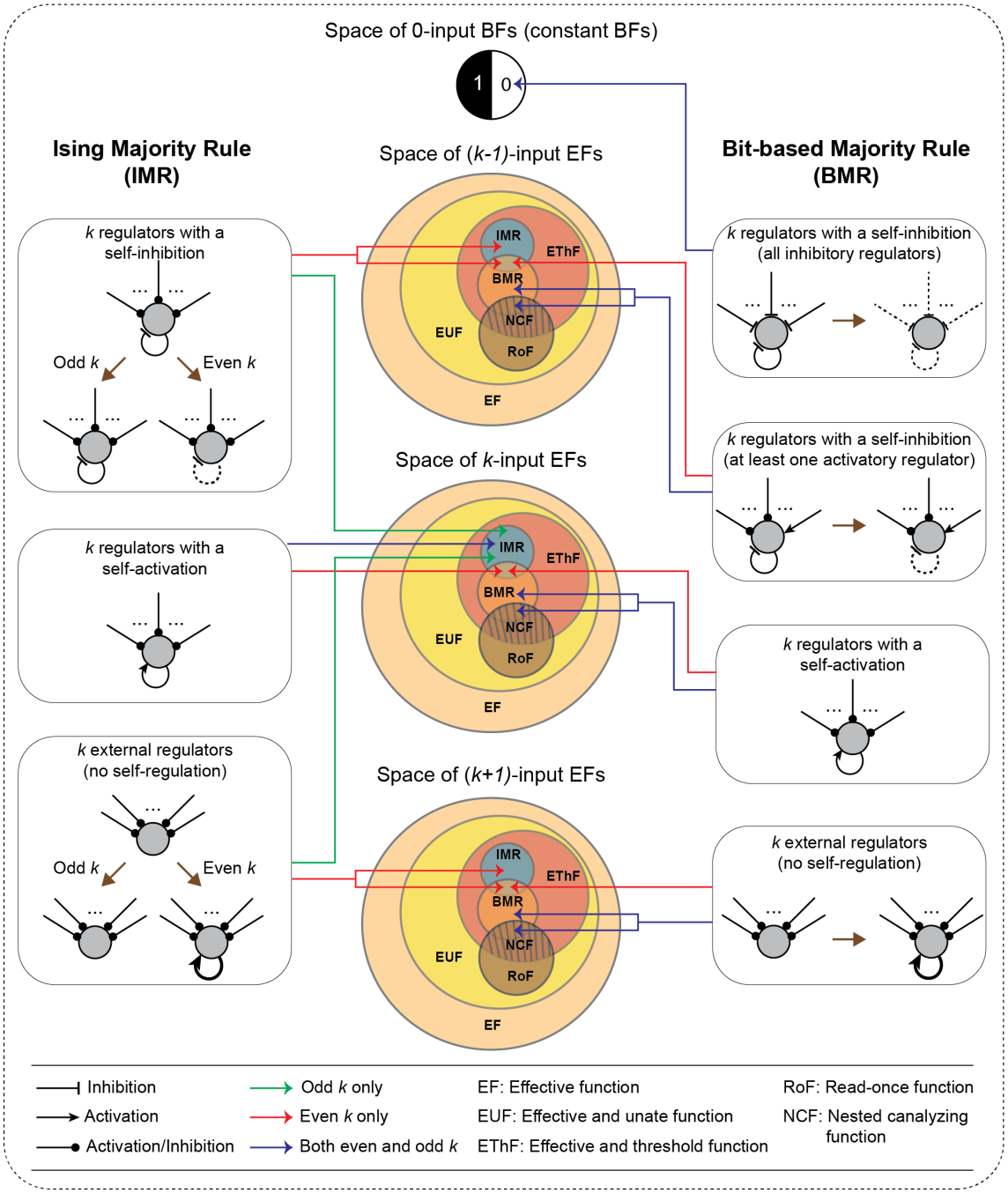


Figure 6.1: The true regulators used by a TMR can differ from the explicit ones, leading to failures in modeling of self regulation. A TMR may have k explicit regulators but when considered as a BF, the number of effective regulators may be $k-1$, k, $k+1$ or 0 as shown by the brown arrows. The target node is depicted via a grey circle. Dashed inhibitory edges (either self-regulatory or external) indicate their ineffectiveness under specific TMR conditions (topology, parity, signs). The absence of a self-regulation can lead to added self-activations as shown with thick black self-loop arrows. Inversely, even though a TMR may include an explicit self-inhibition, the corresponding EF may have no self-interaction. Colored lines connect each topology box to the corresponding regions in the Venn diagrams. Red lines indicate mappings for even k, green lines for odd k and blue lines for both even and odd values of k.

of Boolean complexity and average sensitivity of a BF are provided in Chapter 2, Section 2.6. It was shown that, for any number k of effective inputs, RoFs and therefore NCFs attain the minimum Boolean complexity (equal to k; see Chapter 2, Section 2.6 and Eq. (2.7)). Furthermore, in Feldman's $k[P]$ framework [57, 183] with odd P (P denoting the 'bias' of a BF; see Chapter 2, Section 2.1), NCFs also minimize average sensitivity [51]. Building on this, we compare Boolean complexity and average sensitivity of three EThF subtypes, NCF, IMR and BMR, across varying number of external regulators k. With k external regulators, a BMR corresponds to a $(k+1)$-input EF, while an IMR corresponds to a k-input EF when k is odd and to a $(k+1)$-input EF when k is even (see Section 6.2). Consequently, the bias ranges differ across these subtypes and that is why we compare the mean Boolean complexity and mean average sensitivity over all unique BFs in each subtype for k up to 9, referred to as the 'averaged Boolean complexity' and 'averaged sensitivity'.

From Eq. (2.7) in Chapter 2, it is evident that the Boolean complexity of a k-input NCF is exactly k. For BMRs the theoretical minimum is $(k+1)$, but Figure 6.2(a) reveals that the averaged Boolean complexity is substantially higher than this minimum for all $k > 1$, with the deviation increasing with k. Exhaustive analysis of all sign combinations up to $k = 6$ (see Appendix C, Figure C.1) shows that a BMR achieves the minimum $(k+1)$ only when all regulators are either activatory or inhibitory, precisely the cases in which a BMR is an NCF (see Appendix C, Section C.1, Corollary C.5.1). For all other sign combinations, the Boolean complexity remains significantly higher. IMRs behave similarly: whether interpreted as k- or $(k+1)$-input EFs, their averaged Boolean complexity consistently deviates from the minima for every $k > 1$, and Figure C.1 in Appendix C confirms that IMRs never attain these minimum values for $k = 2$ to 6. Interestingly, as k increases, the averaged Boolean complexities of IMRs and BMRs alternate, with BMRs generally higher for odd k and IMRs higher for even k (see Figure 6.2(a)). This observation highlights that the "simplicity of implementation" does not necessarily imply "low intrinsic logical complexity".

Concerning the averaged sensitivity, for all $k > 1$ the ordering NCF < BMR < IMR holds consistently (see Figure 6.2(b)). The averaged sensitivity of NCFs increases slowly with k, whereas sensitivities of TMRs grow much more rapidly. Figure C.2 in Appendix C shows the averaged sensitivity of all sign combinations of regulators for each subtype, for $k \leq 6$. This figure shows that for given signs of the regulators, the NCF class can produce multiple distinct average sensi-

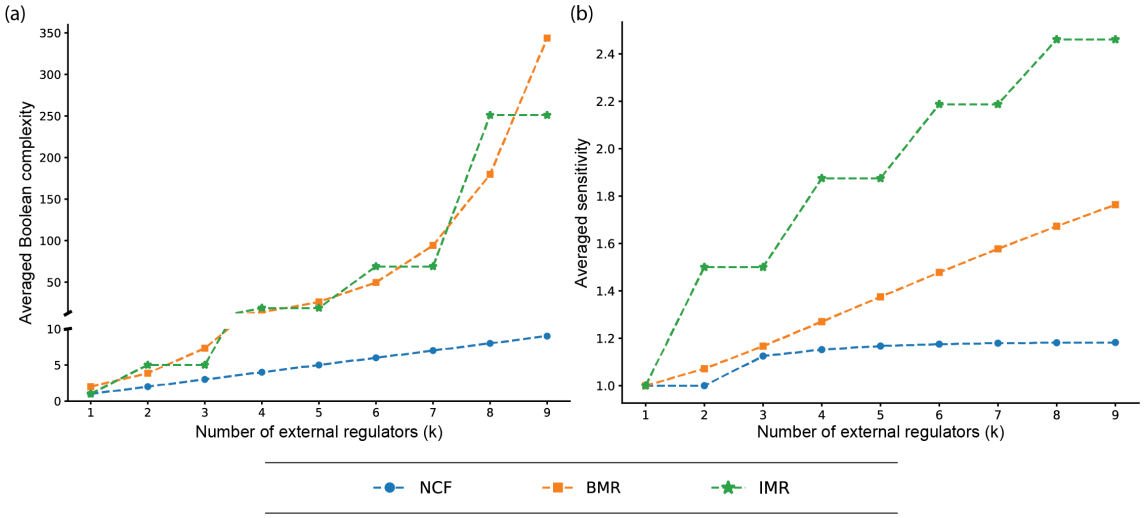


Figure 6.2: TMRs have high averaged Boolean complexity and averaged sensitivity compared to NCFs. The mean of the **(a)** Boolean complexities, and **(b)** average sensitivities, respectively, of all unique NCFs, BMRs and IMRs as a function of the number k of external regulators of a target node. NCFs attain the minimal averaged Boolean complexity (equal to k), whereas both TMR variants exhibit substantially higher values, with deviations increasing with k. For the averaged sensitivity, a consistent ordering NCF < BMR < IMR is observed for all $k > 1$. In part **(a)**, a break has been introduced in the y-axis to accommodate the large gap between the values for NCFs and those for the TMRs.

tivity values because different NCFs with the same sign pattern can have different bias values. In contrast, TMRs exhibit only a single average sensitivity value per sign combination as there is a unique TMR for a given sign combination of the regulators (see Appendix C, Properties C.1, C.2, C.5 and C.6). For some sign patterns, certain BMRs have lower averaged sensitivity than several NCFs. Nevertheless, when all regulator signs are not identical, there always exist NCFs with lower average sensitivity than the corresponding BMR. On the other hand, IMRs consistently show the highest average sensitivity across all sign combinations and in fact attain the maximum average sensitivity among all k-input BFs in the UF family (see Appendix C, Property C.10).

These findings show that, despite their tempting simplicity, TMRs often yield logic that is substantially more complex and more sensitive than NCFs. On average, both IMRs and BMRs exhibit higher Boolean complexity and higher average sensitivity than NCFs. Moreover, for any fixed sign combination of a node's regulators, IMRs achieve the highest average sensitivity among all sign-conforming UFs. Because sensitivity reflects robustness to input perturbations [51, 77], and biological systems are inherently noisy, this renders IMRs the least robust choice.

6.4 Enrichment analysis of threshold majority rules in the empirical dataset

6.4.1 Empirical dataset

For the quantification and enrichment analysis of EThFs and their subtypes, we used the Biodivine Boolean Models (BBM) dataset [52], which compiles 245 reconstructed logical models of biological systems from CellCollective [49], GINsim [184], BioModels [185], the COVID-19 Disease Map [186], and manual curation from literature. The dataset includes models encompassing a broad range of biological processes such as development, signaling, cancer, cell cycle, metabolism, and immune response. Detailed information on each model in BBM dataset and the corresponding source publication is available at `https://github.com/sybila/biodivine-boolean-models`. From those models, we extracted 5990 BFs with up to 10 inputs as described in [187]. Thereafter, we ensured all BFs were effective (EF) by removing any ineffective inputs, and then limited our analysis to BFs with at most 9 effective inputs, since feasible ways to count ThFs currently exist only up to this size [179]. This resulted in a final empirical dataset of 5953 BFs. The distribution of these BFs based on input size and their classification into EThFs and their various subtypes is presented in Table C.2 in Appendix C.

6.4.2 Relative enrichment and statistical significance

To test whether EThFs within all BFs or different subtypes of EThFs within EThFs (e.g., NCF, IMR, BMR) are overrepresented or underrepresented in an empirical datatset, we adopt a relative enrichment framework as presented in Subbaroyan *et al.* [51]. Let T denote a class of BFs (e.g., T = EThF) and $T_s \subset T$ one of its subtypes (e.g., T_s = IMR). Working at a fixed number of effective inputs k, let us define:

- $f_{s,1}$ = Fraction of k-input BFs in the dataset that lie in T_s.
- f_1 = Fraction of k-input BFs in the dataset that lie in T.
- $f_{s,0}$ = Fraction of all possible k-input BFs that lie in T_s.
- f_0 = Fraction of all possible k-input BFs that lie in T.

The relative enrichment of T_s within T is then

$$E_R = \frac{f_{s,1}/f_1}{f_{s,0}/f_0}. \tag{6.5}$$

An E_R value > 1 indicates that the subtype T_s is proportionally more common within T in the empirical dataset than it is in the universe of all k-input BFs. On the other hand E_R value < 1 is indication of underrepresentation of that subtype T_s within T in the empirical dataset. We also accompany each E_R estimate with a formal significance test, as described next.

Under the null hypothesis H_0, we assume that although the empirical dataset may be enriched overall for BFs in T, the selection among members of T is random, that is the probability of elements in T_s is not greater than that of other elements in T. Equivalently, given that M_T of the empirical dataset's M total BFs lie in T, each of those M_T slots is equally likely to be occupied by any member of T, irrespective of whether it belongs or not to T_s. Let, $f_R = |T_s|/|T|$ be the background probability that a randomly chosen BF from T belongs to T_s. Then, under H_0, the number m of T_s members observed among the M_T selected BFs follows a binomial distribution:

$$\Pr(m \mid M_T, f_R) = \binom{M_T}{m} (f_R)^m (1 - f_R)^{M_T - m}$$

The desired one-sided p-value for the observed count m_{obs} is:

$$p = \sum_{m=m_{obs}}^{M_T} \binom{M_T}{m} (f_R)^m (1 - f_R)^{M_T - m} \tag{6.6}$$

A small p-value indicates that the observed enrichment of T_s within T in the empirical dataset is unlikely under the statistics given by H_0. Conversely, if T_s is not enriched within T, or is even underrepresented within T, the calculated p-value will no longer be small or it may even approach 1, signaling that the observed frequency is consistent with H_0 or is associated with depletion.

6.4.3 Threshold majority rules are underrepresented in the BBM empirical dataset

With the mappings described in Section 6.2, we now test whether the BFs that are equal to certain TMRs are preponderant in the BBM empirical dataset of regulatory logic rules (see Section 6.4.1).

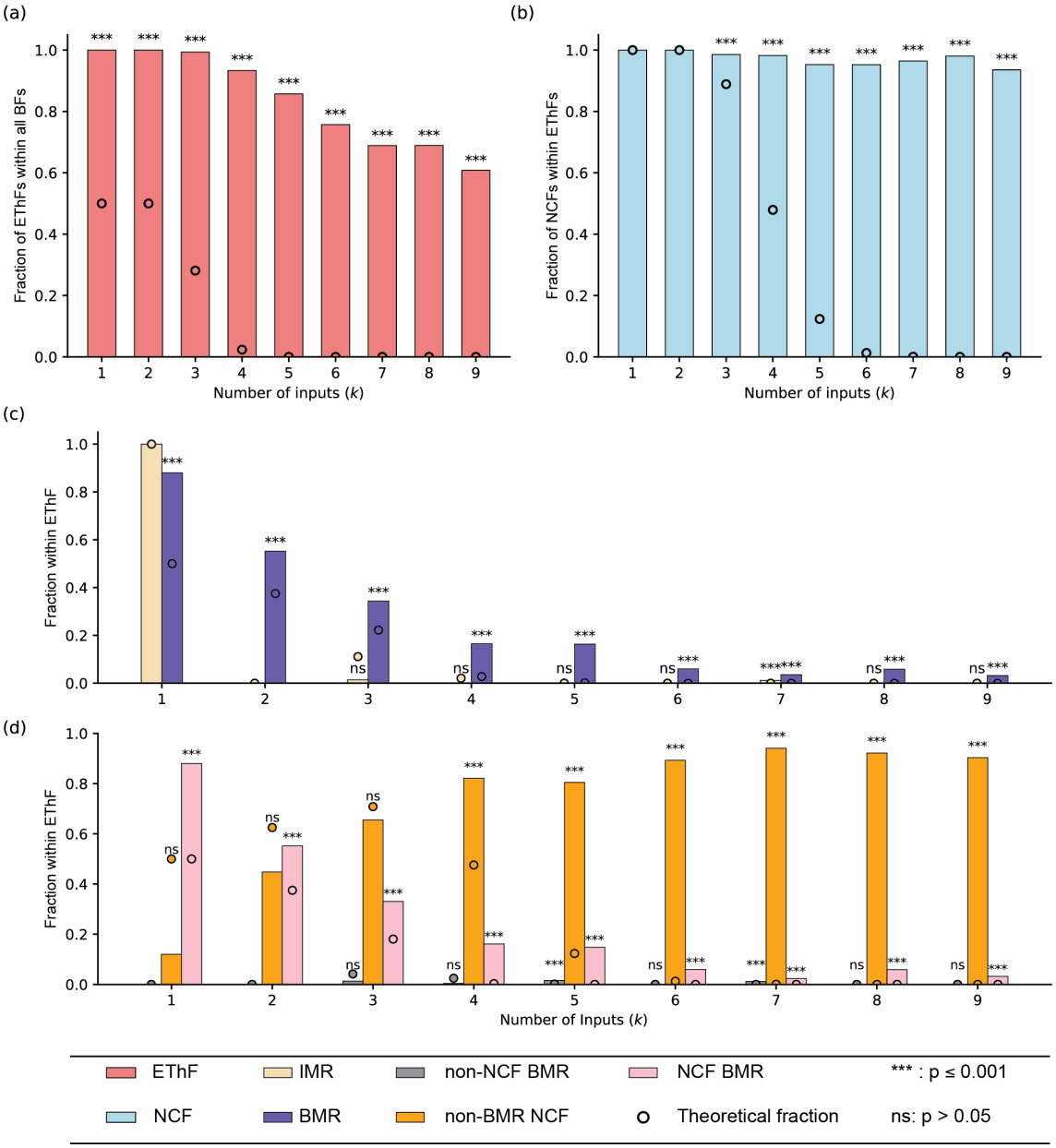


Figure 6.3: Enrichment of EThFs and NCFs, and depletion of IMRs and non-NCF BMRs, in the BBM dataset. Fractions of different subtypes as a function of the number of inputs k: **(a)** EThFs within all BFs, **(b)** NCFs within EThFs, **(c)** BFs equal to some IMR or BMR within EThFs, and **(d)** non-NCF BMRs, NCF BMRs, and non-BMR NCFs within EThFs. Subplots **(a)** and **(b)** show consistent enrichment of EThFs and NCFs, respectively. Subplot **(c)** shows that IMRs are generally absent or strongly depleted relative to theoretical expectations for $k > 2$, while BMRs appear enriched overall. Subplot **(d)** demonstrates that the apparent enrichment of BMRs is largely driven by those that are also NCFs; non-NCF BMRs are underrepresented, whereas non-BMR NCFs are strongly enriched for $k > 3$. In each subplot, dots denote theoretically expected fractions, and colored bars denote empirical fractions from the BBM dataset.

Similar to the empirical dataset, we ensure that BFs that are equal to some IMR or BMR are effective. Subsequently, we categorize these BFs into different classes based on their number of effective regulators k. As discussed in Section 6.1, the IMR, BMR and NCF classes each represent a specific subtype of ThFs. Among these, NCFs are always effective. Therefore, the NCFs and the effective BFs that are equal to some IMR or BMR constitute specific subtypes of EThFs.

For the BBM dataset, we compute the fraction of EThFs within all BFs for each input size k up to $k = 9$. EThFs are consistently enriched across all k, as shown in Figure 6.3(a). A previous study [51] showed strong enrichment of NCFs in empirical regulatory logic rules, motivating us to examine their occurrence within EThFs in the BBM dataset. The corresponding fractions for different k are shown in Figure 6.3(b). The associated significance values (see Section 6.4.2) show that NCFs remain strongly enriched within EThFs for all $k > 2$. For $k = 1$ and 2, the theoretical spaces of NCFs and EThFs coincide, but as k increases the proportion of NCFs in the theoretical EThF space decreases sharply and becomes negligible for large k (see Appendix C, Table C.1).

We next evaluate the same quantities for IMR and BMR subtypes within EThFs (see Figure 6.3(c)). For $k = 1$, EThF, NCF, and IMR coincide, so both the empirical and theoretical fractions of IMRs are 1. For $k = 2$, no EF corresponds to an IMR, so the theoretical fraction itself is 0. However, for $k > 2$, BFs equal to some IMR form a proper subset of EThFs that is disjoint from NCFs (see Appendix C, Property C.3). As shown in Figure 6.3(c) and Table C.2 in Appendix C, the empirical fraction of IMRs within EThFs is generally 0 or very low for $k > 2$, and in nearly all cases falls below theoretical expectations, indicating depletion rather than enrichment (see Appendix C, Table C.3).

For all $k \leq 9$, BFs of the BMR class appear relatively enriched within EThFs in the dataset (see Figure 6.3(c); Appendix C, Table C.3). However, because BMRs and NCFs overlap for all k (see Appendix C, Section C.1, Corollary C.5.1), it is best to assess whether this enrichment is driven by BMRs that are also NCFs (NCF BMRs) or by those that are not (non-NCF BMRs). For each k, we therefore partition BMRs into NCF BMRs and non-NCF BMRs. For consistency, we also consider, for each k, the subset of NCFs that are not equal to any BMR, denoting them as non-BMR NCFs. For $k = 1$ and 2, all BMRs are NCFs (see Appendix C, Table C.1), whereas for $k \geq 3$ non-NCF BMRs appear. For each k, we compute the fractions of these three subtypes within EThFs along with their p-values. As shown in Figure 6.3(d), the observed enrichment of

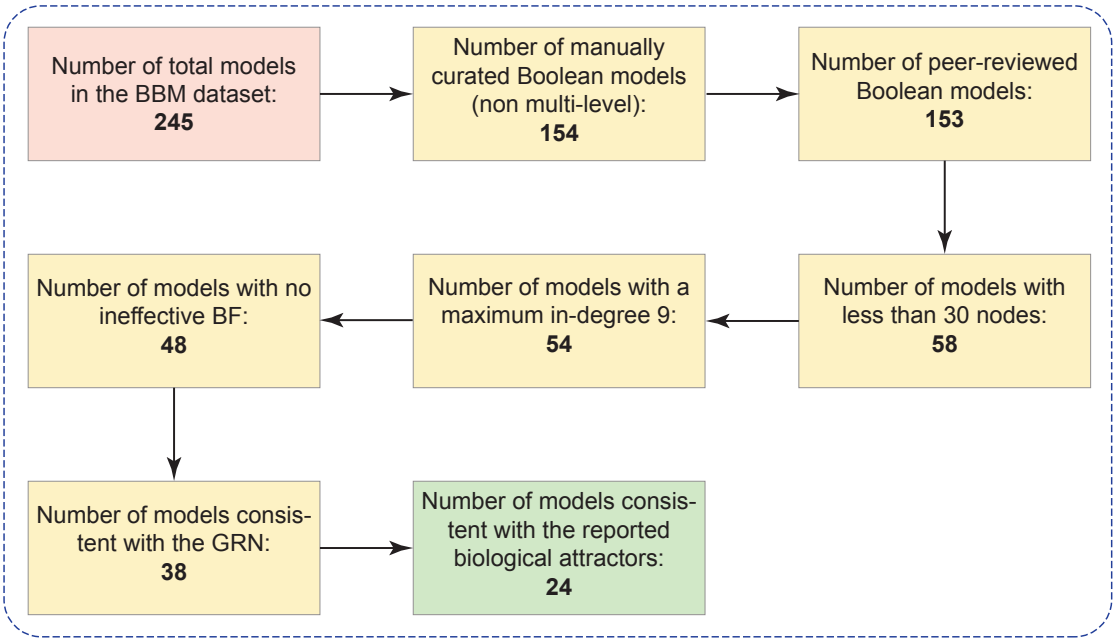


Figure 6.4: Flowchart of the multi-step filtration pipeline applied to the models in the BBM dataset. Starting with 245 models from the BBM dataset, a series of filtration steps (as shown in the flowchart) resulted in a final set of 24 models selected for our detailed analysis of attractor recovery and associated basin sizes.

BMRs is largely driven by NCF BMRs. The empirical fraction of non-NCF BMRs within EThFs is typically lower than the corresponding theoretical expectation, whereas non-BMR NCFs show strong and statistically significant enrichment for all $k > 3$. The associated p-values reveal a consistent hierarchy for $k > 3$: non-BMR NCF < NCF BMR < non-NCF BMR (see Appendix C, Table C.4).

These findings show that IMRs and BMRs are not only underrepresented in the empirical dataset of biological regulatory logic rules but are in fact depleted compared to what would be expected by random selection. The empirical dataset compiles the best available logic rules, each carefully assigned by experts around the world. While individual choices may reflect particular biases, these rules collectively can be considered the most reliable portrait of biological regulatory logic. One may thus speculate that TMRs are not simply rare but might even be fundamentally incompatible with the logic underlying real biological networks.

6.5 Choosing threshold majority rules over nested canalyzing functions impairs recovery of the dynamical outcomes in biological Boolean networks

6.5.1 Imposing structural and dynamical constraints on biological models

We assembled a collection of manually curated BN models to test if simple TMRs can replicate key dynamical behaviors. Our procedure comprises a rigorous multi-step selection as summarized in Figure 6.4. Starting from the BBM dataset [52] of 245 models, we excluded those lacking peer-reviewed references, those using multi-level variables and those that were not manually assigned. To remain compatible with the exhaustive analyses performed using the R package BoolNet [123], we retained models with fewer than 30 nodes. Furthermore, to ensure consistency with the other analyses, we removed any models that contained ineffective BFs or BFs with more than 9 inputs. Each remaining model was then verified against its original publication to confirm GRN topology and regulatory sign consistency. This is essential for subsequent replacement of original BFs with corresponding TMRs. Finally, we restricted to models with explicitly described biological attractors in the publication, and wherein the biological attractors were recovered by simulating the model. This pipeline yielded 24 models (see Appendix C, Table C.5), where 20 contained only NCFs, each of the remaining four models having at least one non-NCF BF.

6.5.2 An attractor-based scoring framework to compare a published model to an alternative model

We would like to evaluate how well an alternative model formed by replacing each BF of the published model by a different rule (specifically, by the corresponding TMR variants) reproduces the given biological outcomes. In particular, we ask: how accurately does the alternative model recover the literature-reported biological attractors, and to what extent does it preserve the associated basin size distributions? To address this, we propose two metrics: *Attractor Recovery Score* and *Distance of basin fraction distribution from the gold standard*.

For any published model, let A_{bio}, A_{model} and A_{alt} denote the set of attractors considered as biologically relevant in the publication (hereafter, referred to as the biological attractors), the set of actual attractors produced by the published model, and the set of attractors of the alternative model, respectively. By construction (see Section 6.5.1), the considered published models always recover the biological attractors, so that $A_{bio} \subseteq A_{model}$. To quantify attractor recovery performance, we introduce the attractor recovery score (ARS), defined as the ratio of two Jaccard indices. First, set

$$\text{Jac}_{model} = \frac{|A_{bio} \cap A_{model}|}{|A_{bio} \cup A_{model}|} = \frac{|A_{bio}|}{|A_{model}|}$$

where the equality follows from $A_{bio} \subseteq A_{model}$. Next define,

$$\text{Jac}_{alt} = \frac{|A_{bio} \cap A_{alt}|}{|A_{bio} \cup A_{alt}|}$$

The attractor recovery score for the alternative model is then

$$\text{ARS}_{alt} = \frac{\text{Jac}_{alt}}{\text{Jac}_{model}} \tag{6.7}$$

A score below one indicates degraded performance relative to the published model, either because the alternative model fails to recover all biological attractors, produces many spurious attractors, or both. A score near zero indicates almost no recovery of the biological attractors or so many attractors that any overlap with biological attractors is random. A score of one or higher indicates comparable or better performance relative to the published model, provided the alternative model recovers all biological attractors.

In synchronous BN dynamics, the entire state space partitions into disjoint basins of attraction (BoA), each associated with a particular attractor. The BoA for a given attractor comprises all initial states from which the system eventually converges to that attractor. Thus, by normalizing each basin's size by the total number of states, one obtains a probability distribution over attractors, where each basin fraction corresponds to the likelihood of the system converging to that attractor from a randomly chosen initial state. To quantify how well a model reproduces the expected basin size distribution, we employ the Jensen–Shannon (JS) distance (see Chapter 4, Section 4.2) [133, 134]. In an ideal scenario, we anticipate that a good model would recover only the biological

attractors, with no spurious ones. However, the published models quite often exhibit additional artefactual attractors, i.e. A_{bio} is a proper subset of A_{model}. Since, the true "gold standard" basin fractions of the biological attractors are unknown a priori, we take the published model's basin fractions for the attractors in A_{bio}, renormalize them over only the biological attractors, and use this renormalized distribution for our reference "gold standard". Formally, let

$$\begin{aligned} A_{bio} &= \{b_i : 1 \leq i \leq l\}, \\ A_{model} &= \{p_j : 1 \leq j \leq m\}, \\ A_{alt} &= \{a_k : 1 \leq k \leq n\} \end{aligned}$$

Now, clearly $l \leq m$ as $A_{bio} \subseteq A_{model}$. Let ω_{p_j} denote the published model's basin fraction of attractor p_j for all $1 \leq j \leq m$. Without loss of generality, assume $b_i = p_i$ for $1 \leq i \leq l$. The renormalized ("gold-standard") basin fraction for each biological attractor is then given by:

$$\omega_{b_i} = \frac{\omega_{p_i}}{\sum_{i=1}^{l} \omega_{p_i}} \quad (1 \leq i \leq l).$$

We now form two probability vectors of length m:

$$\begin{aligned} P_{gold,model} &= (\omega_{b_1}, \ldots, \omega_{b_l}, \underbrace{0, \ldots, 0}_{(m-l)}) \text{ and} \\ P_{model} &= (\omega_{p_1}, \omega_{p_2}, \ldots, \omega_{p_m}) \end{aligned}$$

where the trailing zeros in $P_{gold,model}$ accounts for the $(m - l)$ artefactual attractors of the published model. The JS distance D_{model} between these two vectors quantifies the published model's deviation from the ideal expectation. $D_{model} = 0$ arises only when the published model does not generate any artefactual attractors.

For the alternative model, let $r = |A_{bio} \cap A_{alt}|$ and, again without loss of generality, assume $b_i = a_i$ for $1 \leq i \leq r$. If $l > r$, then the alternative model fails to recover $(l - r)$ biological attractors and if $n > r$ then the alternative model generates $(n - r)$ spurious attractors. Therefore,

we construct two probability vectors of length $l + n - r$:

$$P_{gold,alt} = (\omega_{b_1}, \ldots, \omega_{b_l}, \underbrace{0, \ldots, 0}_{(n-r)}) \text{ and}$$
$$P_{alt} = (\omega_{a_1}, \ldots, \omega_{a_r}, \underbrace{0, \ldots, 0}_{(l-r)}, \omega_{a_{r+1}}, \ldots, \omega_{a_n})$$

where w_{a_k} denotes the alternative model's basin fraction for attractor a_k, $1 \leq k \leq n$. The JS distance D_{alt} between these vectors is the alternative model's deviation from the ideal expectation. By comparing D_{model} and D_{alt}, both of which lie in $[0, 1]$, we can assess relative performance of the alternative model. A lower distance corresponds to better performance.

We now examine the dynamical implications of using TMRs, asking whether they can substitute for the BFs of published BN models while preserving key dynamical behaviors. From the 245 models in the BBM dataset [52], our multi-step filtration pipeline yielded 24 high-confidence models satisfying the required structural and dynamical criteria (see Section 6.5.1). We then ask how well the corresponding alternative models, obtained by replacing every BF with a TMR consistent with the activatory or inhibitory sign of each interaction, recover the biological attractors and maintain their basin size distributions. To quantify this, we compute the scores ARS_{alt}, D_{model} and D_{alt} (see Section 6.5.2). For each of the 24 models, we generate two sign-conforming alternative variants, one using IMRs and one using BMRs, and simulate all three models to obtain their complete attractor sets and associated basins of attraction.

6.5.3 Threshold majority rules perform poorly in recovering the biologically meaningful attractors

The attractors in biological BNs correspond to key biological outcomes such as cell types, apoptosis, proliferation, or cell-cycle states, so a useful model must be able to reproduce them. Figure 6.5 displays the ARS for the two alternative models obtained by using IMRs and BMRs at each node for each of the 24 networks. For most networks, both IMR- and BMR-based models yield scores near or equal to zero, demonstrating their inability to capture the biologically desired dynamical behavior. For a couple of small networks, IMRs recover all biological attractors and attains a score of exactly one. For network 7 (network numbers represent the identifier of each network in the BBM dataset), the extremely small state space may allow even random rules to recover the bio-

logical attractors by chance. For network 110, which has nine nodes, six have in-degree one and all BFs in the published model are NCFs. Since IMRs coincide with NCFs for $k = 1$, replacing the rules with IMRs preserves two-thirds of the original BFs. As network size and connectivity increase, IMR performance rapidly deteriorates.

6.5.4 The basin fractions of biologically meaningful attractors when using threshold majority rules are far from the desired target

Next we evaluate how closely the basin fraction distribution of the published models and their corresponding IMR and BMR variants align with a reconstructed "gold standard" (see Section 6.5.2). This comparison is quantified using the JS distance, which measures the divergence between the basin size distribution of a model and the idealized distribution derived from the published model, considering only the biological attractors. The JS distance ranges from 0 to 1, where a value of 0 indicates perfect alignment with the gold standard distribution, and a value close to 1 indicates a very large deviation thereof. As can be seen in Figure 6.6, several published models themselves exhibit considerable distances from the gold standard, largely due to the presence of spurious attractors whose basins occupy a substantial portion of the state space. This indicates that some published models are unsatisfactory, a point that we have emphasized in one of our previous works [64]. Nevertheless, models constructed using TMRs consistently show poorer alignment with the gold standard compared to the published models (i.e., $D_{TMR} > D_{model}$), as indicated by their uniformly high JS distances across most cases. The sole exception is model 88 with IMR rules, which has a network size of only 6 nodes. Overall, these findings underscore a key limitation of using TMRs for modeling biological networks in general. Although they represent a very simple modeling choice, such simplicity generally comes at the cost of failing to preserve essential dynamical features of biological systems.

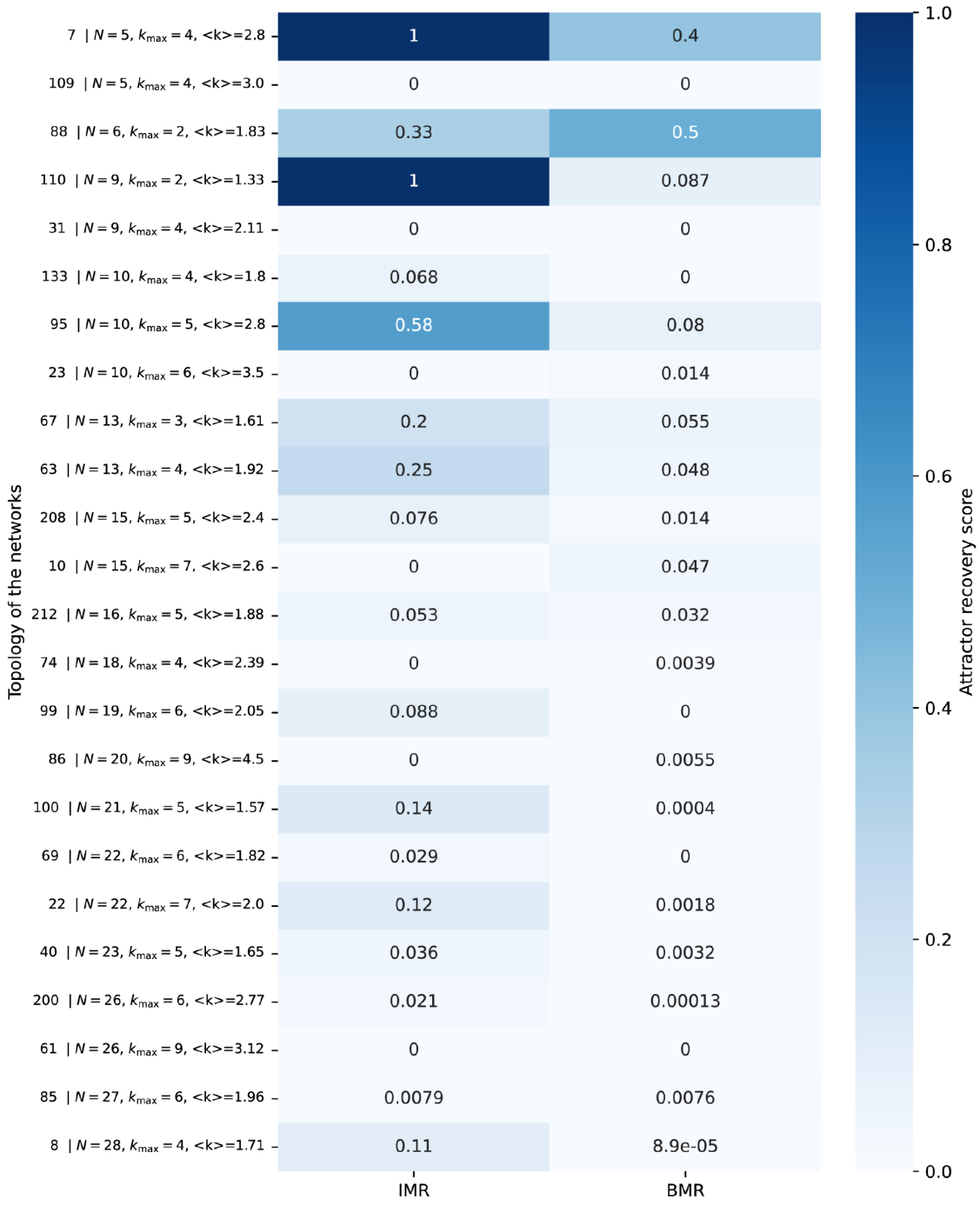


Figure 6.5: TMR-based models generally fail to recover biological attractors. This heatmap displays the attractor recovery scores for the alternative models obtained by replacing all BFs in 24 published BBM models with sign-conforming IMRs or BMRs. Each row corresponds to a specific network from the BBM dataset. On the far left, the number indicates the network ID as assigned in the BBM dataset. The next columns list the number of nodes (N), the maximum in-degree ($k_{\max}$), and the average in-degree ($\langle k \rangle$) of the network. A score near 0 indicates poor recovery of biological attractors, while higher scores indicate better performance. For most networks, both IMR- and BMR-based models show scores near 0.

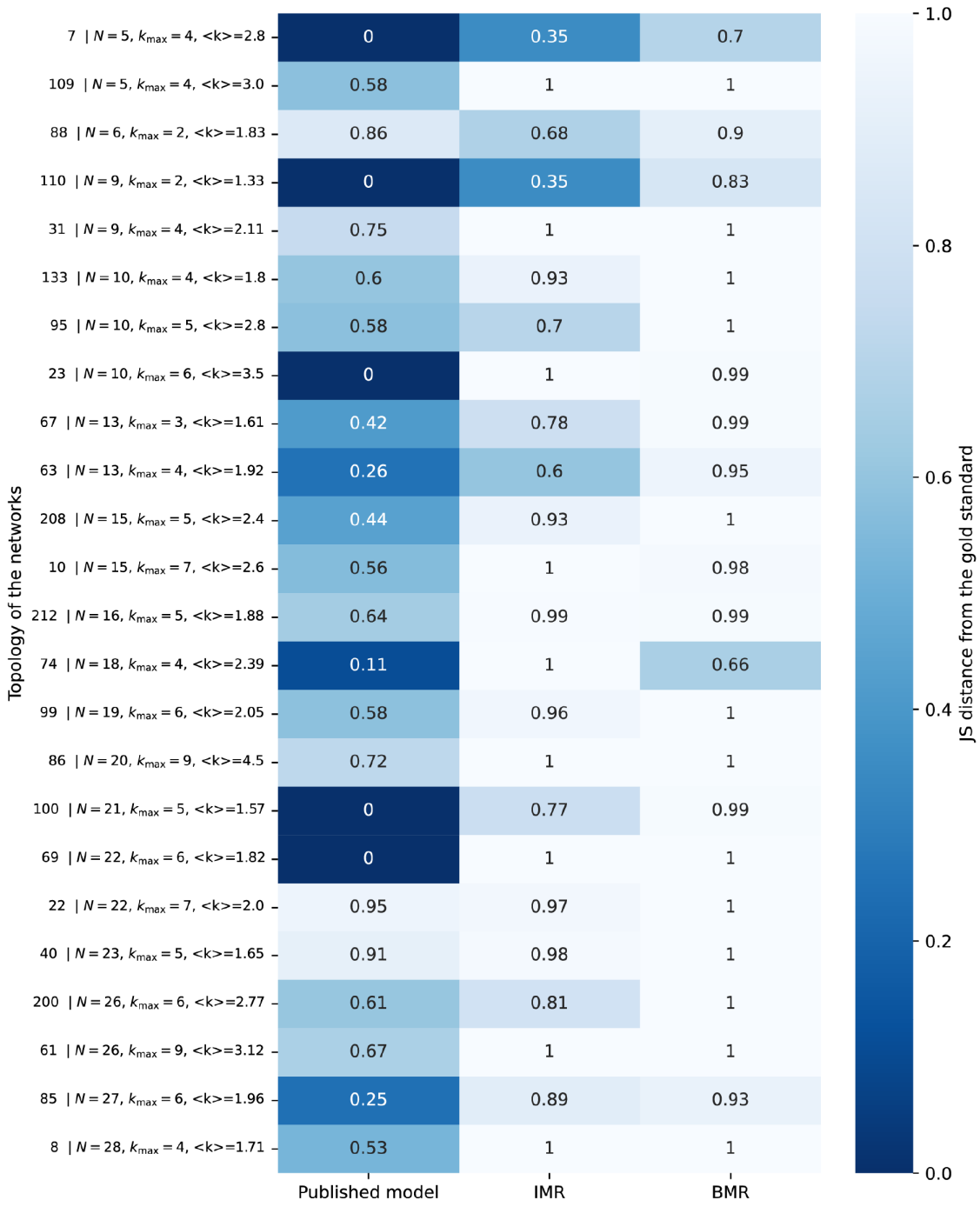


Figure 6.6: JS distance from gold standard reveals poor preservation of basin size distributions of biological attractors by TMR-based models. The heatmap shows the Jensen-Shannon (JS) distance between the basin fraction distributions of the 24 published BBM models and their IMR and BMR variants relative to the corresponding gold standards. Each row corresponds to a specific network from the BBM dataset. On the far left, the number indicates the network ID as assigned in the BBM dataset. The next columns list the number of nodes (N), the maximum in-degree ($k_{\max}$), and the average in-degree ($\langle k \rangle$) of the network. A JS distance of 0 indicates perfect agreement with the gold standard, while values closer to 1 indicate increasingly divergent basin size distributions. Across most networks, IMR- and BMR-based models typically exhibit high JS distances, indicating poor preservation of the expected basin size distribution of the biological attractors.

6.5.5 Substituting non-nested canalyzing functions with suitable nested canalyzing functions in published models can yield superior dynamical outcomes.

It has been shown that NCFs are highly enriched in the empirical dataset of BFs (see Section 6.4.3) [51, 187]. With no surprise, only four of the 24 models (IDs 61, 69, 95, and 212), contain at least one non-NCF. This prompts two questions: whether such non-NCFs can be replaced by appropriate scNCFs while preserving all biological attractors, and whether the resulting models can match or outperform the published ones in terms of ARS and JS distance. Because a given sign pattern can admit multiple scNCFs, we exhaustively examined all valid replacements in these four networks while enforcing recovery of the biological attractors (see Appendix C, Section C.2) [64, 148]. The number of candidate models satisfying these constraints was 686244, 96, 2116 and 48 for network IDs 61, 69, 95 and 212, respectively. For network 61, we randomly sampled 5000 candidate models and exhaustively simulated them to compute ARS_{repl} and D_{repl} (see Section 6.5.2), finding that for 1640 models, $\mathrm{ARS}_{repl} \geq 1$ and 4064 models matched or improved upon D_{model}. For the remaining three networks also, subsets of candidate models satisfy $\mathrm{ARS}_{repl} \geq 1$ and $D_{repl} \leq D_{model}$. In nearly all cases, these NCF-replaced models outperformed the corresponding TMR-based alternatives. A detailed summary of model counts is provided in Appendix C, Table C.6. Taken together, these results reinforce that using TMRs instead of NCFs leads to serious failures in preserving biological network dynamics. Unlike NCFs, where multiple sign-conforming BFs allow flexibility to preserve known biological attractors, TMRs uniquely fix the logic once the network topology and interaction signs are specified. As a result, TMR-based models often fail to recover biological attractors and generate numerous spurious ones, and their basin-size distributions rarely align with the desired ones, making TMRs generally inappropriate for modeling biological regulatory networks.

6.6 Comparison of dynamical properties in random networks when using different subtypes of threshold functions

For the study presented in this section, we consider two different network topology ensembles of random Boolean networks (RBNs): (i) the P-P ensemble, in which directed edges are assigned randomly between node pairs while maintaining a fixed average number of edges λ per node, and (ii) the R-R ensemble, in which both the in-degree and out-degree of each node are fixed to a constant value (see Chapter 4, Section 4.1) [77].

We generate and analyze ensembles of RBNs under both topological configurations: P-P and R-R. The analysis setup is as follows:

(i) We consider networks with $N = 12$ and $N = 16$ nodes.

(ii) The average in-degree (λ) or fixed in-degree (k_{in}) considered takes values $2, 3, 4$ and 5.

(iii) For networks with $N = 12$, we generate 10000 RBNs per setting, while for $N = 16$, we generate 1000 RBNs per setting.

(iv) In each case, for every RBN, we construct three different models: one in which all BFs are scNCFs, one in which all BFs are BMRs, and one in which all BFs are IMRs.

(v) Under this setup, we compute selected stability measures for each Boolean model, as described in Chapter 3.

6.6.1 Ising majority rules drive network dynamics toward a chaotic regime

We analyzed how the choice between the three different ThF subtypes (scNCF, BMR and IMR) can affect the dynamical stability of RBNs. For that, we first consider 12-node RBNs with varying mean connectivity under the P-P topology. We find that the scNCF models consistently yield the fewest attractors, whereas BMR and IMR models produce an order of magnitude more, with BMR exhibiting the highest average attractor count (see Figure 6.7). This proliferation of attractors in BMR models has been noted previously [188,189], and it was also found that there is an abundance of fixed point attractors for these models. Next, we consider the bushiness and convergence rate

stability measures [79] of the STGs associated with each type of model, wherein high values correspond to stable dynamics (see Chapter 3, Section 3.4.1). Models using IMRs exhibit the lowest average G-density, whereas models using scNCFs and BMRs both yield distributions with similarly high mean values. Notably, BMR shows a much wider range of G-density, whereas the G-densities for all scNCF models are well clustered at higher levels. In contrast, the measure of convergence rate λ_{GoE} has its highest values for the BMR models, corroborating an earlier finding that BMR rules induce very short transients [189].

Turning to damage spreading metrics, both Derrida coefficient (δ) and final Hamming distance (h^{∞}) (see Chapter 3, Section 3.1) [60] place the mean of the distributions for scNCF and BMR models very close to the critical boundary at 1 ($\delta = 1$ and $h^{\infty} = 1$, respectively), whereas the distributions for IMR models clearly fall in the chaotic regime. Notably, although the average δ for BMR is closer to the critical line than that of scNCFs, BMRs exhibit a much wider spread of δ values, reflecting a greater sensitivity to the underlying network topology. Finally, we consider the measure fragility (ϕ) [60], the phase-insensitive analogue of h^{∞} (see Chapter 3, Section 3.1). We find that scNCFs yield the lowest mean ϕ, BMRs are intermediate, and IMRs have the highest mean, well above the critical threshold at $\phi = 1$. The means for scNCF and BMR models both lie in the ordered regime. All these ordering patterns for each observable remain consistent as one varies the average connectivity (λ) over the values $2, 3, 4$ and 5.

The qualitative patterns observed for 12-node networks persist when we analyze larger networks with 16-nodes (see Appendix C, Figure C.3). In this case, the differences in the distributions of number of attractors and G-density become even more noticeable for networks with higher connectivity. Across all scenarios, IMRs consistently perform the worst, while scNCFs and BMRs behave quite similar in most cases. These trends also hold when we switch from the P-P to the R-R topology (see Appendix C, Figures C.4 and C.5). One notable nuance appears in the distribution of the number of attractors. Under the P-P topology, BMRs produce the highest average attractor count for every tested value of the average connectivity. Under the R-R topology, where the in- and out-degree for each node is fixed ($k_{in} = k_{out}$), the mean attractor count under BMR rules is maximum when k_{in} is odd. When k_{in} is even, IMR models take the lead in the average attractor count. In all cases, scNCF models maintain the fewest attractors. Lastly, the behaviour of all other stability measures under the R-R topology matches closely with that seen under P-P.

These results illustrate why we use multiple stability metrics: different measures can change the stability ranking of rule types. Nevertheless, by nearly all criteria, IMRs induce dynamical instability to the system while scNCFs and BMRs often tie and drive the dynamics more toward the stable regime. For some measures, such as number of attractors and ϕ, scNCFs clearly do best, and for λ_{GoE}, BMRs stand out. Even though BMRs appear to drive the system toward stable dynamics, our primary concern in the context of modeling real biological networks remains the accurate recovery of known biological attractors (as discussed in Section 6.5.3). If a logic rule cannot achieve this fundamental requirement, its superior stability is of little practical utility. Finally, it is important to recognize that, even though we start with the same nominal network wiring, applying different types of rules can alter the effective architecture. Under scNCFs, the working topology matches the original exactly. However, with TMRs, a node that originally has k regulators may functionally behave as if it has $k-1, k, k+1$ or even 0 inputs, depending on the parity of k, the presence and signs of self-regulations, and the specific type of TMR used.

In this context, it is important to note that, in non-biological settings, such as formal analyses of BN dynamical properties in RBNs, TMRs can indeed be a practical option since their simplicity allows one to bypass the daunting task of model selection and enables efficient simulations for very large networks. In modeling biological networks, however, one must raise the question that whether this simplicity incurs unacceptable compromises in biological realism. If so, their widespread use warrants reconsideration.

6.7 Discussion

Boolean modeling is, of course, a substantial approximation of biological reality, yet even within this framework selecting the most appropriate option is far from easy. To assess the suitability of TMRs for modeling biological regulatory networks, this chapter presents a systematic comparative analysis of TMRs and NCFs across structural, empirical, and dynamical dimensions. We summarize our findings as follows. First, IMRs tend to produce even-biased BFs, and both TMR variants permit little scope for canalyzation. Moreover, TMRs mishandle self-inhibition: BMRs ignore it entirely, while IMRs do so whenever the number of regulators is even, despite the known importance of negative autoregulation in GRNs. Second, for any given number of regulators, NCFs exhibit lower averaged Boolean complexity and sensitivity than either TMR variant. Third, IMRs

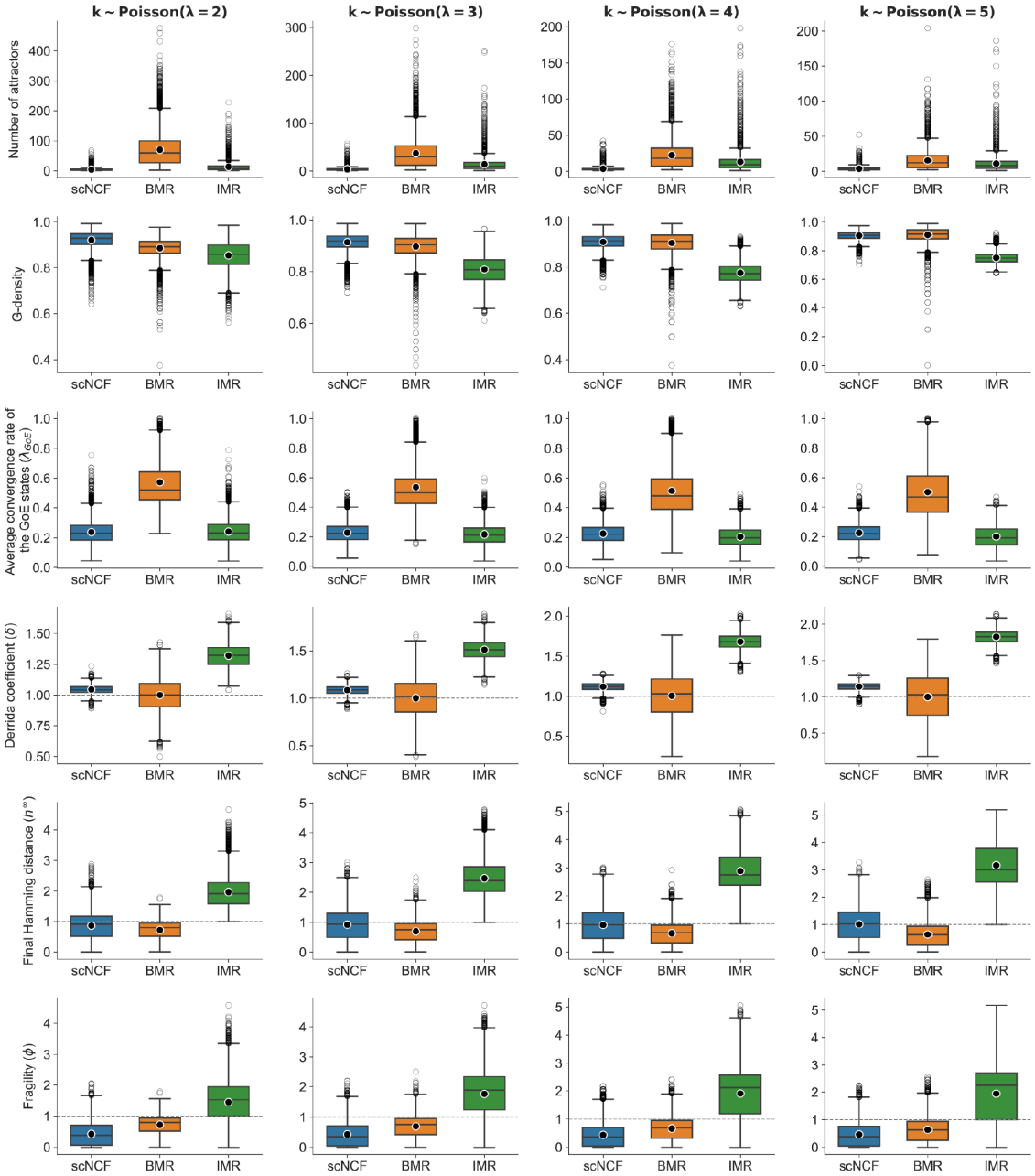


Figure 6.7: BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the P-P network topology with $N = 12$. Each row corresponds to one stability observable and each of the four columns represents different values of the mean in-degree parameter ($\lambda = 2, 3, 4, 5$). In each subplot, the three boxplots present the distributions of the observables across three subtypes of ThFs: scNCFs, BMRs and IMRs. Across all observables and values of λ, IMR-based models consistently display the least stable behavior, while scNCF and BMR models typically cluster closer to the ordered or critical regimes. Outliers in each distribution are shown as grey circles while the black dots outlined in white denote the mean of each group.

are underrepresented in the empirical BF dataset relative to theoretical expectations. The apparent enrichment of BMRs is largely driven by BMRs that are also NCFs, whereas non-NCF BMRs are as scarce as IMRs. Finally, we asked whether TMRs still serve as a practical replacement for BFs in reconstructed Boolean GRN models. **For a set of 24 high-confidence published models, we found that the corresponding TMR variants generally fail to recover the biological attractors and their basin-size distributions, whereas substitution of any non-NCF BFs with NCFs may yield alternative models that match or even outperform the original published models.** Note that although we employed a synchronous update scheme in our analysis, many of our claims are robust to update scheme as most biological attractors are fixed points, which are unchanged under synchronous or asynchronous updating. We also performed a dynamical stability analysis using ensembles of RBNs, showing that BMRs and NCFs typically keep dynamics near the ordered or critical regime, while IMRs push dynamics toward chaos. Both TMR types tend to produce many attractors with very small basins, a phenomenon also seen in well-known cell-cycle Boolean models of *S. cerevisiae* [67] and *S. pombe* [34], where small-basin attractors were discarded as biologically irrelevant.

While this work was in progress, a preprint by Kadelka *et al.* [190] (now published as [191]) appeared, which also analyzed the two TMR variants–termed the Ising (IMR) and 01 (BMR) formalisms. The authors report that, in small and weakly connected networks, TMRs reproduce dynamics similar to real systems, consistent with our observation of higher ARS in small networks. They also report a 65–80% mean similarity between TMRs and biological rules for up to six regulators based on functional and state space similarity measures. However, our results indicate that such overall similarity is misleading. Despite apparent agreement, TMRs frequently fail to recover known biological attractors, the primary benchmark for Boolean GRN models. Moreover, Kadelka *et al.* do not assess TMRs' intrinsic complexity or robustness. We find that TMRs are substantially more complex and sensitive than NCFs, and that IMRs drive dynamics toward chaos. Kadelka *et al.* also proposed modified variants, Ising* and 01*. Testing these on our 24 selected models yields only marginal improvements: cases with ARS = 0 remain unchanged, and gains elsewhere are minimal (see Appendix C, Figures C.6 and C.7). Although Ising* produces odd bias when k is a power of 2, it remains non-canalyzing for $k > 2$ (see Appendix C, Properties C.11 and C.12), while 01* continues to render some regulations ineffective (see Appendix C, Property

C.13). Overall, neither the original nor the modified TMRs provide a reliable default logic for modeling GRNs.

Data and code availability statement

All data and codes needed to reproduce the results in this chapter are deposited in GitHub and are available at: https://github.com/asamallab/TMRvsBioLogic.

Chapter 7

Summary and future outlook

Only theory can turn a heap of facts into a tower of knowledge.

– Andreas Wagner

7.1 Summary

This thesis addressed the central question of how the choice of regulatory logic influences the dynamics and stability of Boolean models of gene regulatory networks (GRNs). Despite the widespread adoption of Boolean models in systems biology, this question has received less systematic attention than complementary questions concerning network topology and update schemes. To address this gap, the thesis systematically investigates how different classes of Boolean functions (BFs) shape the dynamics, and stability of Boolean models of GRNs. In particular, this thesis examined how different classes of BFs influence network stability, the structure of the state transition graph (STG), and the suitability of certain simple threshold-based logic rules for modeling biological regulatory systems. The three core studies presented in this thesis, spanning abstract random Boolean networks (RBNs), reconstructed biological GRNs, and a critical evaluation of threshold-based modeling rules, collectively demonstrate that the choice of regulatory logic class is a key determinant of Boolean network (BN) dynamics.

Chapter 2 introduces the theoretical and methodological foundations of Boolean modeling of GRNs and their associated logical update rules. We present the basic elements of the framework, including network states, STGs, attractors, and update schemes, and provide a detailed account

of several classes of biologically meaningful BFs, namely effective functions (EFs), effective and unate functions (EUFs), read-once functions (RoFs), nested canalyzing functions (NCFs), and link operator functions (LOFs). We also review the random Boolean network (RBN) paradigm in the pre-genomic context. The chapter further discusses reconstructed Boolean models of biological systems that emerged in the post-genomic era and highlights how these developments enabled the creation of repositories and resources of Boolean models. In addition, we discuss two complexity measures of BFs, Boolean complexity and average sensitivity, and present strategies for sampling Boolean models under biological constraints, including a hypercube-based iterative sampling method that enables scalable sampling of EUFs for nodes with higher input degrees.

Next, in Chapter 3, we provide a systematic overview of stability measures for BN dynamics and organize the diverse notions of stability into a coherent framework. We review perturbation-based measures derived from damage spreading arguments, including the Derrida coefficient (δ), the final Hamming distance (h^{∞}), and fragility (ϕ), and describe stability measures based on properties of attractors and their basins of attraction. We also discuss measures associated with polynomial approximations of regulatory rules and the resulting notion of approximability of network dynamics. Finally, we introduce cellular automata (CA)-inspired stability measures that characterize global properties of the STG, namely bushiness and convergence. Together, Chapters 2 and 3 establish the conceptual vocabulary and methodological toolbox drawn upon in the subsequent chapters.

In Chapter 4, we address the first objective of this thesis and systematically examine how the choice of BF class modulates the dynamical regime of RBNs. Using five classes of BFs ranging from almost random (EF) to biologically meaningful (EUF, RoF, NCF, LOF), and a comprehensive suite of stability measures spanning damage spreading, attractor properties, and approximability, we show that a clear and robust stability hierarchy holds across all cases: EF < EUF < (RoF, NCF, LOF), in increasing order of stability. This hierarchy is insensitive to network size (N), average connectivity (k), and degree distribution, indicating that regulatory logic plays a decisive role in shaping network dynamics rather than the effect being an artifact of a particular topological ensemble. Furthermore, whereas random BFs (EFs) push network dynamics toward the chaotic regime as connectivity or network size increases, biologically meaningful BF classes exhibit minimal fluctuations across most stability measures, demonstrating a robustness to topological vari-

ation not shared by random logic. A holistic summary of these trends across different network topologies is provided in Figure 7.1. Taken together, the results of Chapter 4 show that restricting regulatory logic to biologically meaningful classes is not merely a matter of biological fidelity but also confers a dynamical advantage: such networks are inherently more stable and more robust to perturbations and structural variation.

In Chapter 5, we shift the focus from abstract RBNs to reconstructed Boolean GRNs and investigate how different classes of BFs influence the structural features of the STG. Using ensembles in which network topology and biological attractors are held fixed while only the regulatory logic is varied, we quantify two complementary global properties of the STG inspired by cellular automata: bushiness, measured by the G-density, and convergence rate, measured by λ_{GoE}. We find that biologically meaningful BFs consistently give rise to more bushy and more convergent STGs than random EFs, indicating more robust network dynamics. This result holds across all ten reconstructed GRNs analysed and therefore exhibits broad generality. Because the computation of such global measures becomes computationally expensive for larger networks, we further develop scalable proxies. By adapting Wuensche's Z-parameter from CA to BNs and introducing four variants, we identify $\overline{Z}_{mid}$ as the best proxy for G-density, along with the network sensitivity ($\overline{s}$). For convergence, we show that λ_{random}, the average convergence rate computed from randomly sampled initial states rather than GoE states, serves as an excellent and computationally feasible proxy for λ_{GoE}. In summary, Chapter 5 provides three scalable proxies: $\overline{Z}_{mid}$, a novel CA-inspired quantity introduced to the BN community, and $\overline{s}$, both of which reflect the bushiness or contraction rate of the STG, and λ_{random}, which provides a reliable estimate of λ_{GoE}. These developments offer practical tools for studying global dynamical properties of larger BNs, complement existing measures for quantifying the robustness of network dynamics, and thereby inform model selection.

While the analyses presented in Chapters 4 and 5 highlight the critical role of regulatory logic in determining network dynamics, in realistic biological contexts the exact regulatory rules governing gene expression are often unknown. As a result, modelers commonly adopt simplified update rules for practical convenience. One such commonly used class of rules is the threshold majority rule (TMR), which has gained popularity because of its conceptual simplicity and ease of implementation.

In Chapter 6, we systematically assess the suitability of TMRs as a modeling framework for

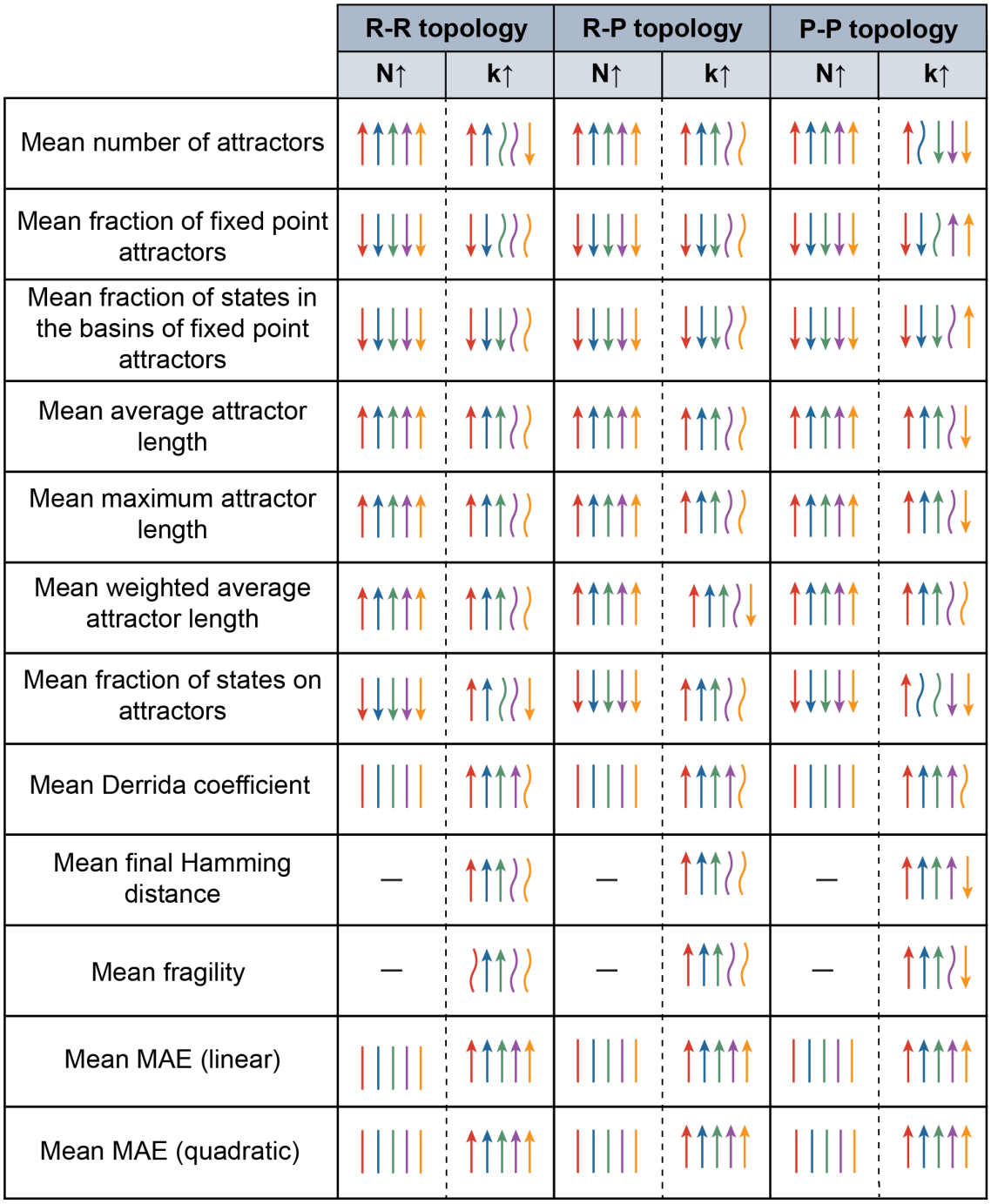


Figure 7.1: Trends of different stability measures with varying N and k. The figure consists of 12 rows, each corresponding to a different stability measure. The columns represent three network topologies: R-R, R-P and P-P. For each topology we provide two sub-columns: one for increasing network size (N) and another for increasing average in-degree (k). In each case, the following notations have been used to indicate the different trends: an upward arrow (↑) indicates an increasing trend, a downward arrow (↓) indicates a decreasing trend, a curved line (ᔓ) represents an inconsistent trend, a vertical bar (|) represents an absence of significant variation, and a dash (-) denotes insufficient data for trend inference (e.g. computations not accessible with our facilities). In each block, the colors of the symbols correspond to different classes: red for EF, blue for EUF, green for RoF, purple for NCF, and orange for LOF.

Boolean GRNs by comparing the two commonly used TMR variants with NCFs across structural, empirical, and dynamical criteria. Our analyses show that TMRs exhibit limited canalyzation, display discordant bias patterns, and frequently eliminate self-inhibitions. They are also underrepresented in empirical dataset. Moreover, when compared with NCFs, TMRs exhibit greater complexity and higher sensitivity to perturbations. At the network level, TMRs almost always fail to reproduce biological attractors and their associated basin size distributions, whereas models incorporating suitable NCFs consistently perform better. Overall, the analyses presented in Chapter 6 reveal that TMRs, despite their conceptual appeal, fall short as a general strategy within the Boolean modeling framework for capturing biologically realistic regulatory logic. While they may perform satisfactorily in certain cases, their rigid, one-size-fits-all logic lacks flexibility and often misaligns with biological regulatory mechanisms. In contrast, NCFs offer multiple sign-conforming choices, exhibit minimal Boolean complexity and average sensitivity, and recover biological attractors with a much higher success rate and greater dynamical stability. Their enrichment in empirical datasets further suggests a potential evolutionary preference. We therefore recommend moving beyond TMRs in Boolean modeling despite their convenience.

Taken together, the three studies presented in this thesis converge on a unified conclusion: the class of regulatory logic assigned to individual genes in a Boolean model is a primary determinant of both local and global network dynamics. Biologically meaningful BF classes, particularly NCFs, consistently outperform random BFs or simplified alternatives (such as TMRs) across the criteria examined. These findings have direct implications for the reconstruction, analysis, and interpretation of Boolean GRN models.

7.2 Future outlook

The analyses presented in this thesis open several avenues for future investigation, spanning methodological improvements, theoretical extensions, and broader biological applications.

The analyses conducted in Chapter 4 reveal that biologically meaningful BF classes consistently stabilize network dynamics across a range of network sizes and connectivities, and that networks employing such functions can largely withstand variations in topology while preserving overall stability. However, the computational analyses presented in Chapter 4 are necessarily restricted to modest values of N and k due to the high computational demand. Whether the observed

stability hierarchy persists in the limit of very large networks or very high connectivity, therefore, remains an open question. Thus, an immediate priority is to develop reliable proxy measures that extend the analysis of dynamical stability to larger networks. A particularly interesting direction would be to derive analytical results or approximations to determine if the stability hierarchy persists as networks become large and/or highly connected, since addressing this question purely through computation appears infeasible. Exploiting known properties such as the minimal average sensitivity of NCFs may provide a route toward such analytical predictions, though whether such an analytical extension is mathematically tractable is itself an open and interesting question. Furthermore, the analysis in Chapter 4 was conducted on a limited set of random network topologies. It would therefore be worthwhile to examine whether the presence of specific network motifs, such as feedback loops or feed-forward loops commonly observed in biological regulatory networks, modulates the stabilizing effect of biologically meaningful BF classes. Finally, all analyses in Chapter 4 were performed under synchronous updating; investigating whether similar conclusions hold under asynchronous dynamics remains an important direction for future work.

Chapter 5 also opens several avenues for future methodological improvement. Notwithstanding the developments of new methods and results, there are some aspects that could be addressed in the future. Computing Z-parameters for nodes with more than six inputs becomes computationally expensive due to the factorial growth in the number of input permutations that must be enumerated. Developing efficient algorithms for computing Z-parameters at higher input degrees would therefore significantly extend the applicability of these local descriptors. Furthermore, all global STG measures in this chapter are defined with respect to synchronous update dynamics. Extending these measures to asynchronous update schemes presents both theoretical and computational challenges. Finally, while λ_{random} serves as an excellent proxy for λ_{GoE}, the proxy still relies on information at the network level. An important open question is whether equally effective proxies can be constructed solely from properties of the individual BFs, without requiring simulation of network dynamics, as was possible for the bushiness measures.

In both Chapters 4 and 5, we have sampled EUFs using an algorithm proposed in Chapter 2 of this thesis. The current sampling scheme is not completely uniform and may therefore introduce a small bias. Although this bias is negligible for nodes with small input degrees, developing algorithms that sample EUFs more uniformly would strengthen the conclusions for networks con-

taining nodes with larger input degrees.

The enrichment analysis presented in Chapter 6 assumes a uniform baseline in which all BFs are equally probable. A promising yet challenging future direction is to develop a biophysical model that incorporates mechanistic details of transcriptional regulation together with the underlying thermodynamic constraints to infer plausible regulatory logic rules. Such a framework could serve as a more realistic null model for enrichment analysis, and a possible starting point may be the work of Buchler *et al.* [192]. In Chapter 6, we recommend moving beyond TMRs in Boolean modeling, despite their convenience. However, even when the analysis is restricted to biologically meaningful classes such as NCFs, model selection remains a nontrivial challenge. For a given combination of regulatory signs, multiple NCFs are typically consistent with the available biological information, and attractor recovery alone is often insufficient to uniquely determine the appropriate rule. Developing additional, well-motivated selection criteria to navigate this degeneracy, therefore, remains an important open problem. For example, the network-level dynamical properties introduced in Chapter 5, namely the bushiness and convergence of the STG, may serve as useful criteria for ranking candidate models. Furthermore, our earlier work [64] showed that constraints based on relative stability or developmental landscapes can guide the ranking and selection of plausible models. Integrating such criteria with quantitative experimental data, such as time-series gene expression, perturbation experiments, and single-cell multi-omics data, could enable a more systematic and biologically grounded approach to Boolean GRN reconstruction. Recent advances in GRN inference from single-cell and multi-omics data, together with machine learning based approaches, provide promising opportunities to infer regulatory interactions and constrain the space of plausible Boolean regulatory rules [47, 193, 194]. Combining these data-driven approaches with biologically meaningful classes of BFs and model selection criteria could facilitate the reconstruction of more predictive and biologically realistic Boolean GRN models. Rather than sacrificing biological realism by resorting to simplified rules such as TMRs, the challenge ahead lies in developing principled strategies for navigating the space of biologically meaningful BFs that capture combinatorial regulation at the single-gene level and identifying the most appropriate model for a given biological system.

Empirical analyses of regulatory logic have shown that certain classes of BFs, including NCFs and RoFs, occur far more frequently in curated biological datasets than would be expected by

chance [64]. The results presented in this thesis show that regulatory networks governed by such rules can benefit dynamically, exhibiting properties that promote stable and organized system behavior. A natural question that follows is how such regulatory rules emerge during the course of biological evolution. Addressing this question would likely require comparative inference of regulatory logic across homologous genes in different species, since currently available reconstructed Boolean models do not provide sufficient information to resolve these evolutionary patterns. Pursuing this line of inquiry further, an evolutionary perspective on regulatory logic may also enable the application of evolution-inspired methodologies, such as directed evolution, for the synthetic design of regulatory circuits.

Finally, the findings of this thesis are framed within the Boolean approximation, where gene expression states are represented as binary values. Whether the insights gained here carry over to frameworks such as multi-valued logical models that allow finer-grained representation of gene activity levels remains an open and potentially fruitful question. Similarly, incorporating stochasticity through noisy Boolean dynamics or hybrid Boolean-continuous models may broaden the applicability of the regulatory logic perspective to phenomena such as gene expression variability and cellular heterogeneity.

In summary, the broader outlook emerging from this thesis lies in integrating improved computational methods, biophysically grounded modeling approaches, and principled reconstruction and model selection strategies. Together, these developments could enable Boolean GRN models that remain computationally tractable while more faithfully capturing the biological mechanisms of combinatorial gene regulation.

Appendix A

Supplementary Material for Chapter 4

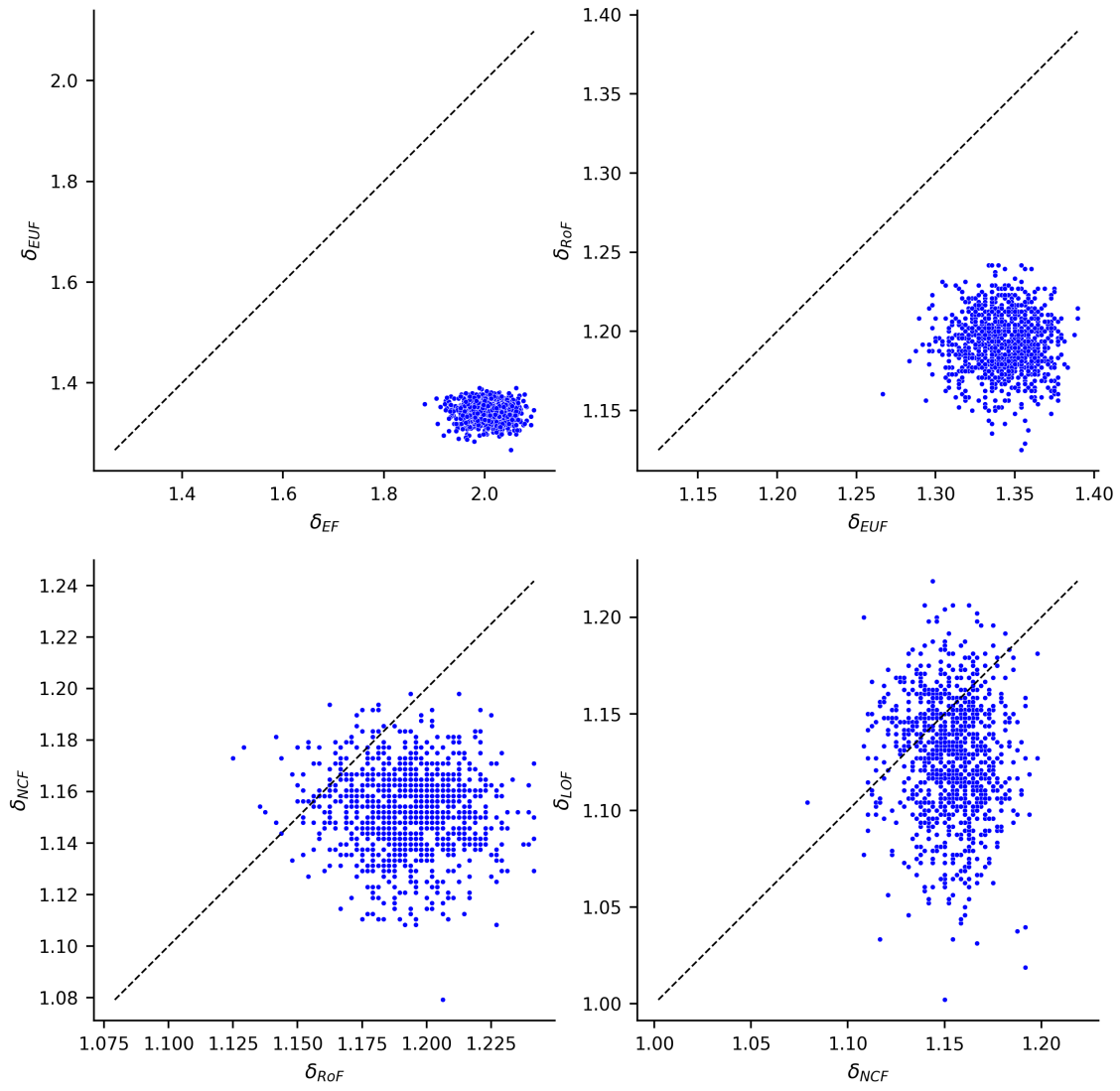


Figure A.1: Comparison of mean values of δ when using different classes of BFs on identical networks. For each of the 1000 different networks with k-regular (R-R) topology ($k_{in} = k_{out} = 4$, $N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The values of δ were then averaged over the 10 models and plotted. Each sub-figure provides comparisons across two successive classes of the putative hierarchy. The dashed line is the X=Y boundary.

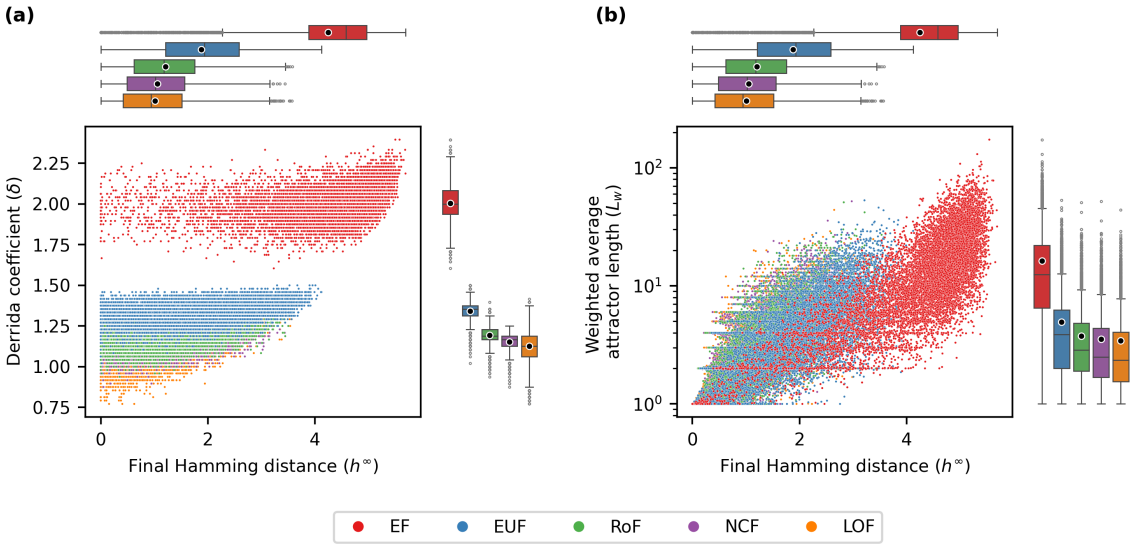


Figure A.2: Comparing the distributions of h^∞ against δ and h^∞ against L_w. For each of the 1000 different networks with k-regular (R-R) topology ($k_{in} = k_{out} = 4$, $N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities h^∞, δ and L_W were computed for each instance. The sub-figure **(a)** illustrates the joint values of δ and h^∞ along with their marginal distributions for different ensembles. Similarly, the sub-figure **(b)** presents the joint values of δ and L_W across ensembles.

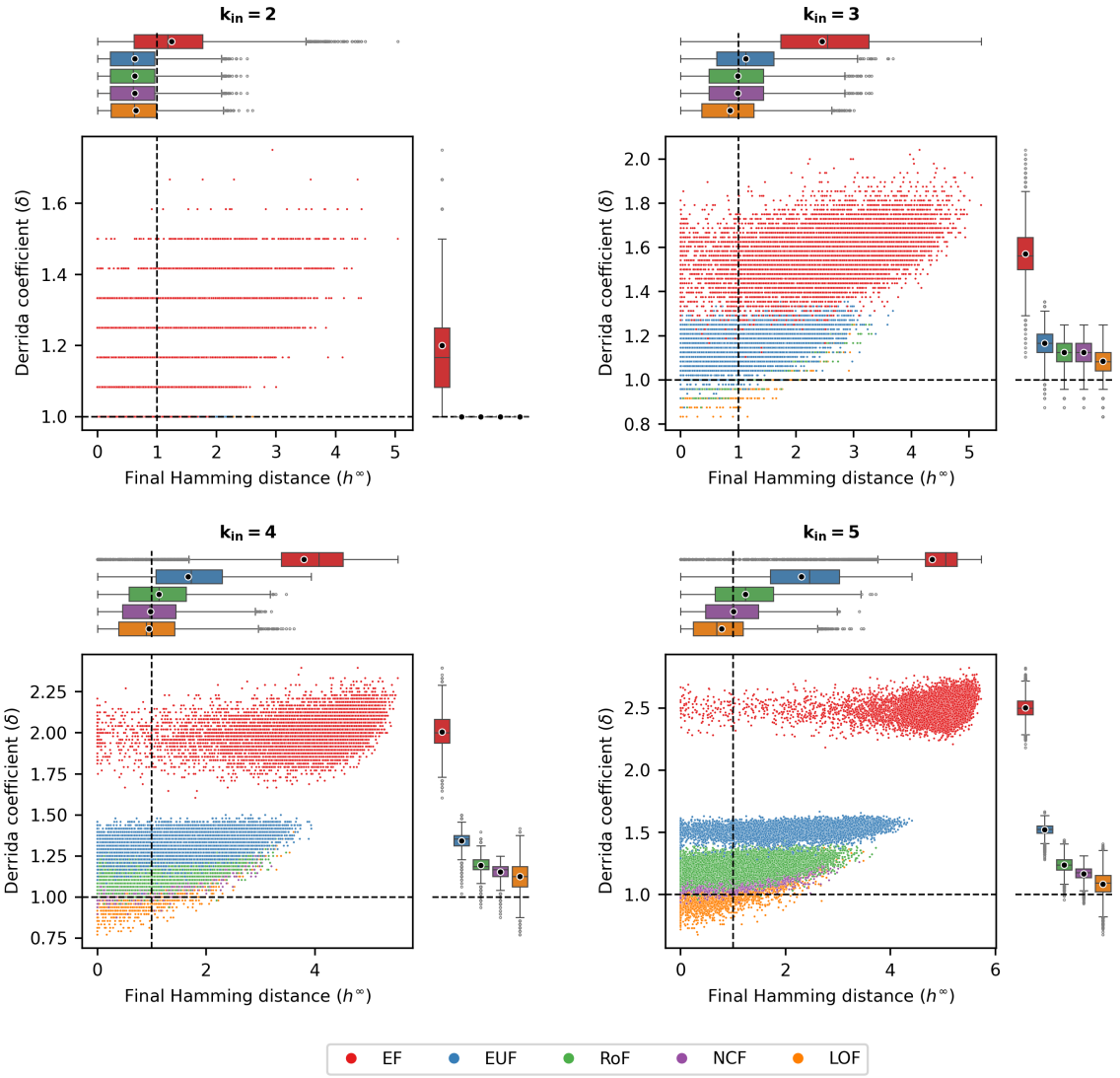


Figure A.3: The distributions of h^∞ against δ for R-P networks. For each of the 1000 different networks with R-P topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities h^∞, δ were computed for each instance. Each sub-figure corresponds to a different number of inputs per node, ranging from 2 to 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $h^\infty = 1$ respectively.

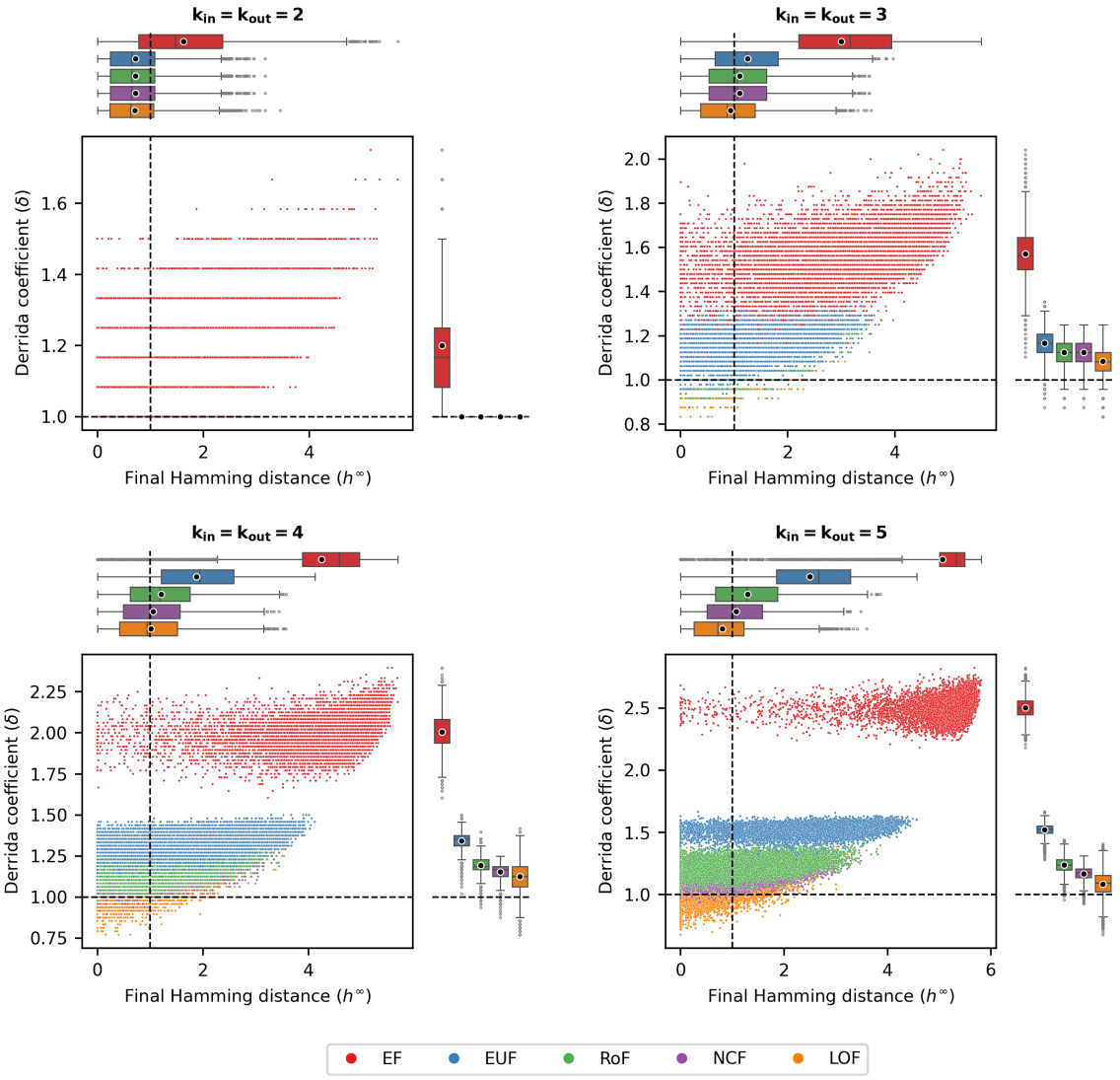


Figure A.4: The distributions of h^∞ against δ for R-R networks. For each of the 1000 different networks with R-R topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities h^∞, δ were computed for each instance. Each sub-figure corresponds to a different number of inputs per node, ranging from 2 to 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $h^\infty = 1$ respectively.

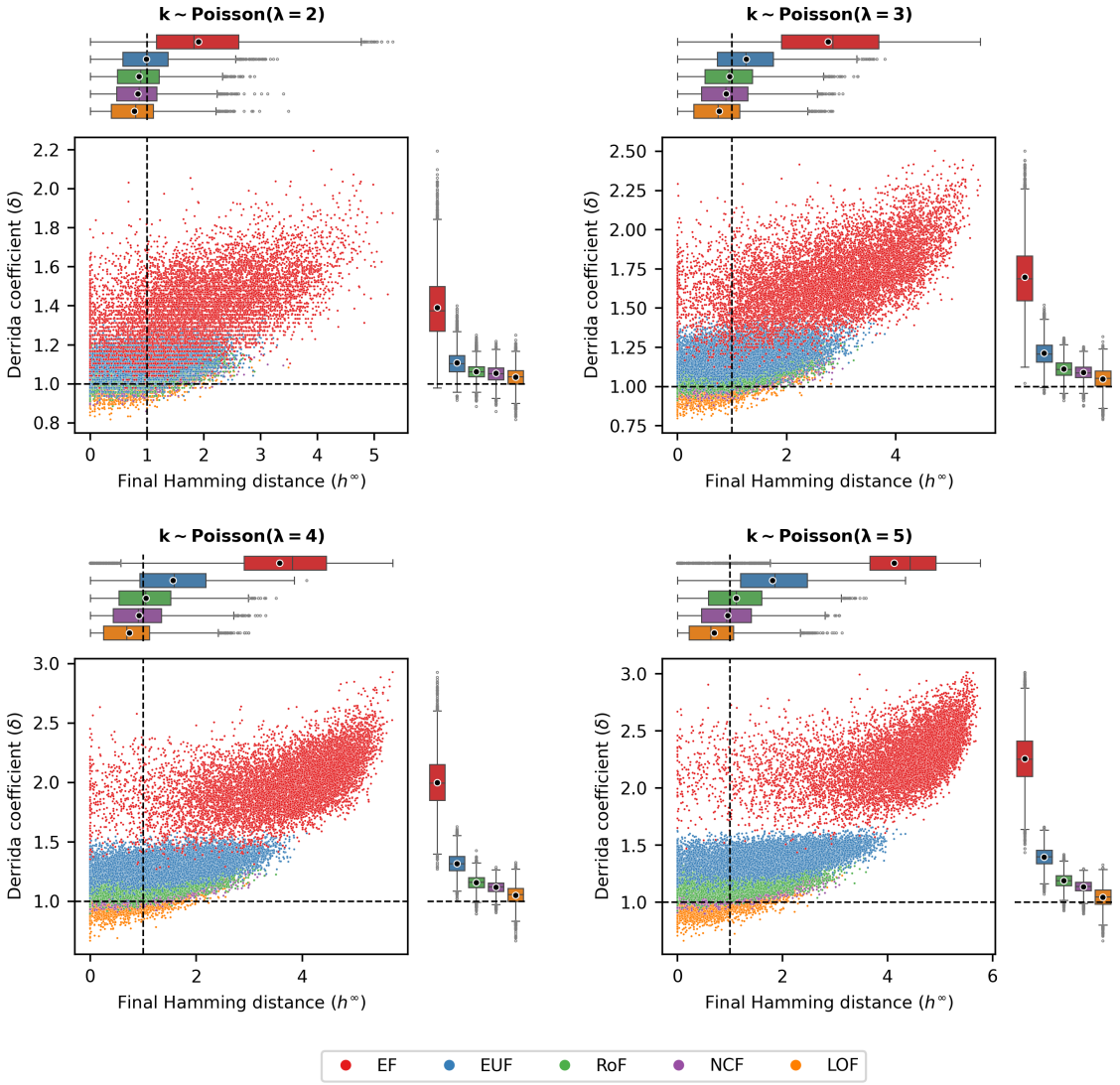


Figure A.5: The distributions of h^∞ against δ for P-P networks. For each of the 1000 different networks with P-P topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities h^∞, δ were computed for each instance. Each sub-figure corresponds to a different average number of inputs per node, with the degrees drawn from Poisson distributions with means 2, 3, 4 and 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $h^\infty = 1$ respectively.

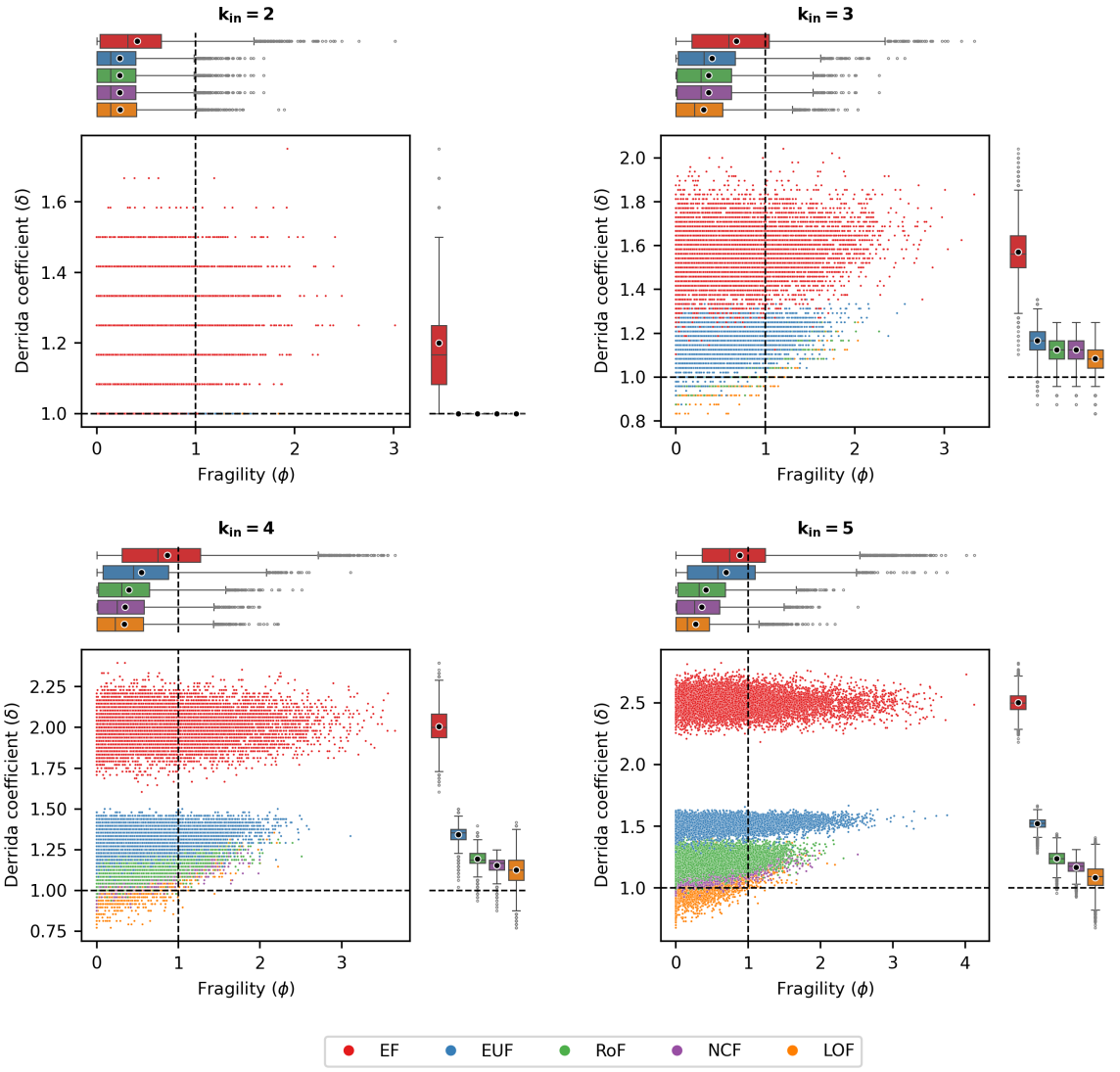


Figure A.6: The distributions of ϕ against δ for R-P networks. For each of the 1000 different networks with R-P topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities ϕ, δ were computed for each instance. Each sub-figure corresponds to a different number of inputs per node, ranging from 2 to 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $\phi = 1$ respectively.

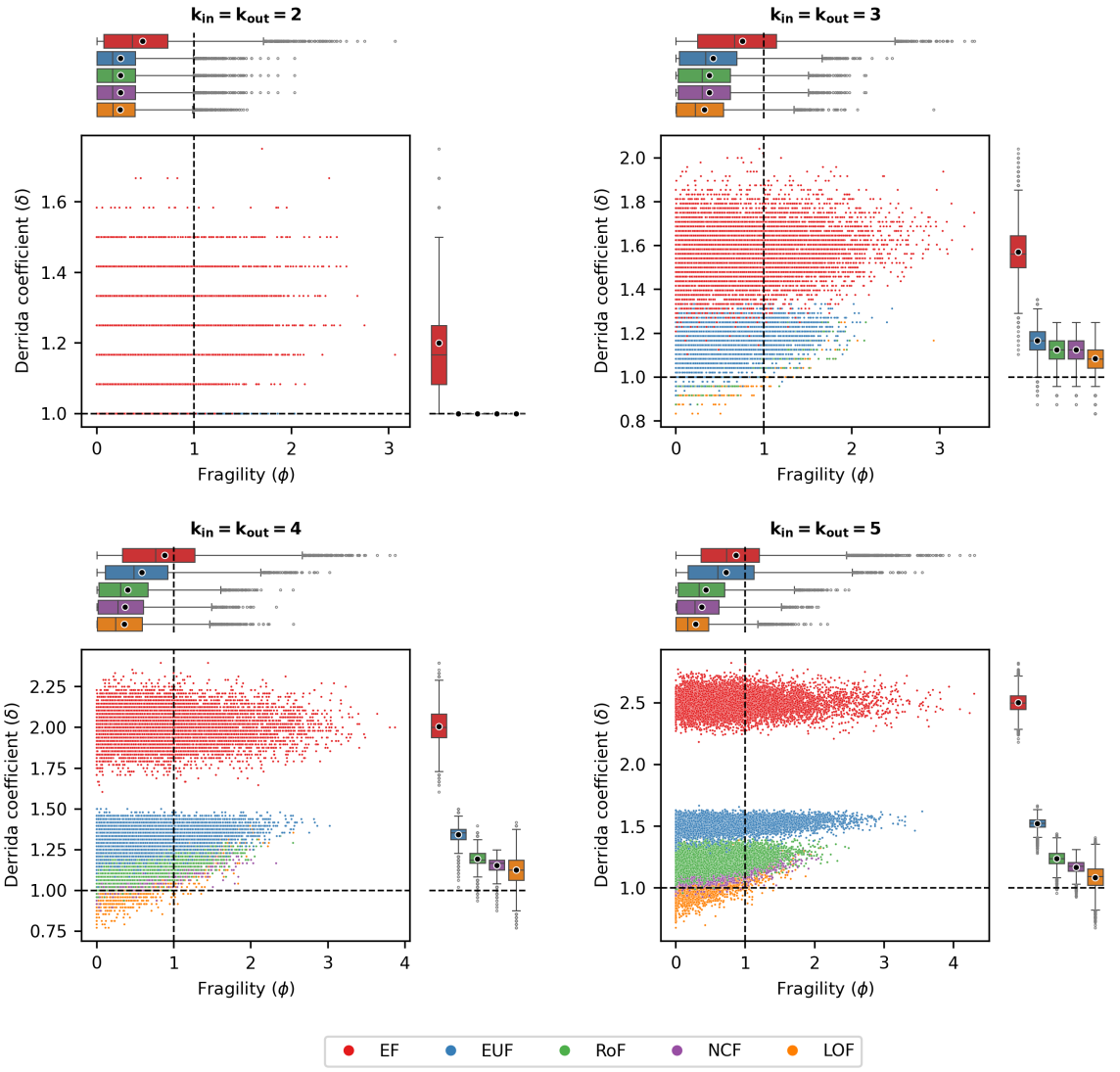


Figure A.7: The distributions of ϕ against δ for R-R networks. For each of the 1000 different networks with R-R topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities ϕ, δ were computed for each instance. Each sub-figure corresponds to a different number of inputs per node, ranging from 2 to 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $\phi = 1$ respectively.

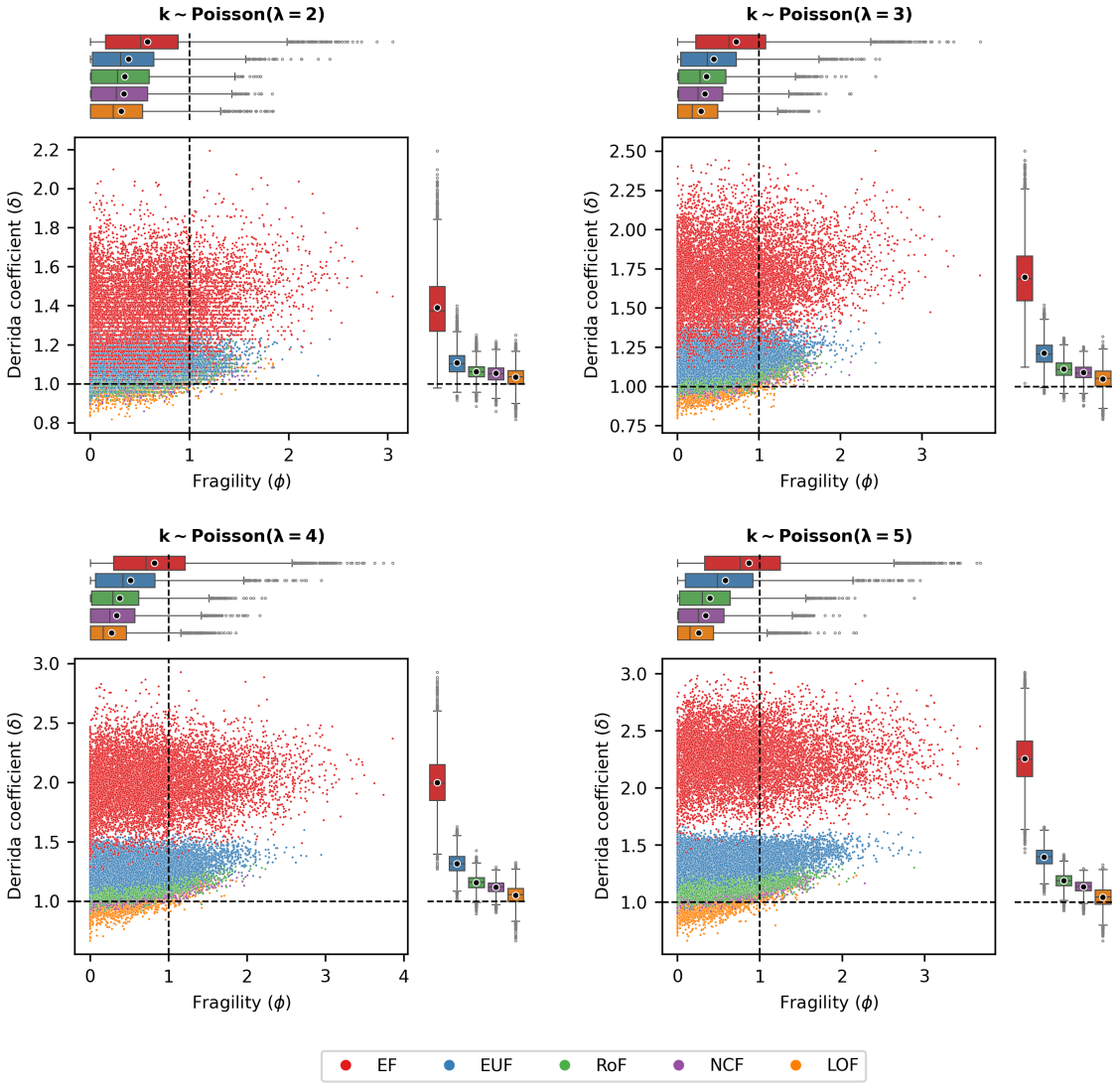


Figure A.8: The distributions of ϕ against δ for P-P networks. For each of the 1000 different networks with P-P topology ($N = 12$), 10 models were randomly generated, for each of the 5 different ensembles of models (EF, EUF, RoF, NCF and LOF). The quantities ϕ, δ were computed for each instance. Each sub-figure corresponds to a different average number of inputs per node, with the degrees drawn from Poisson distributions with means 2, 3, 4 and 5. In each sub-figure, the horizontal and vertical dashed black lines correspond to $\delta = 1$ and $\phi = 1$ respectively.

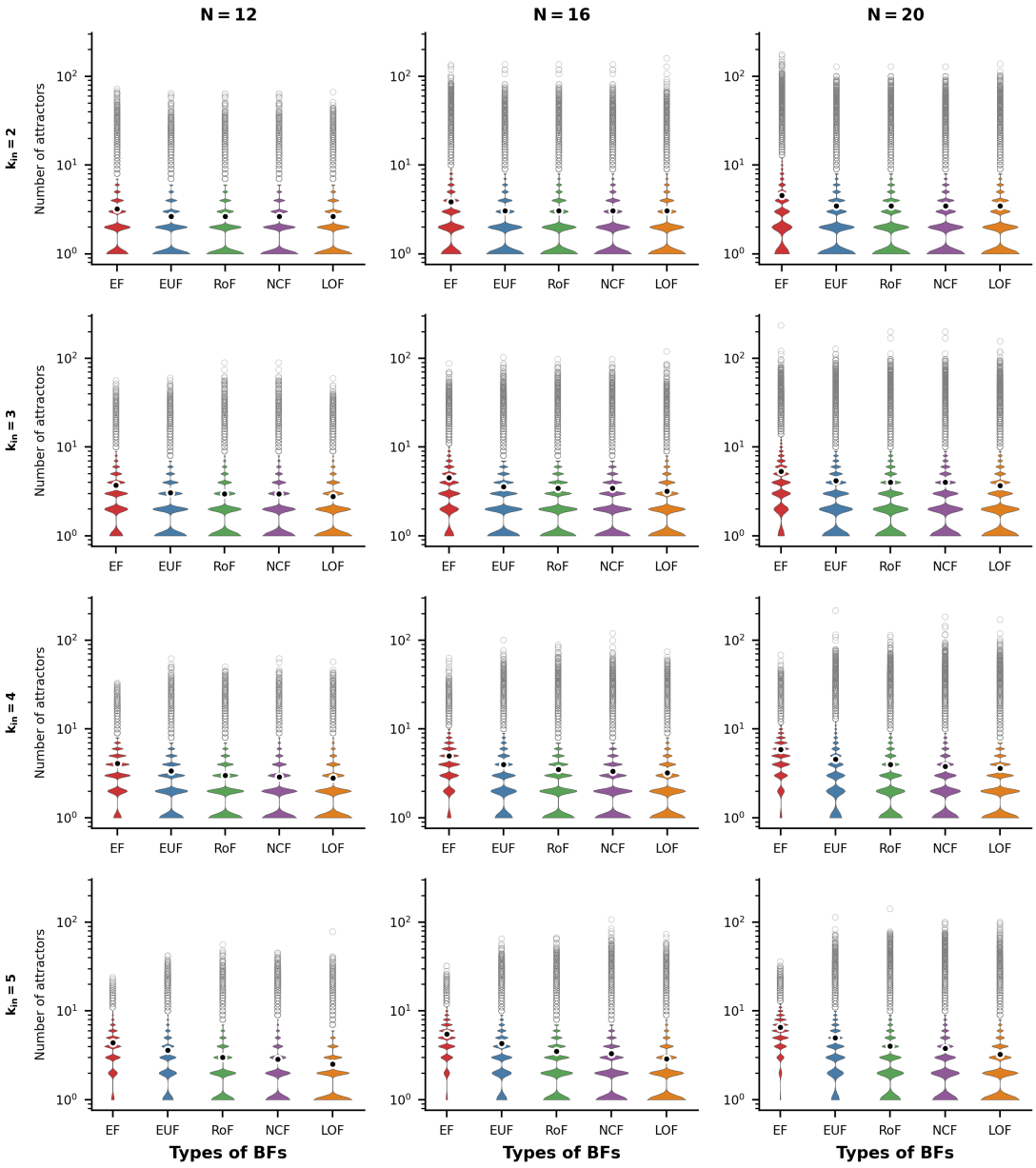


Figure A.9: Distributions of number of attractors for R-P networks. The violin plots display the distribution of the number of attractors for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the number of attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

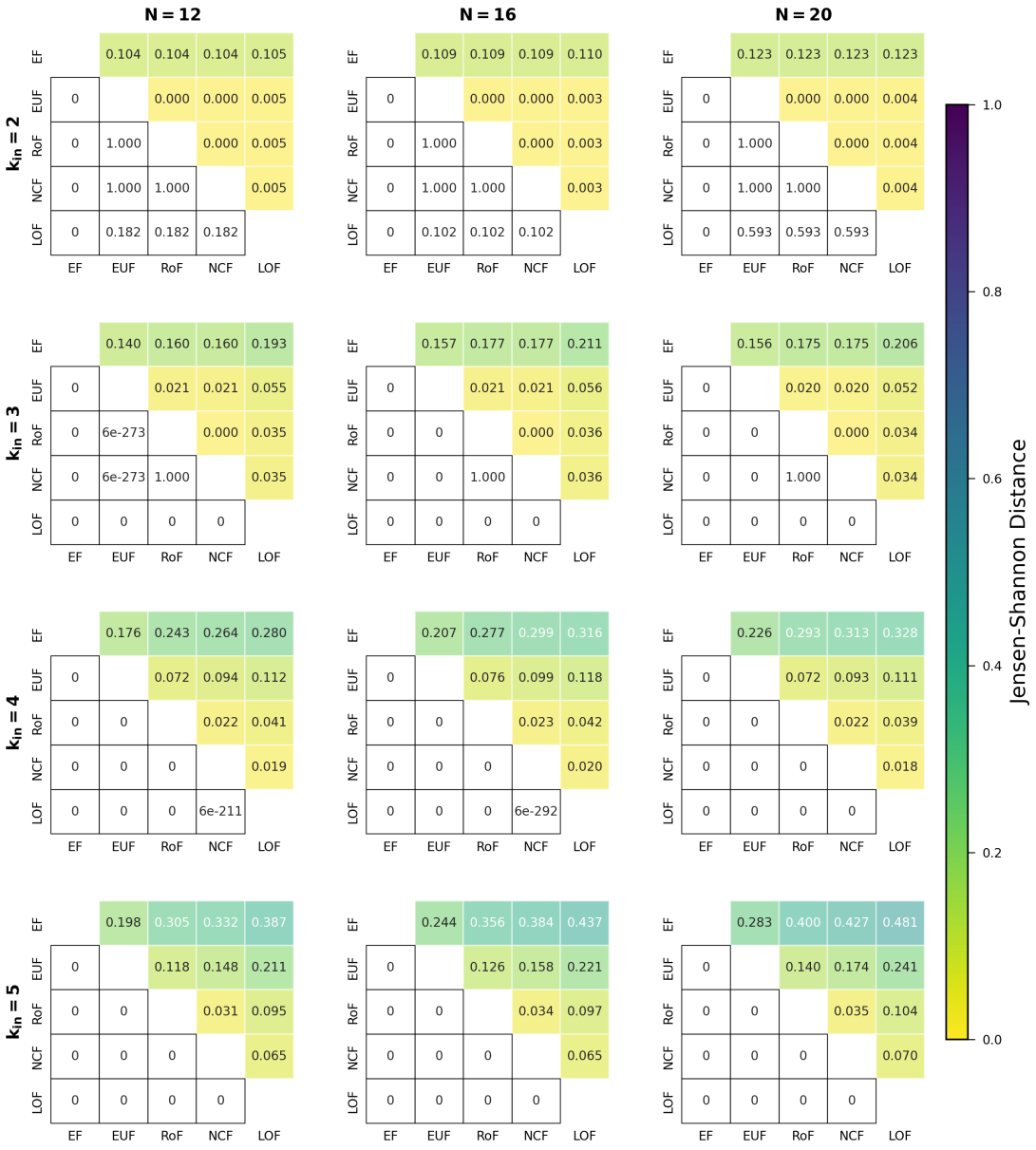


Figure A.10: Comparing distributions for the number of attractors when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the number of attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

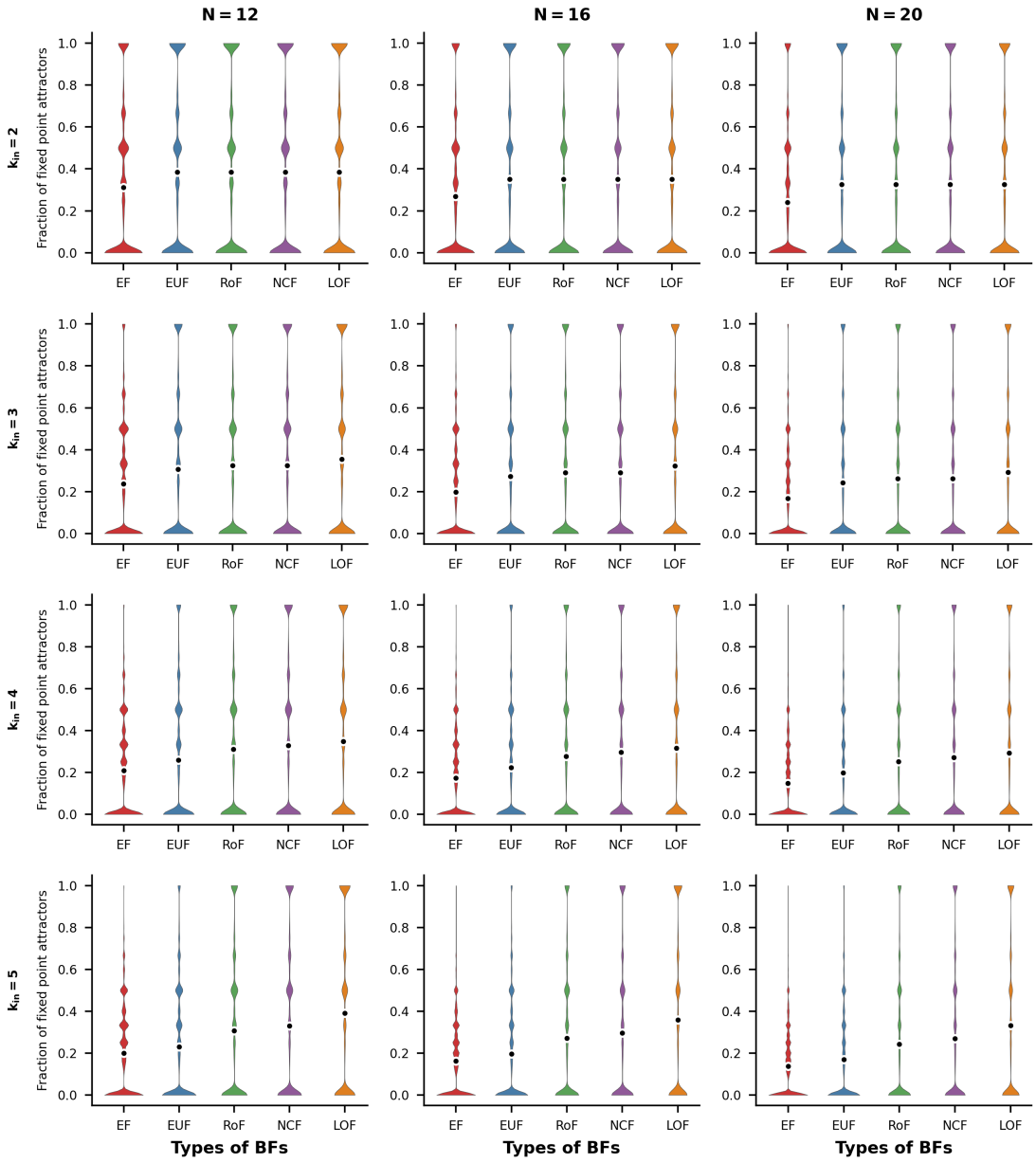


Figure A.11: Distributions of fraction of fixed point attractors for R-P networks. The violin plots display the distribution of the fraction of fixed point attractors for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

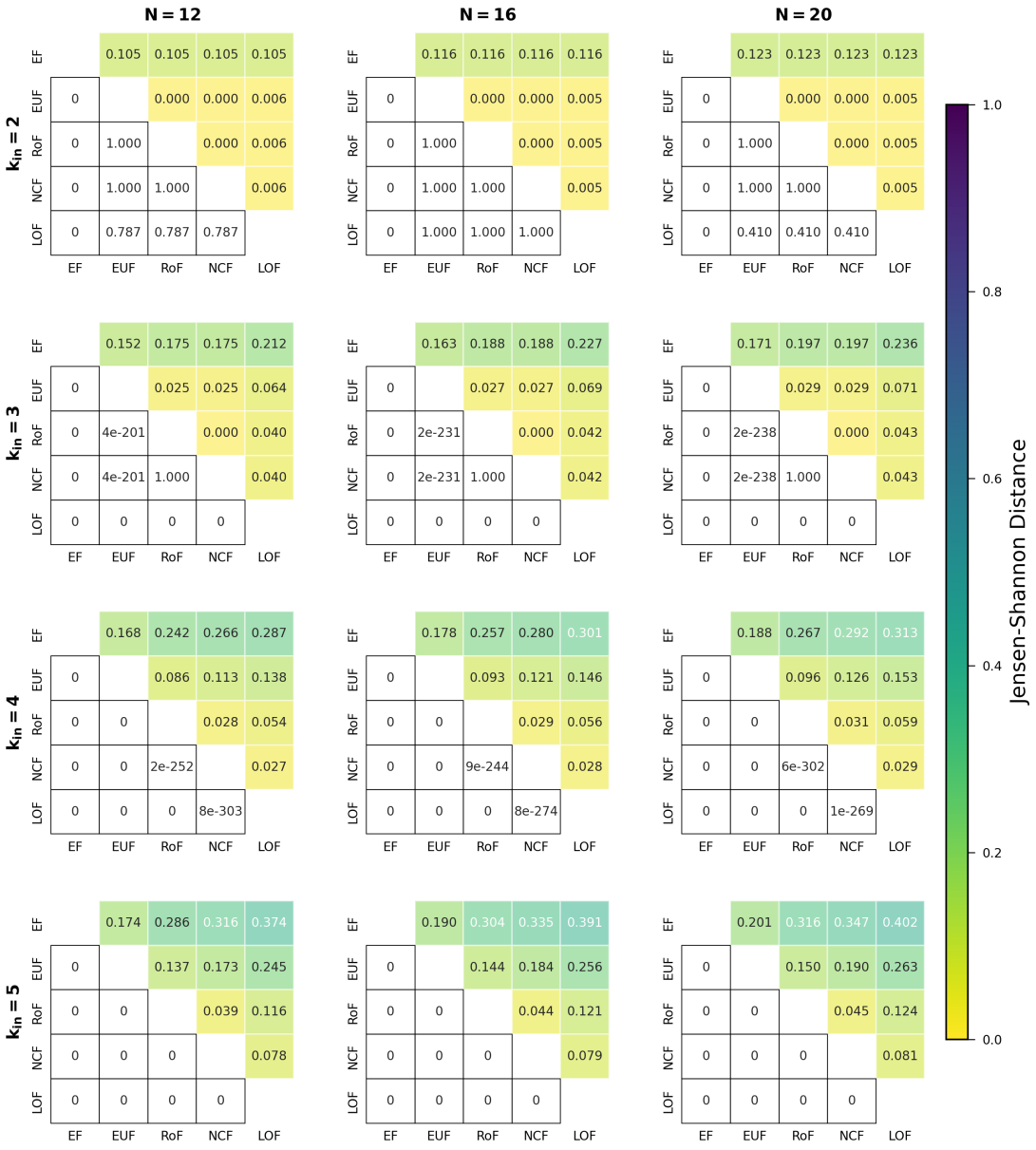


Figure A.12: Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the fraction of fixed point attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

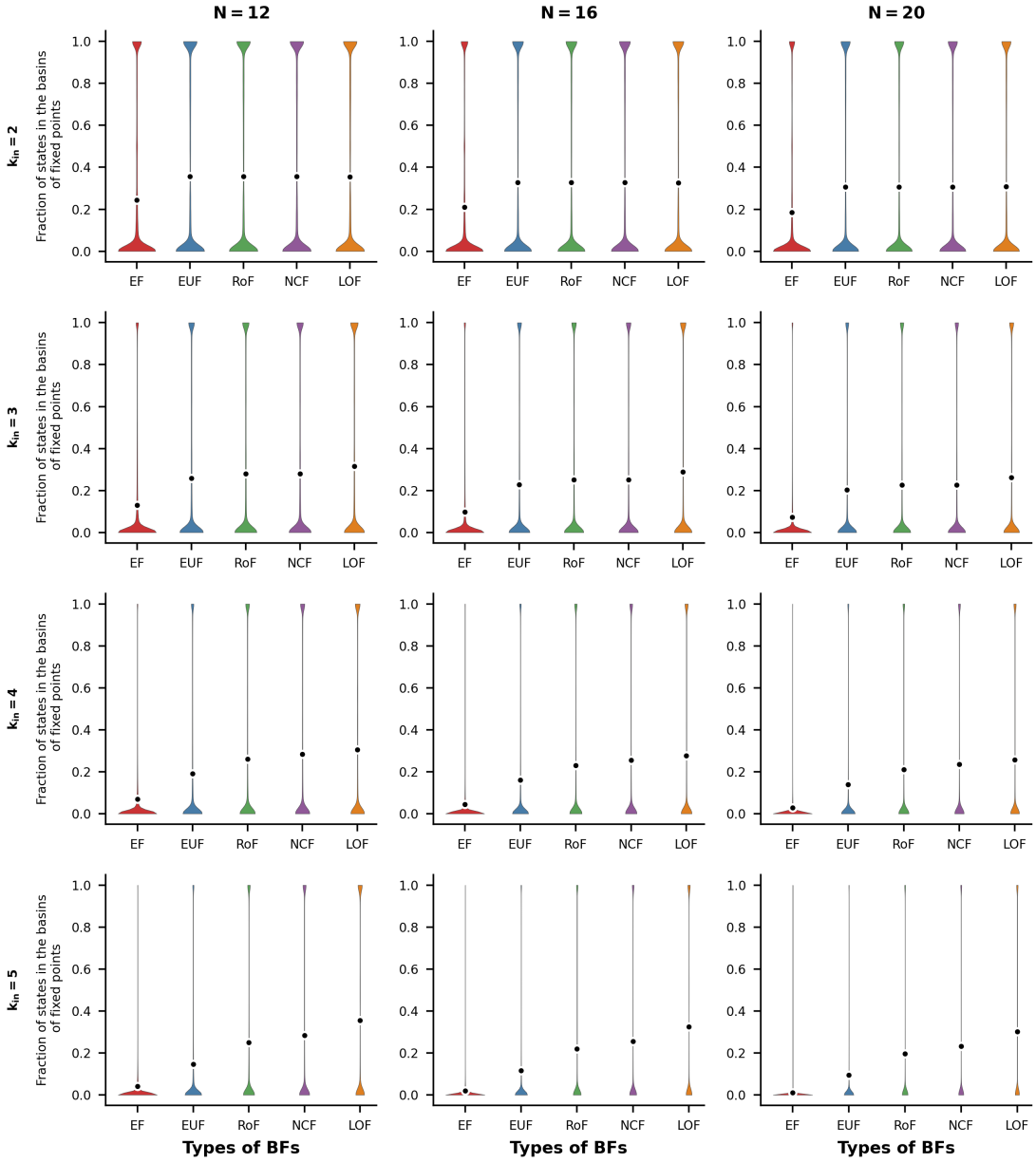


Figure A.13: Distributions of fraction of states in the basins of fixed points for R-P networks. The violin plots display the distribution of the fraction of states in the basins of fixed point attractors for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states in the basins of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

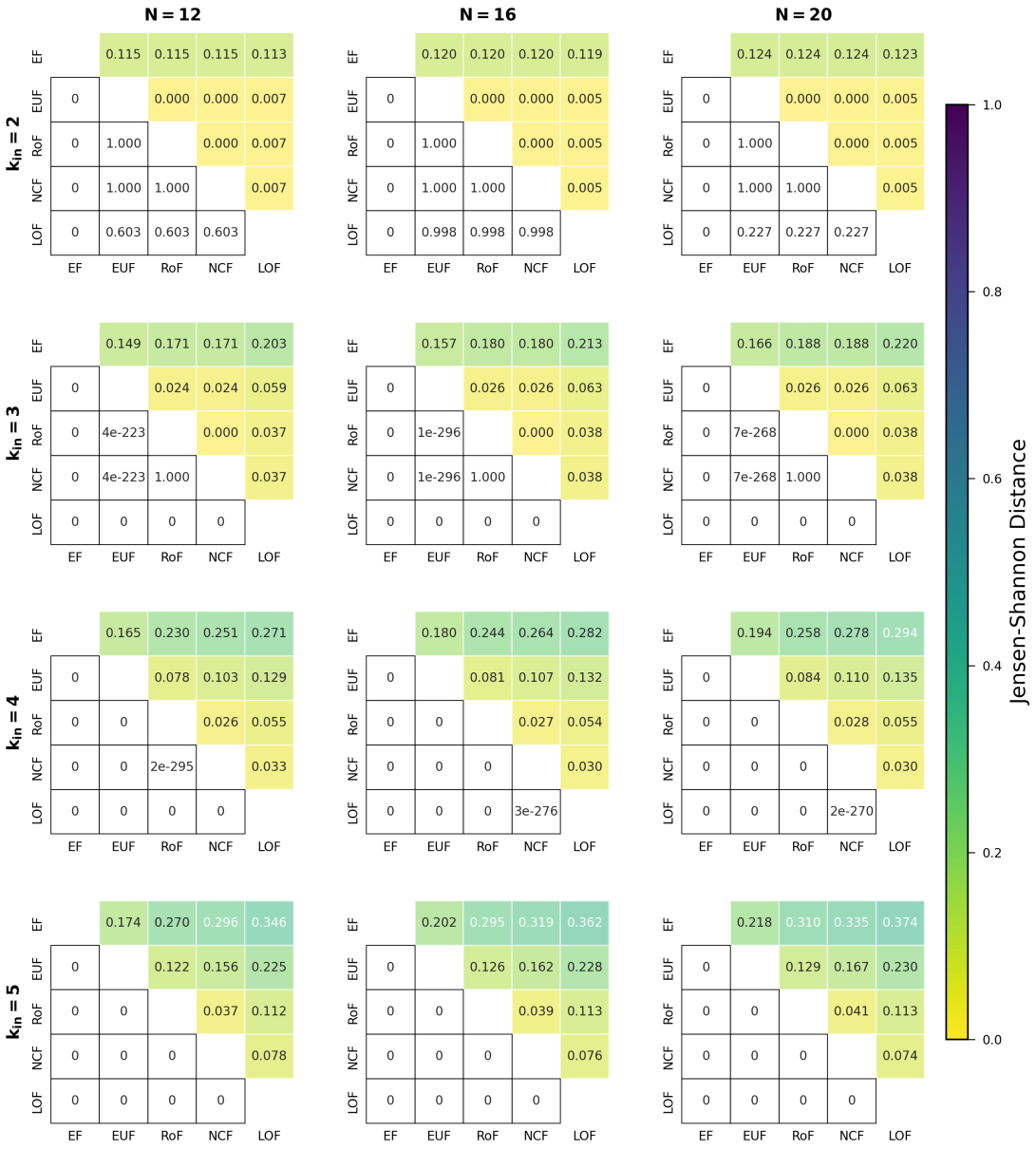


Figure A.14: Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the fraction of states in the basins of fixed points for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

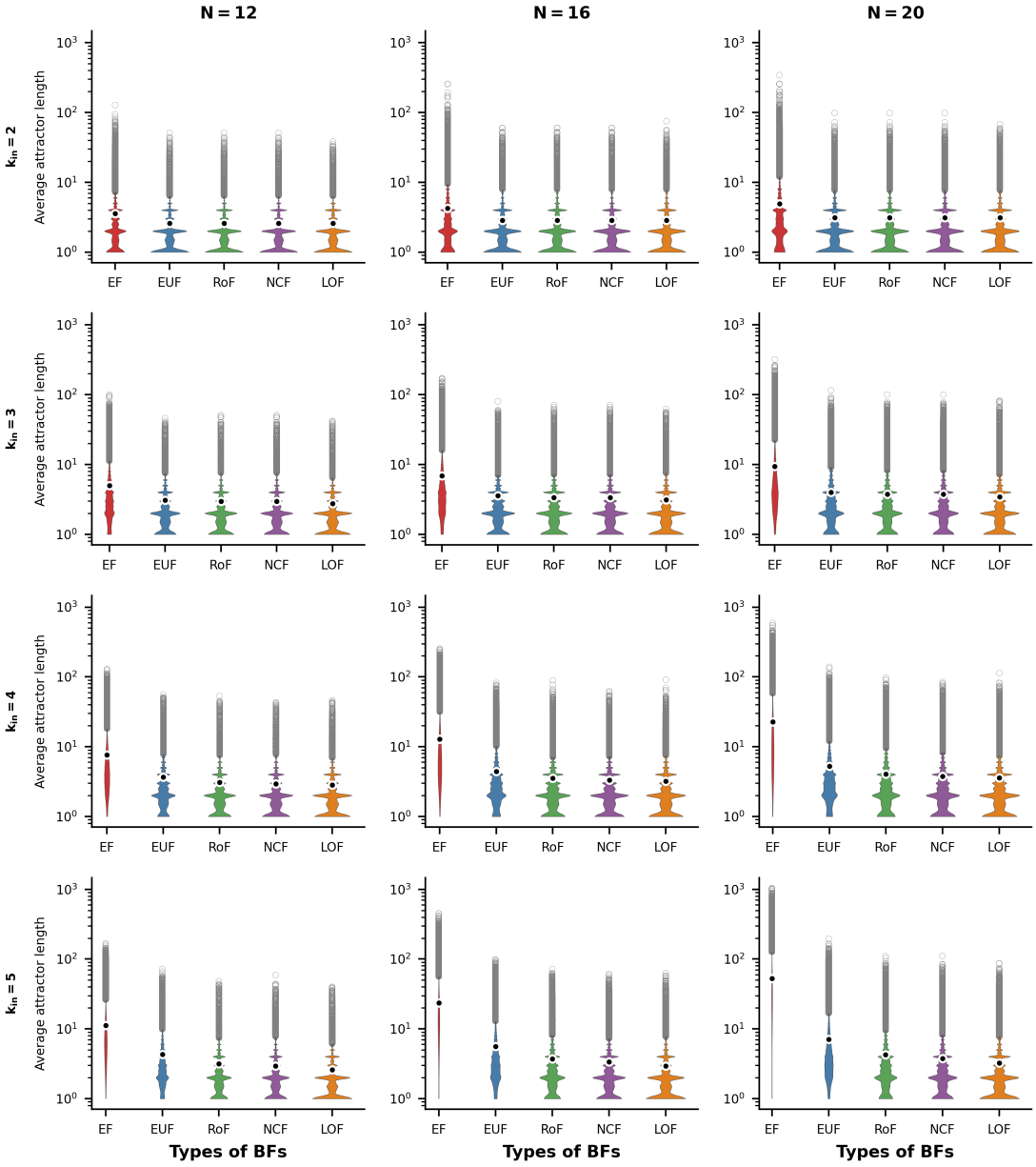


Figure A.15: Distributions of average attractor lengths for R-P networks. The violin plots display the distribution of the average attractor length for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

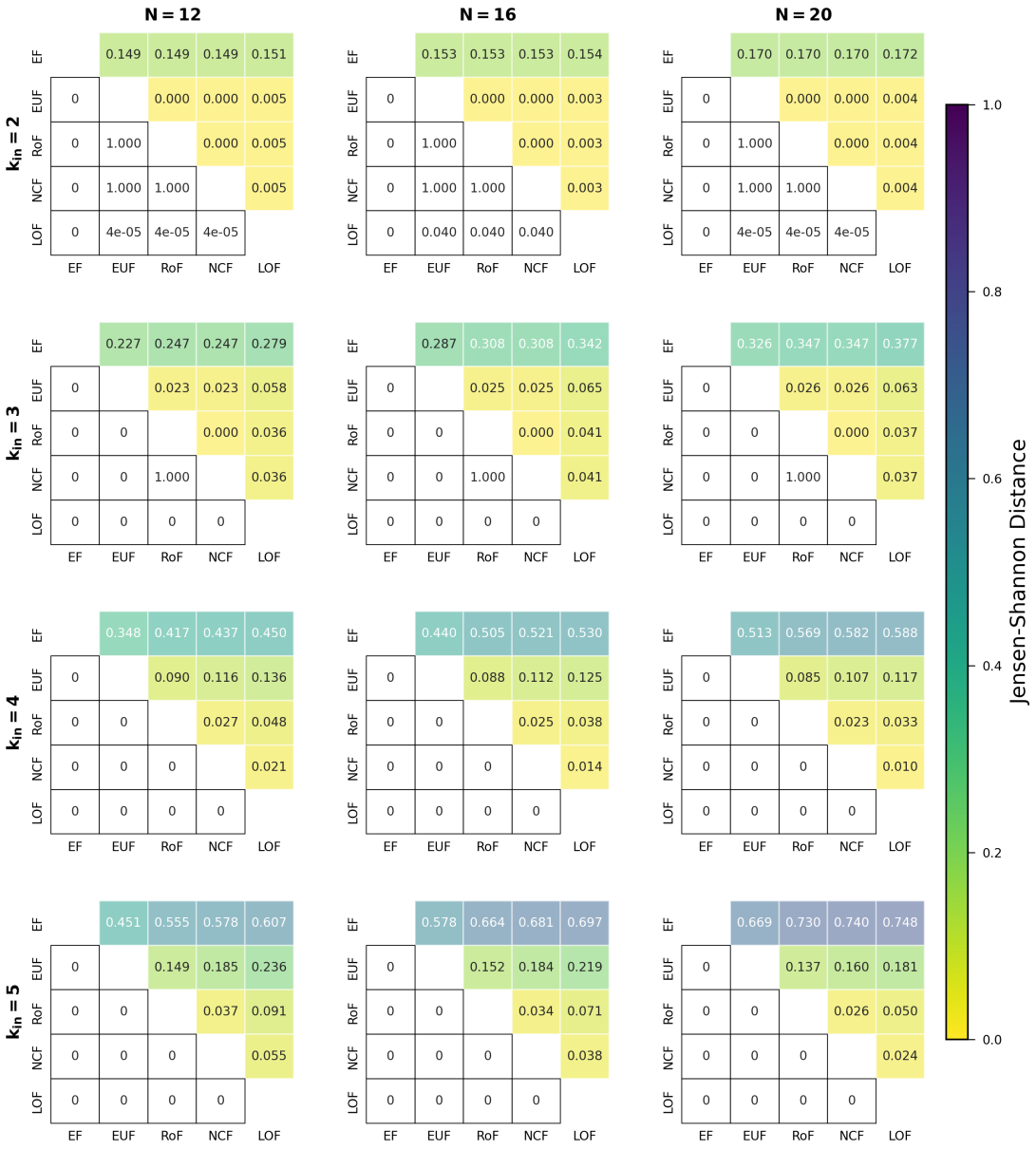


Figure A.16: Comparing distributions for the average attractor lengths when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

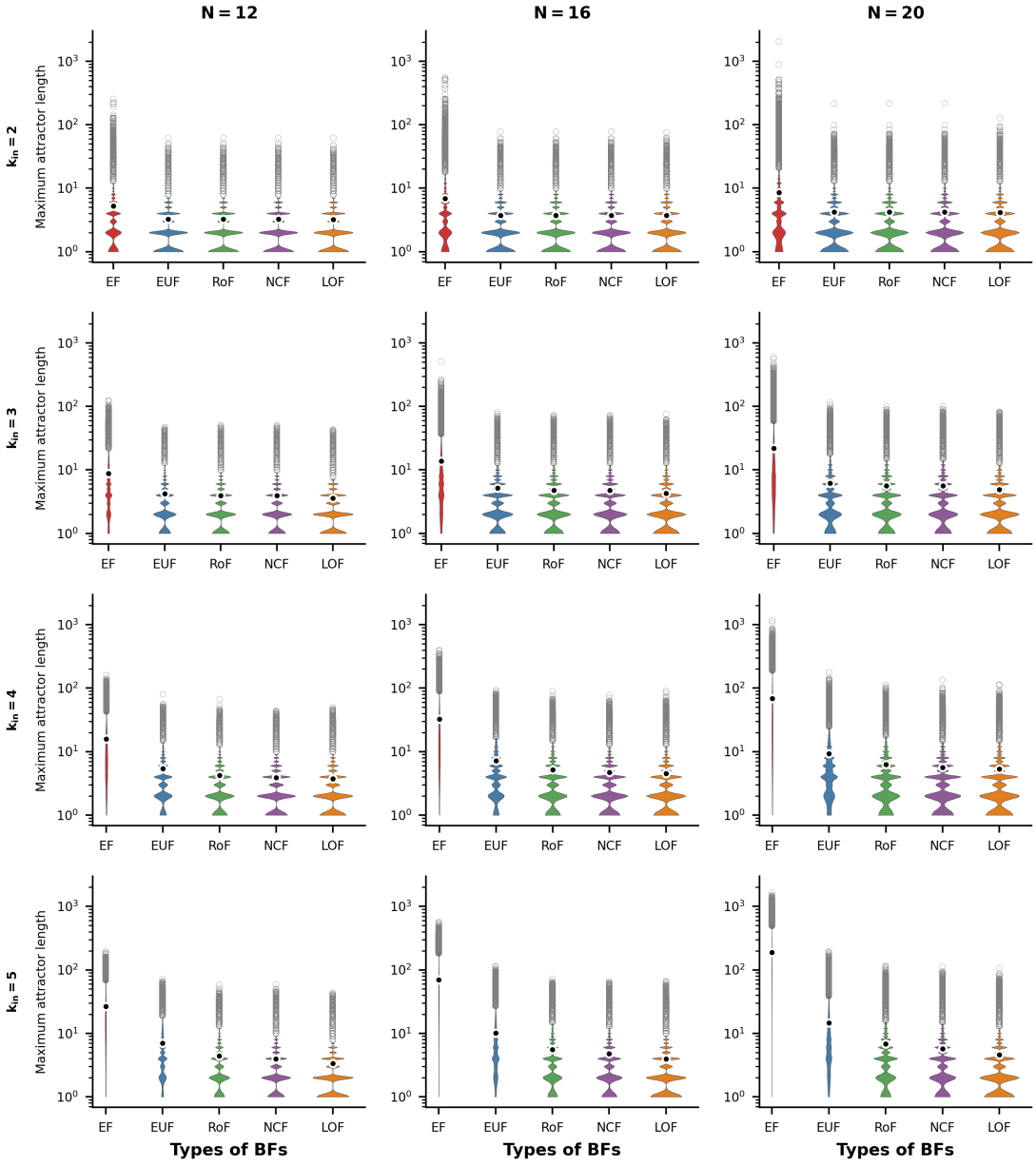


Figure A.17: Distributions of maximum attractor lengths for R-P networks. The violin plots display the distribution of the maximum attractor length for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the maximum attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

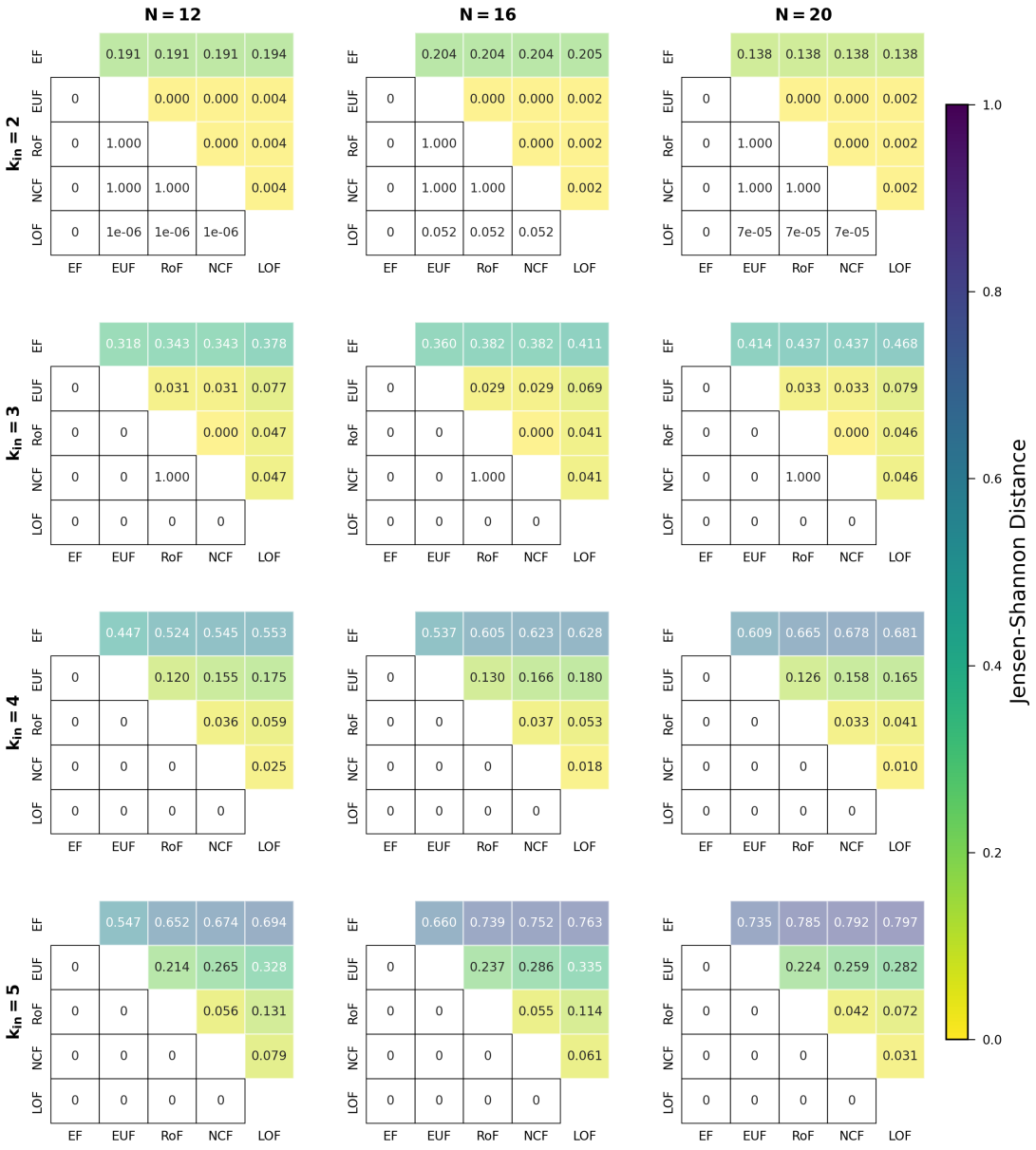


Figure A.18: Comparing distributions for the maximum attractor lengths when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the maximum attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

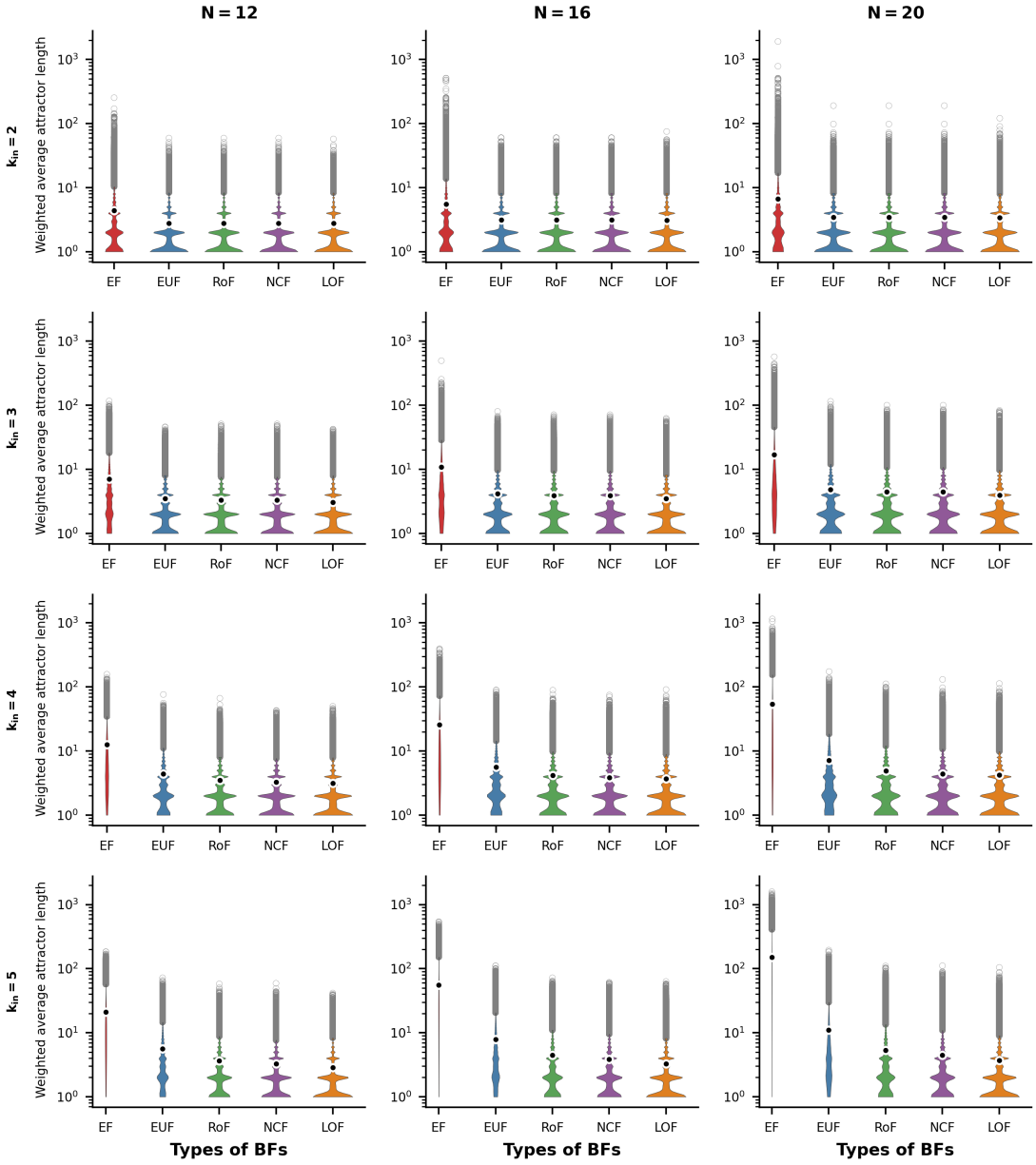


Figure A.19: Distributions of weighted average attractor lengths for R-P networks. The violin plots display the distribution of the weighted average attractor length for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the weighted average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

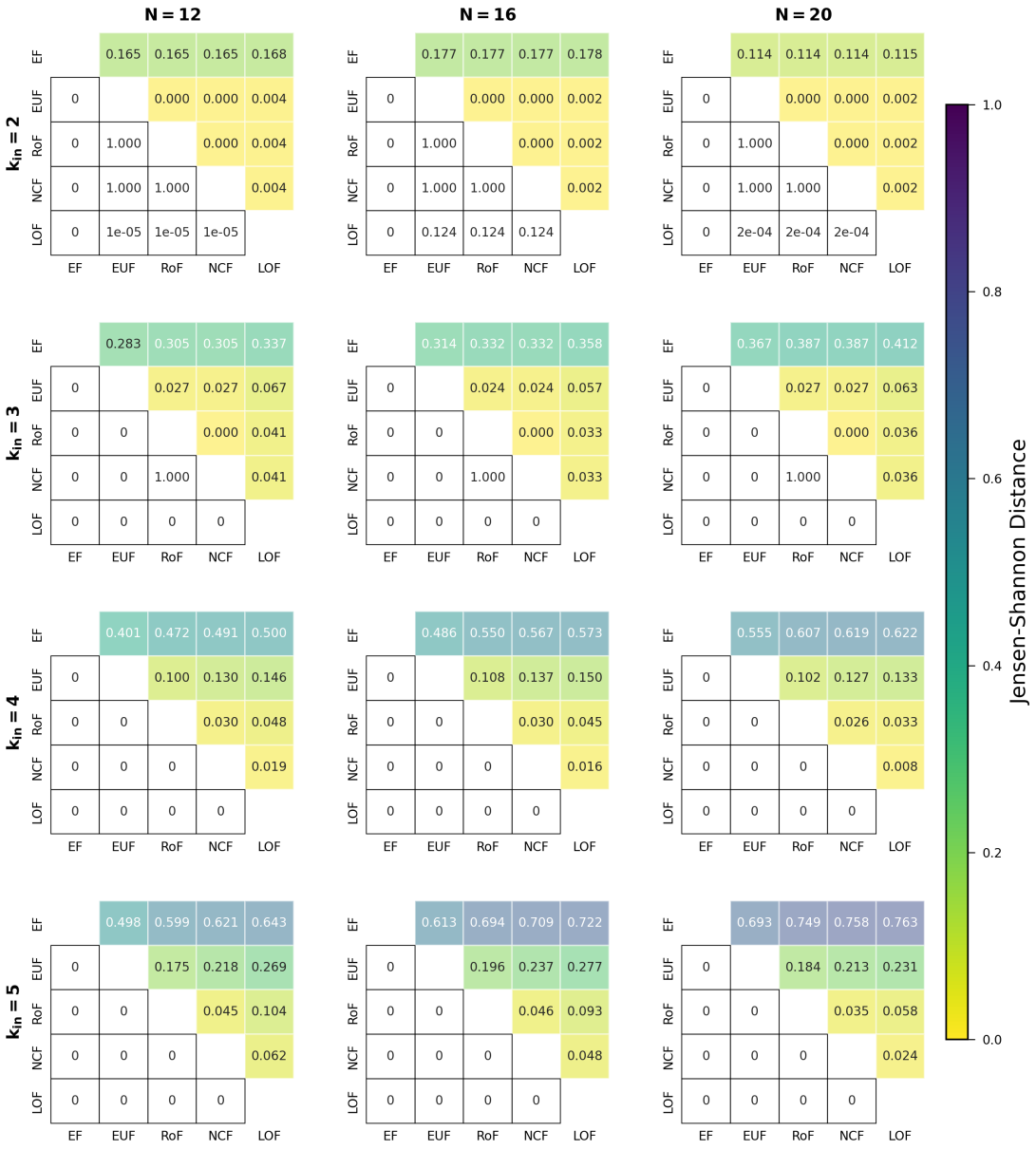


Figure A.20: Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the weighted average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

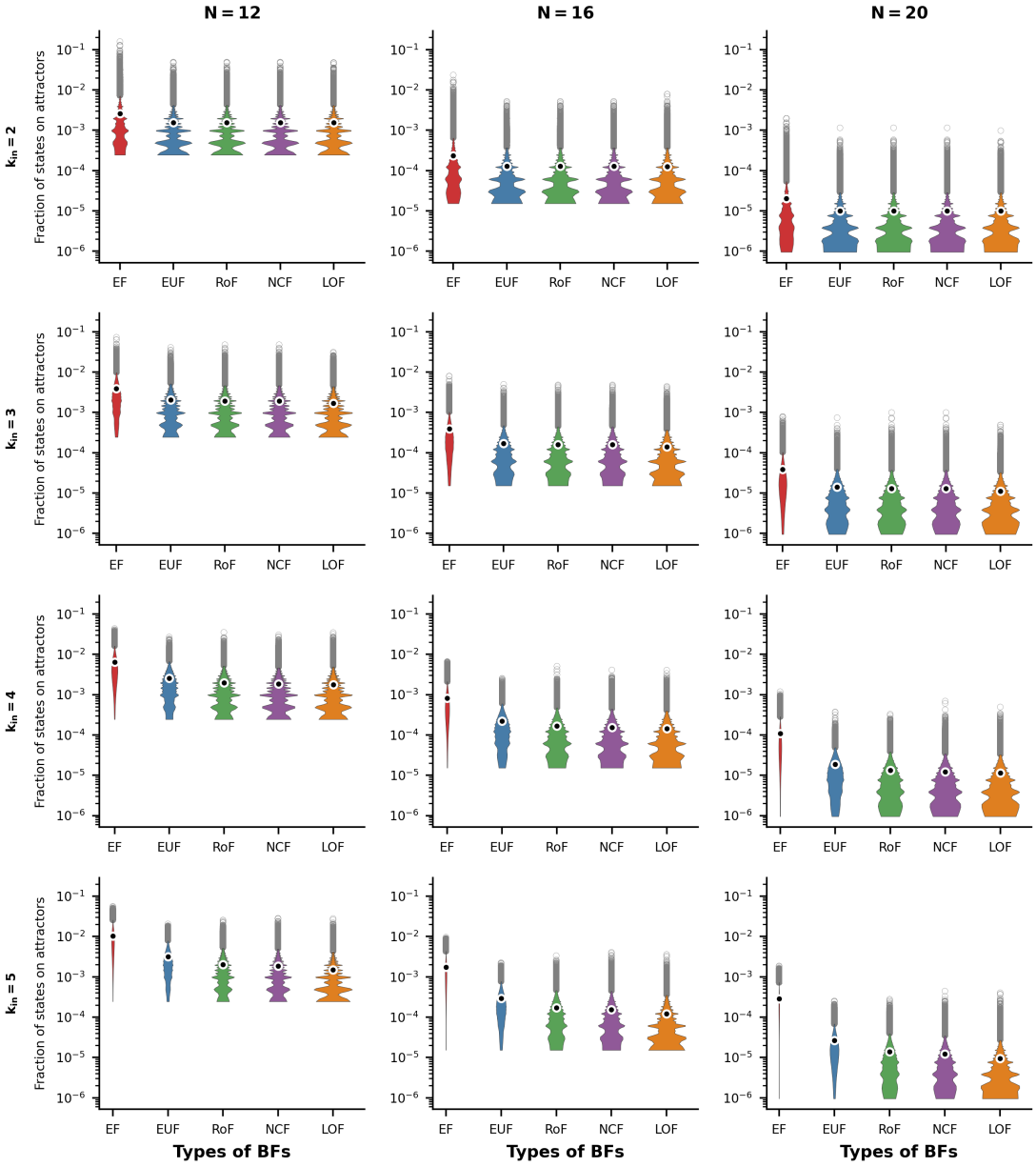


Figure A.21: Distributions of fraction of states on attractors for R-P networks. The violin plots display the distribution of the fraction of states on attractors for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states on attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

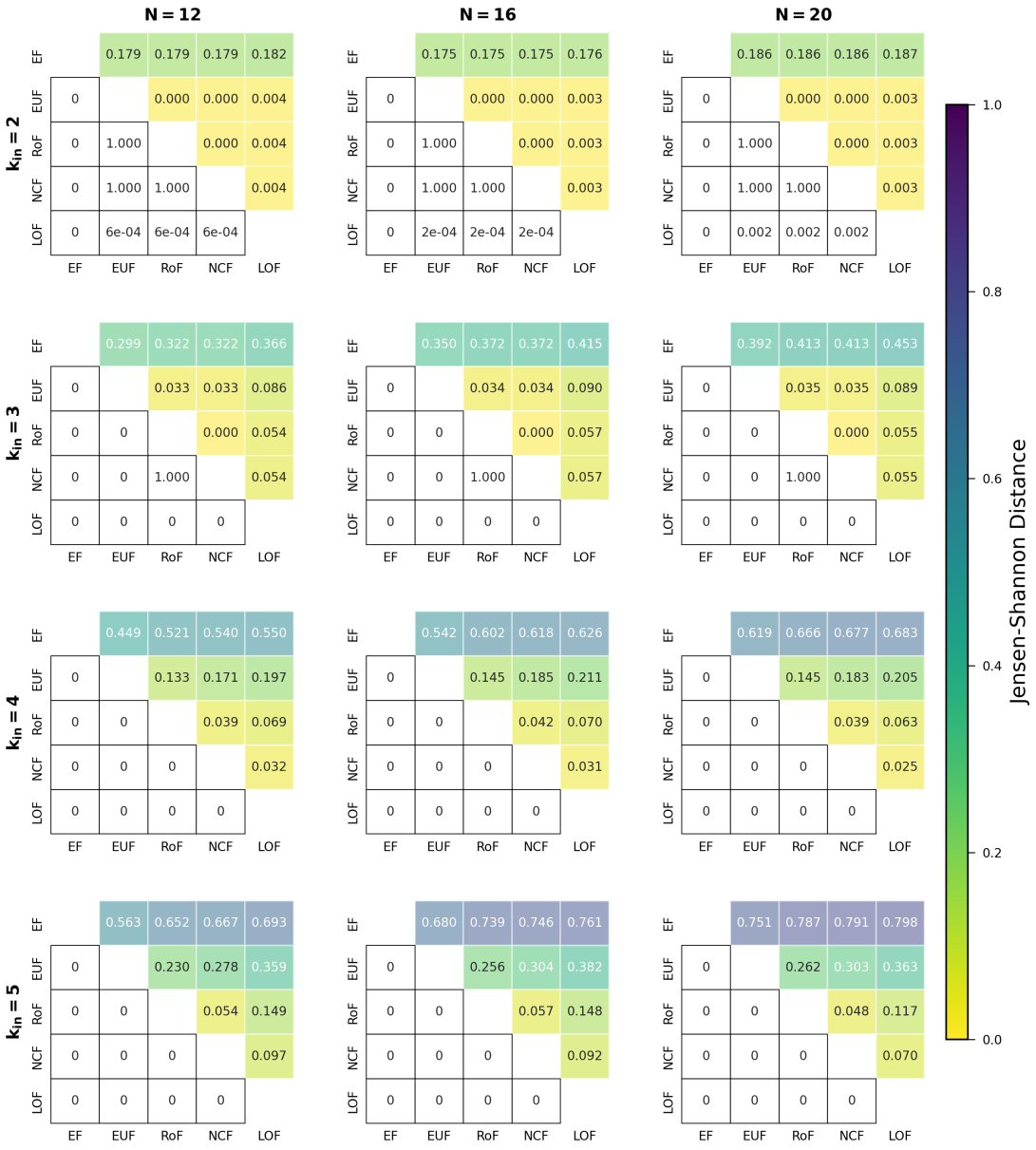


Figure A.22: Comparing distributions for the fraction of states on attractors when using different classes of BFs within the R-P networks. The different sub-figures are organised according to the values of k_{in} (the in-degree) and N (the number of nodes in the network). At given k_{in} and N, we determined the distribution of the fraction of states on attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

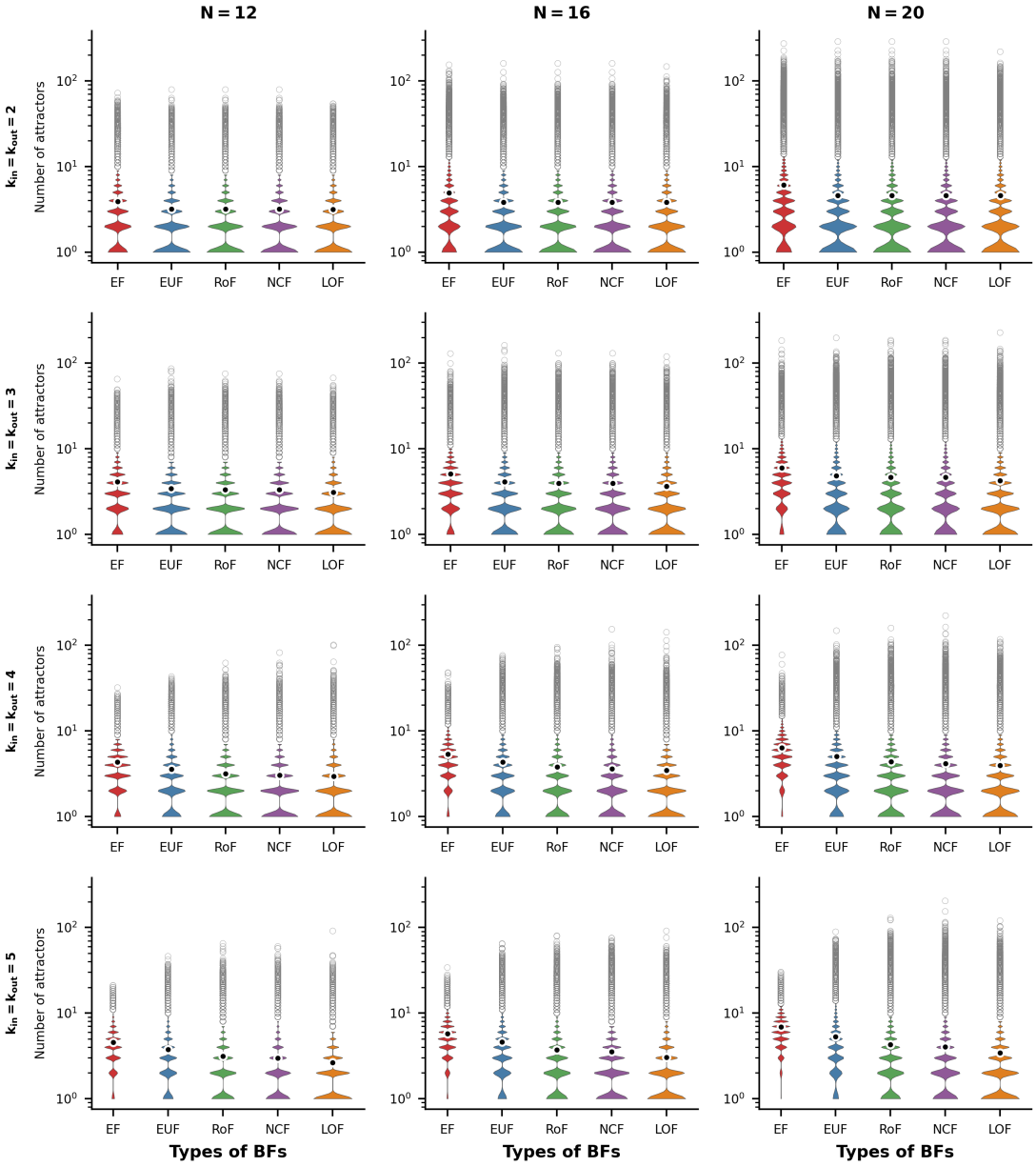


Figure A.23: Distributions of number of attractors for R-R networks. The violin plots display the distribution of the number of attractors for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12$, 16 and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the number of attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

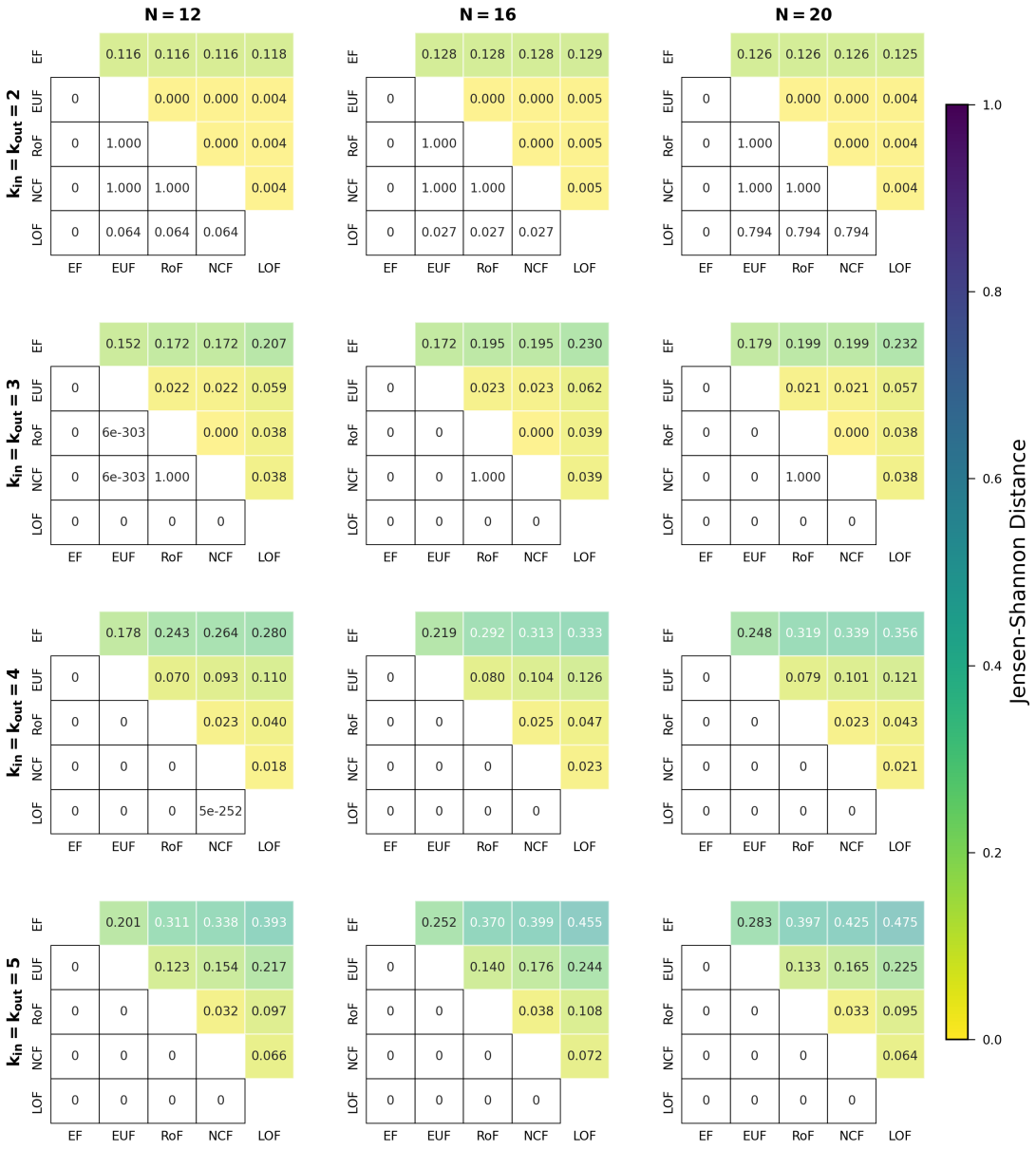


Figure A.24: Comparing distributions for the number of attractors when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the number of attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

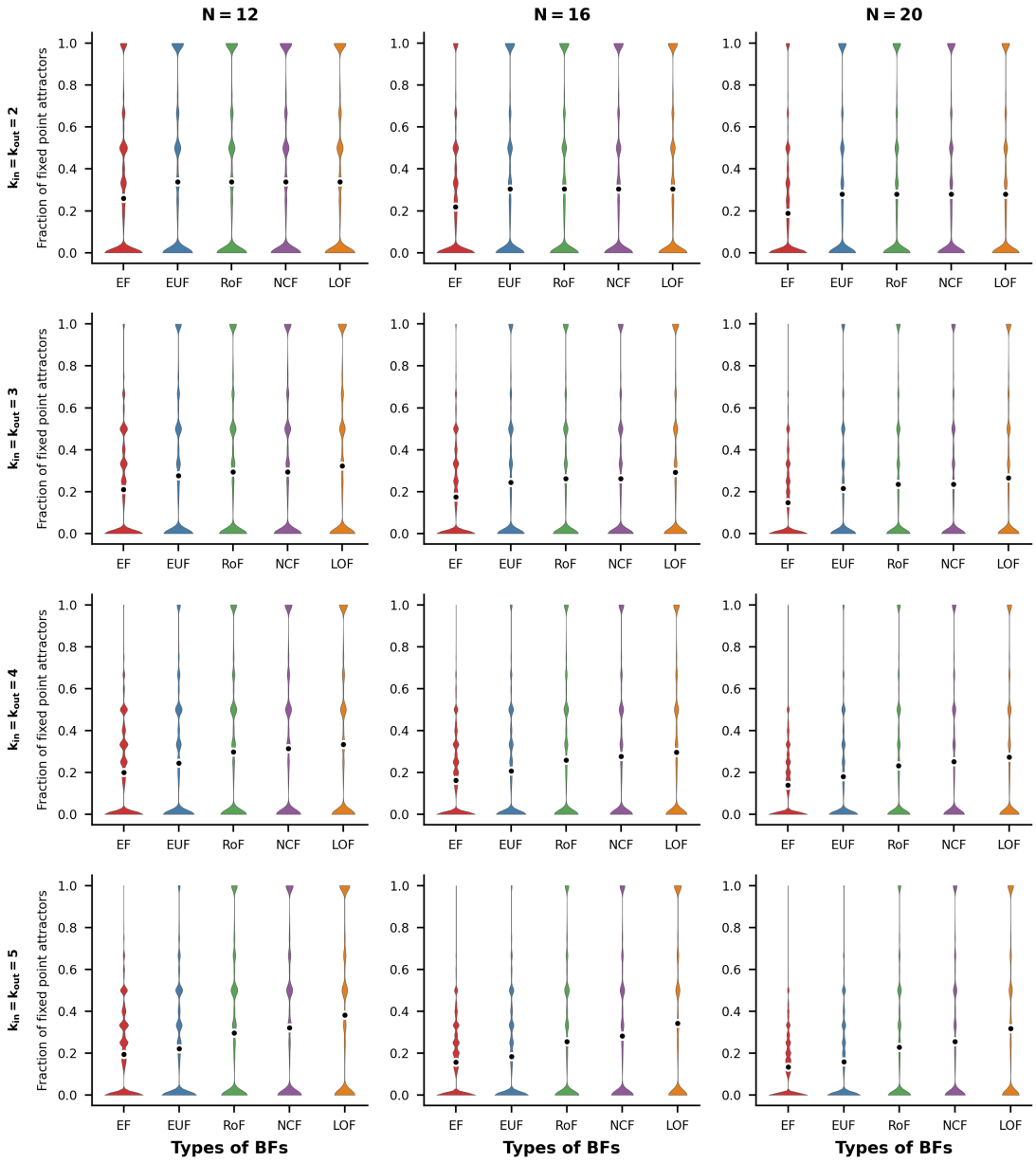


Figure A.25: Distributions of fraction of fixed point attractors for R-R networks. The violin plots display the distribution of the fraction of fixed point attractors for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

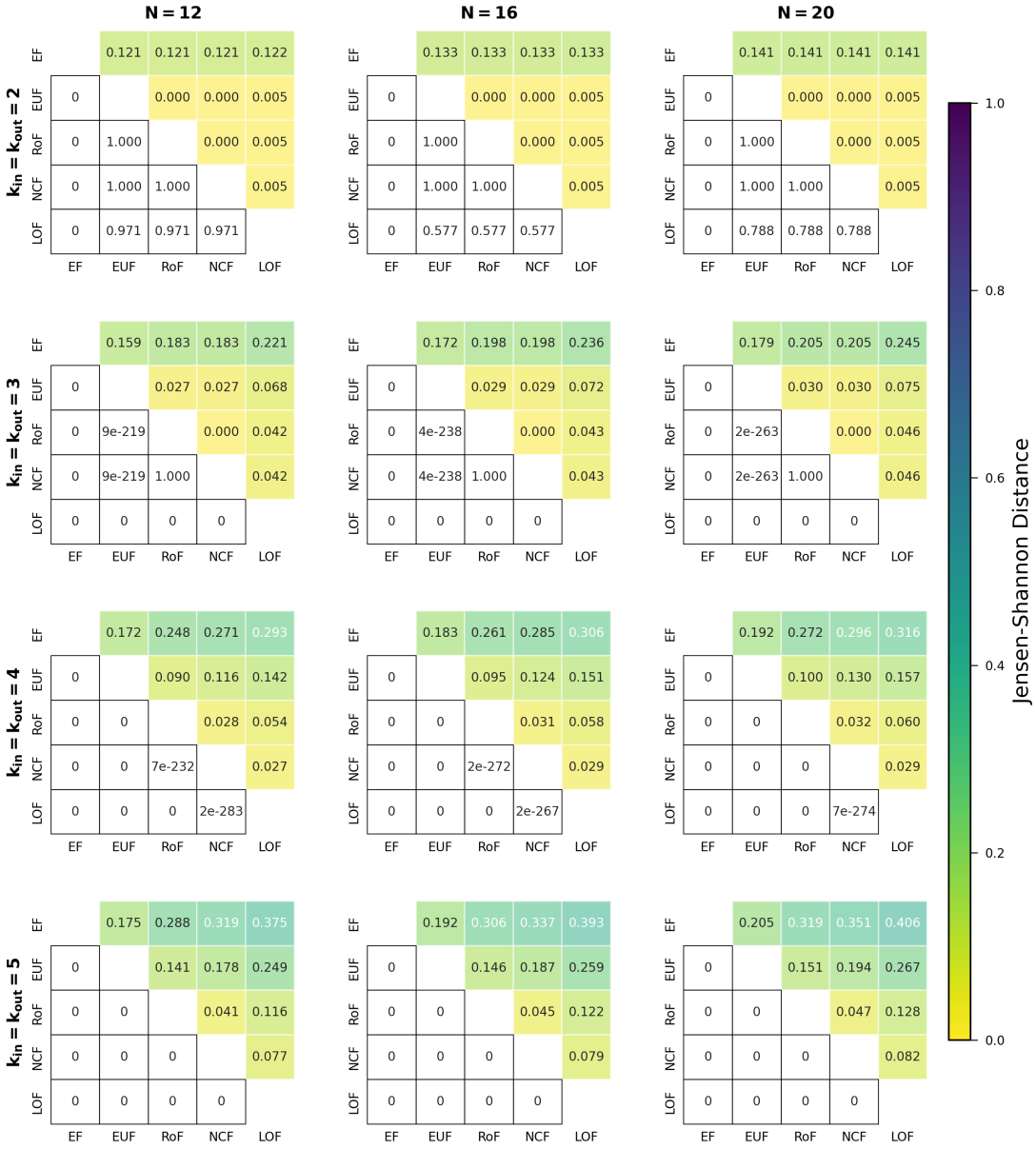


Figure A.26: Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the fraction of fixed point attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

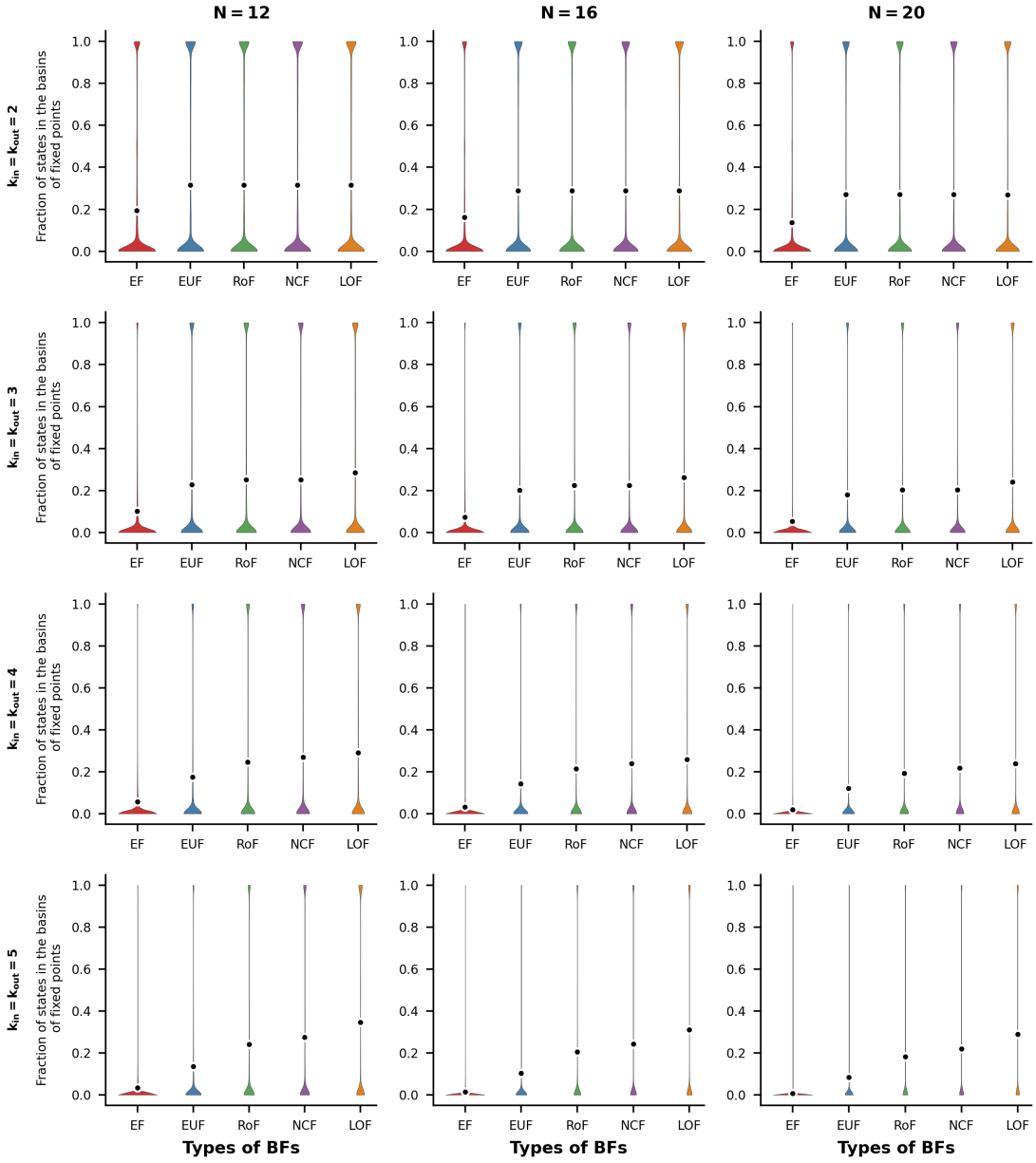


Figure A.27: Distributions of fraction of states in the basins of fixed points for R-R networks. The violin plots display the distribution of the fraction of states in the basins of fixed point attractors for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states in the basins of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

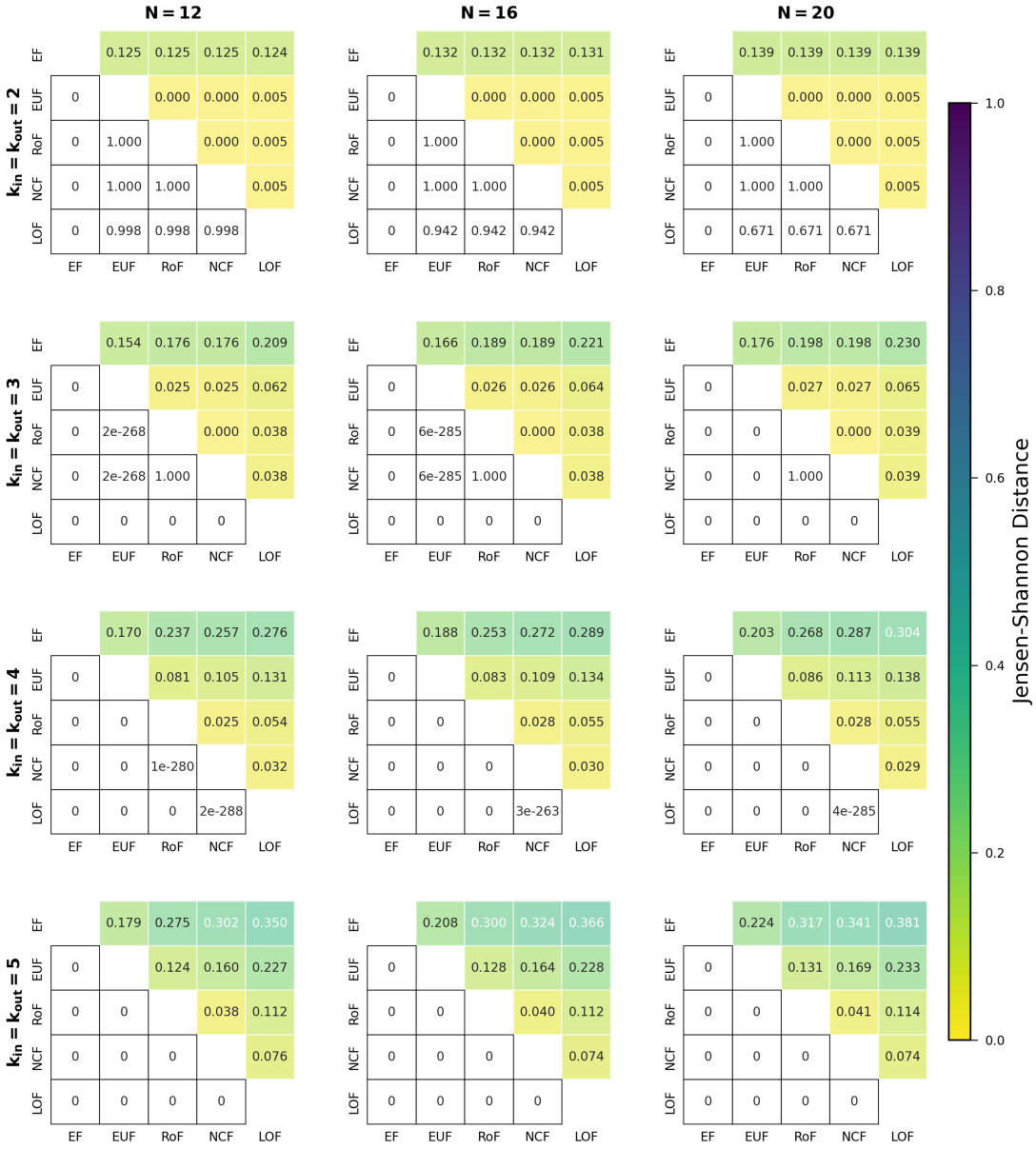


Figure A.28: Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the fraction of states in the basins of fixed points for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

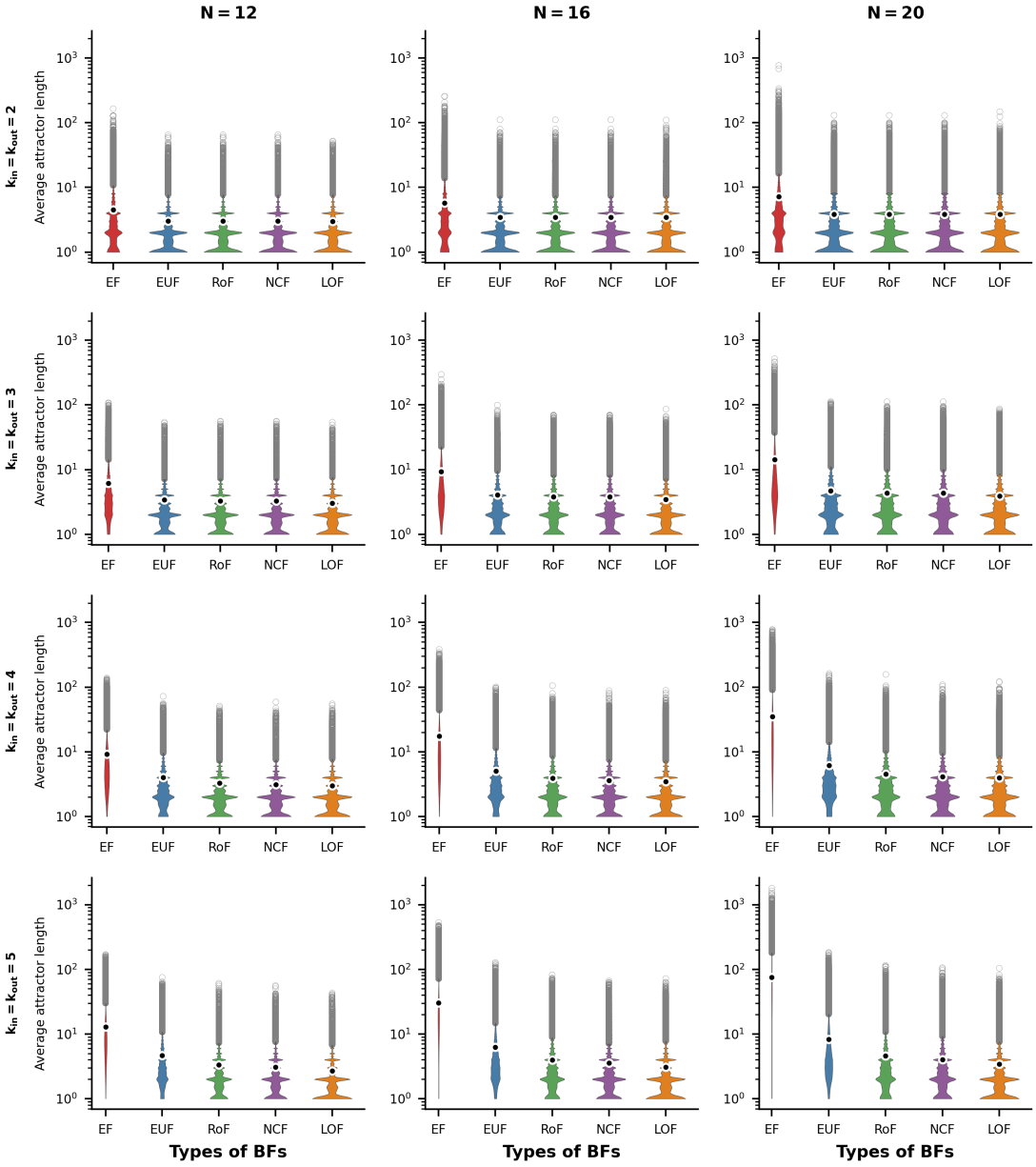


Figure A.29: Distributions of average attractor lengths for R-R networks. The violin plots display the distribution of the average attractor length for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

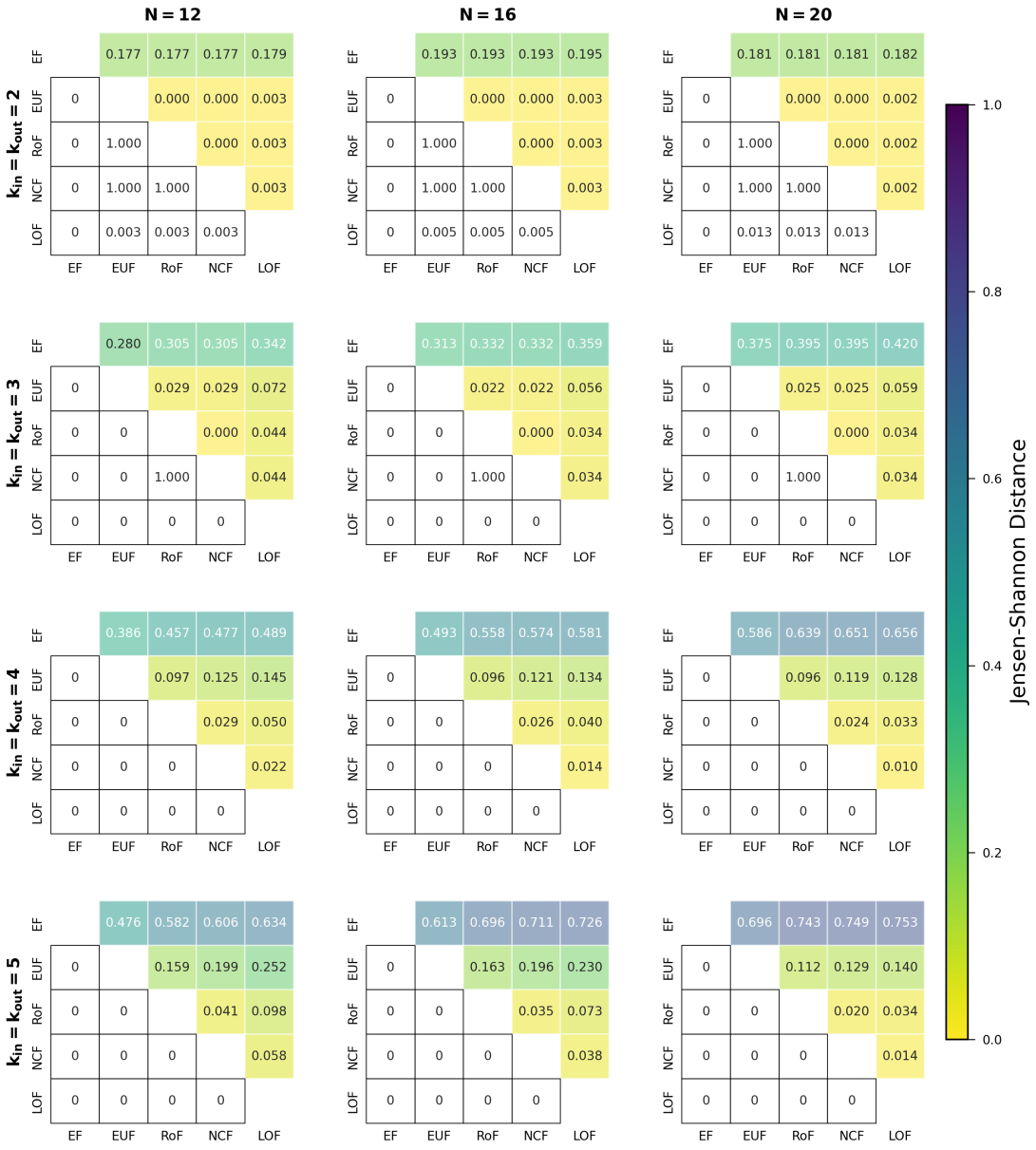


Figure A.30: Comparing distributions for the average attractor lengths when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

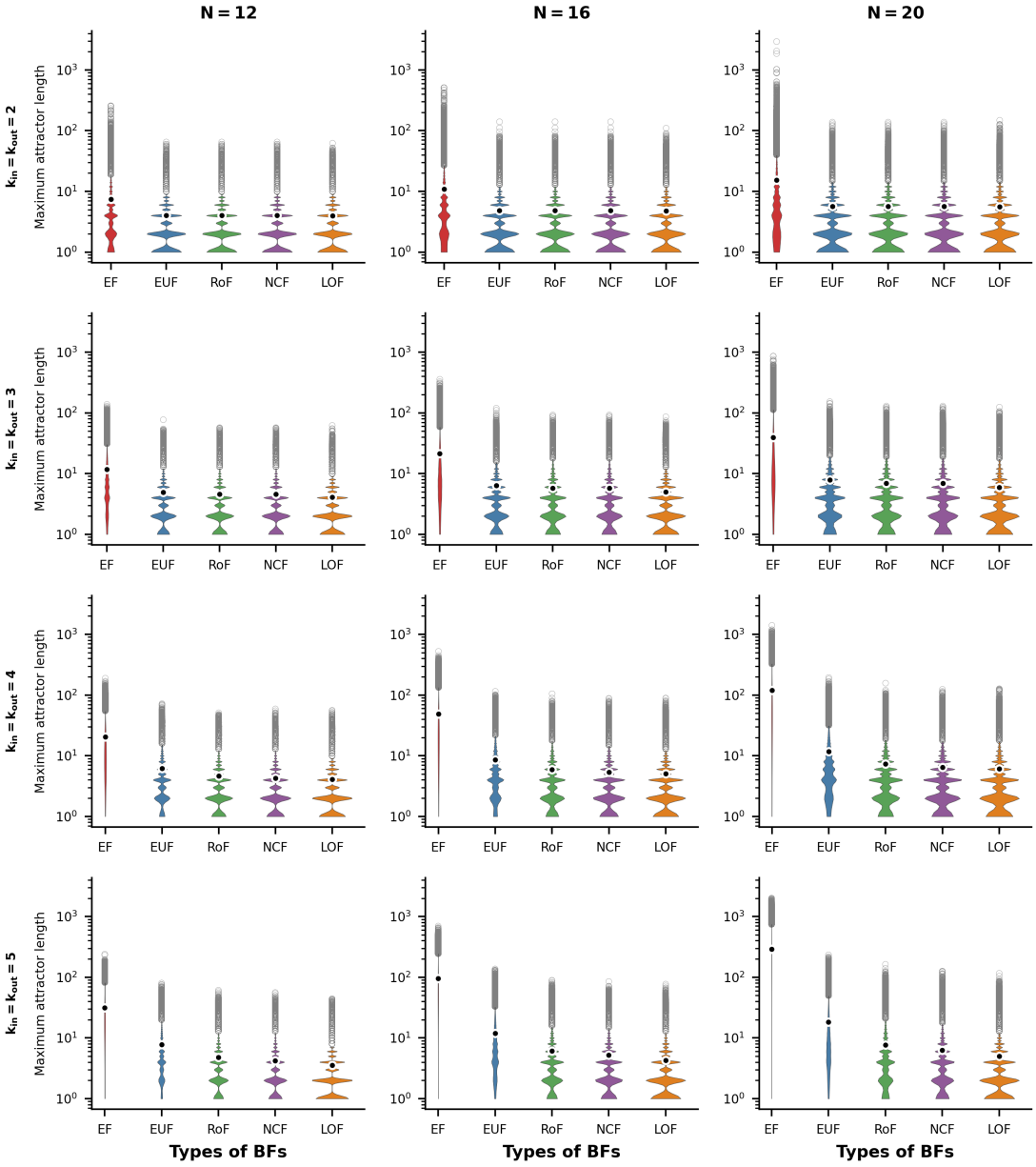


Figure A.31: Distributions of maximum attractor lengths for R-R networks. The violin plots display the distribution of the maximum attractor length for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the maximum attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

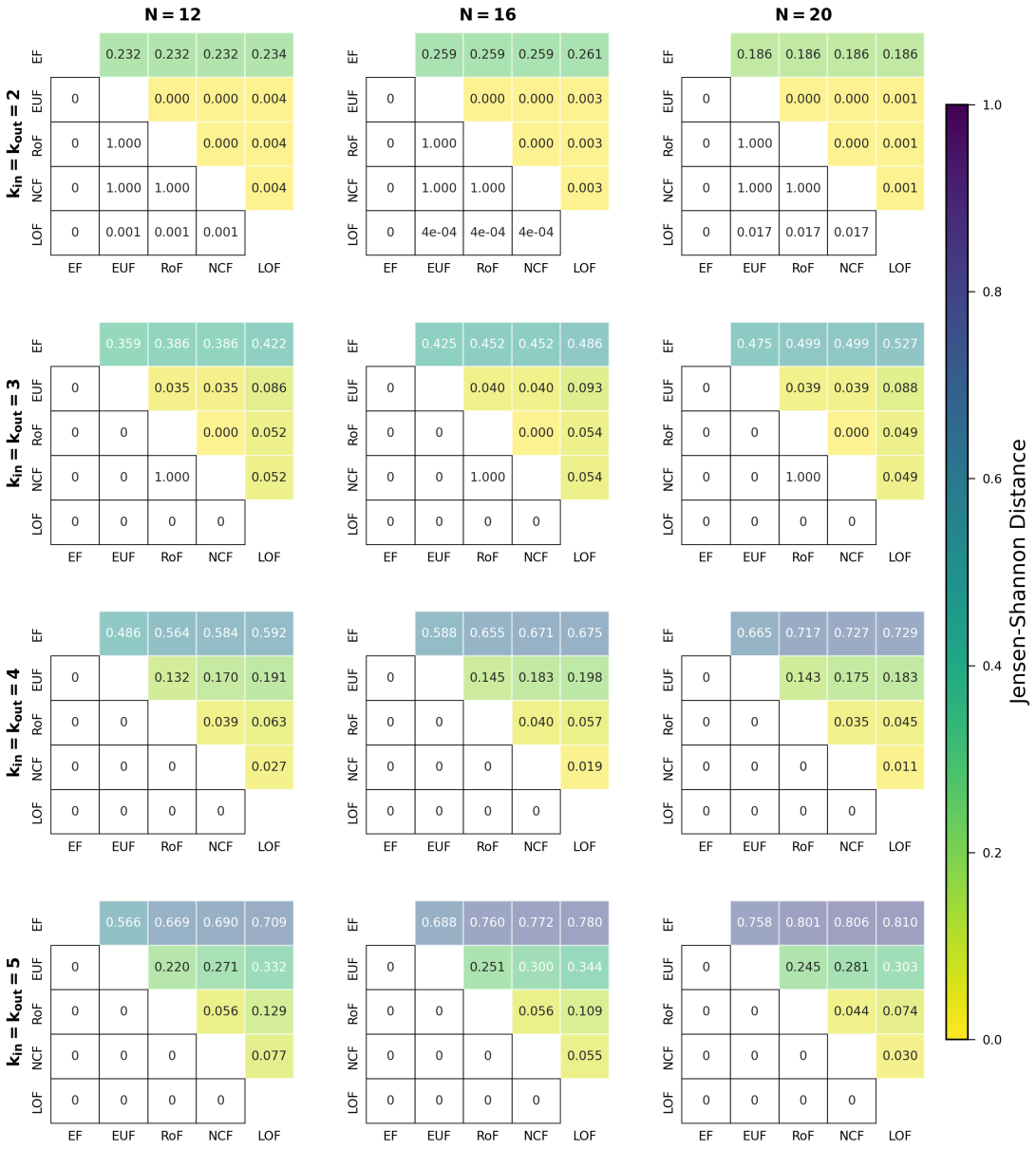


Figure A.32: Comparing distributions for the maximum attractor lengths when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the maximum attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

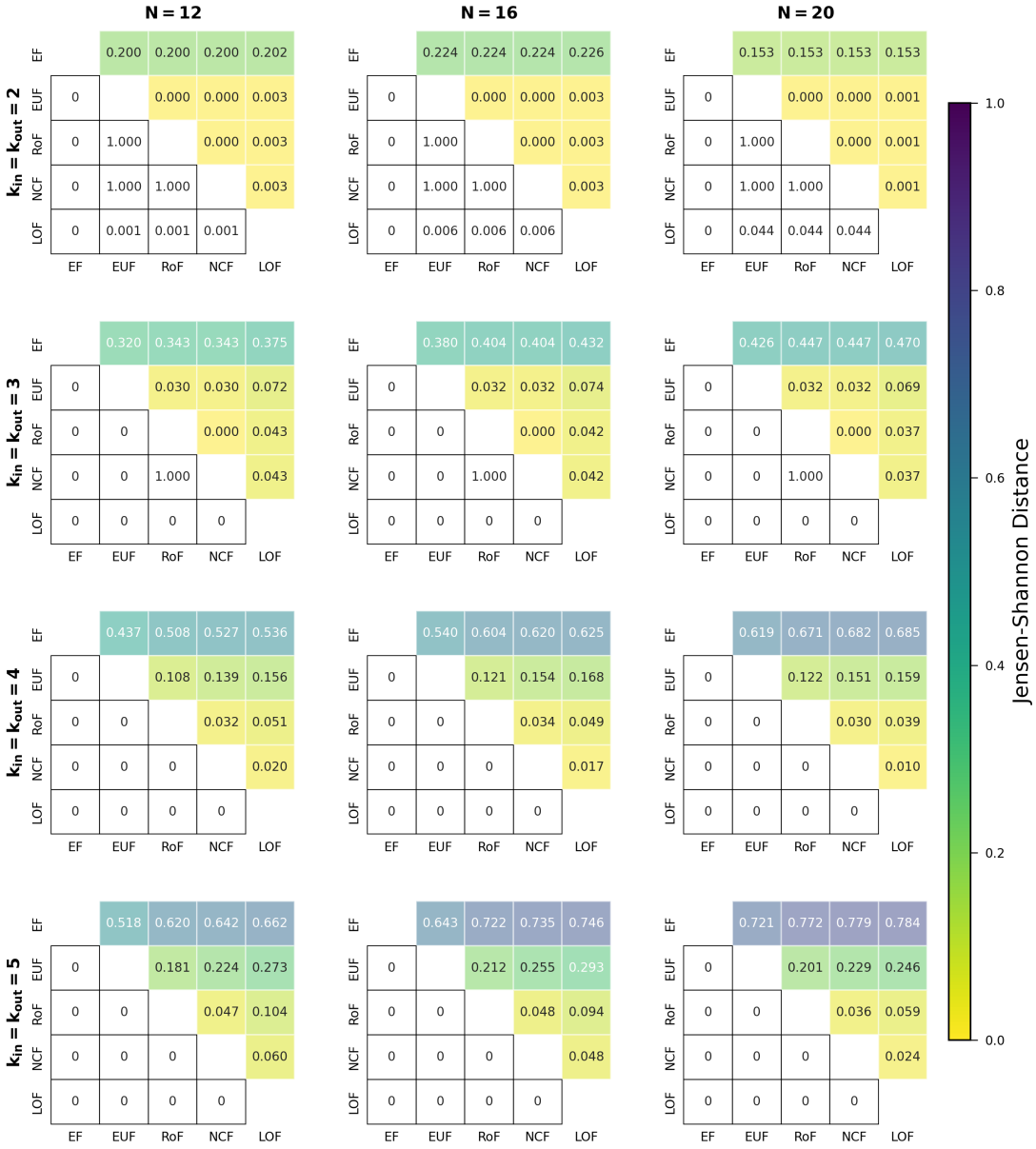


Figure A.33: Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the weighted average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

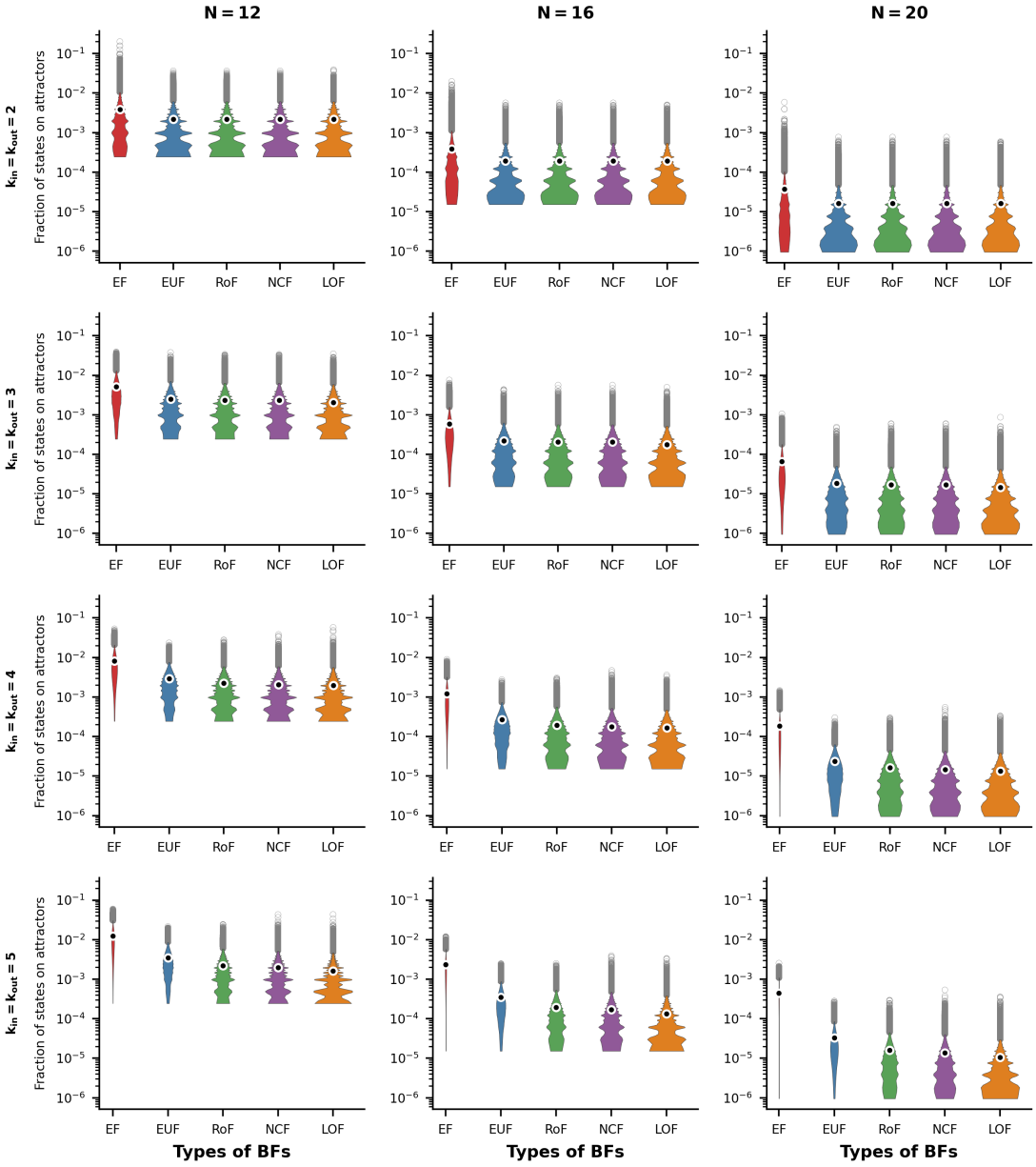


Figure A.34: Distributions of fraction of states on attractors for R-R networks. The violin plots display the distribution of the fraction of states on attractors for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different number of inputs (and outputs) per node, ranging from 2 to 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states on attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

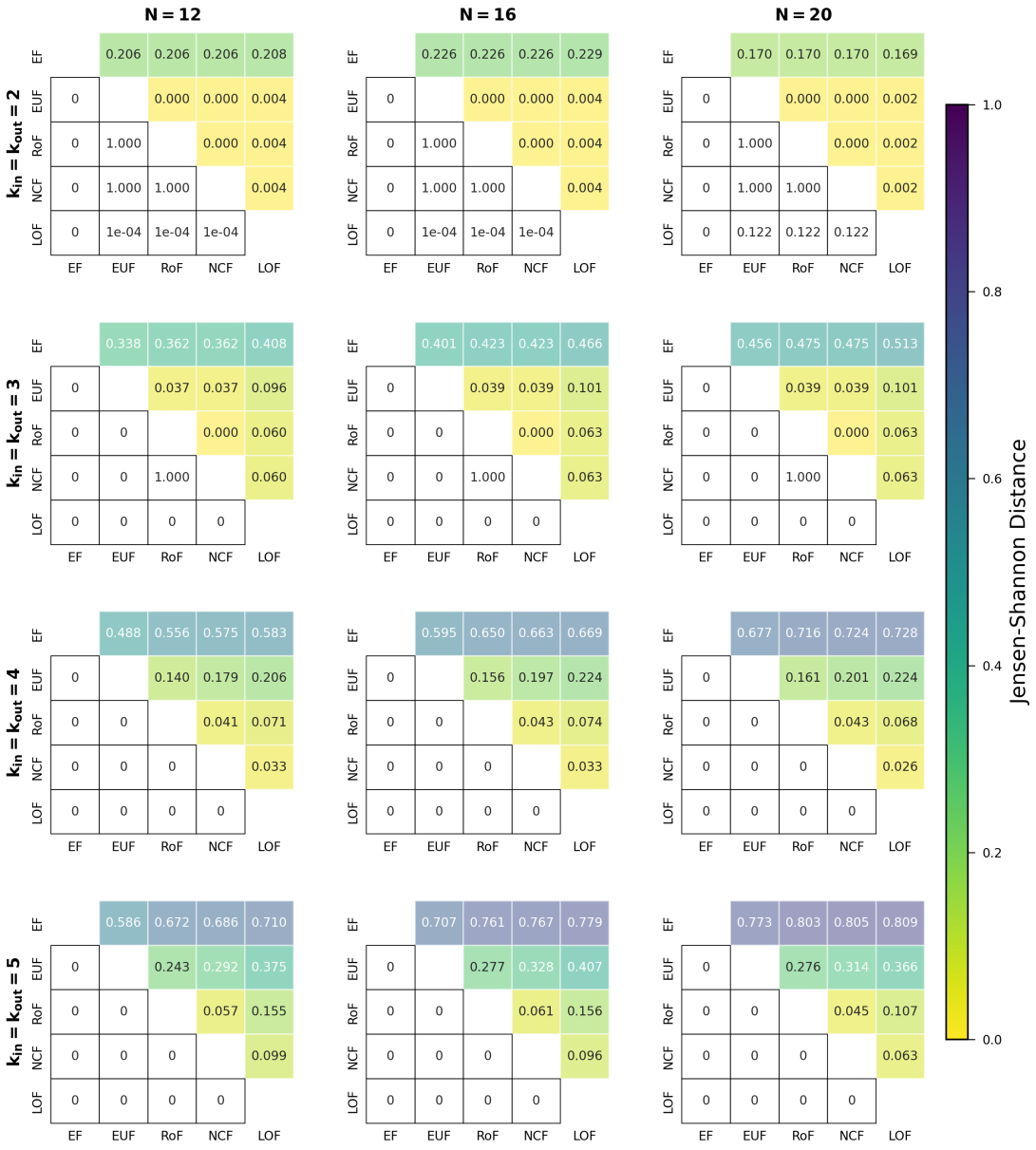


Figure A.35: Comparing distributions for the fraction of states on attractors when using different classes of BFs within the R-R networks. The different sub-figures are organised according to the values of the in-degree and out-degree (k_{in} and k_{out} both equal), and N (the number of nodes in the network). At given k_{in} (or k_{out}) and N, we determined the distribution of the fraction of states on attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

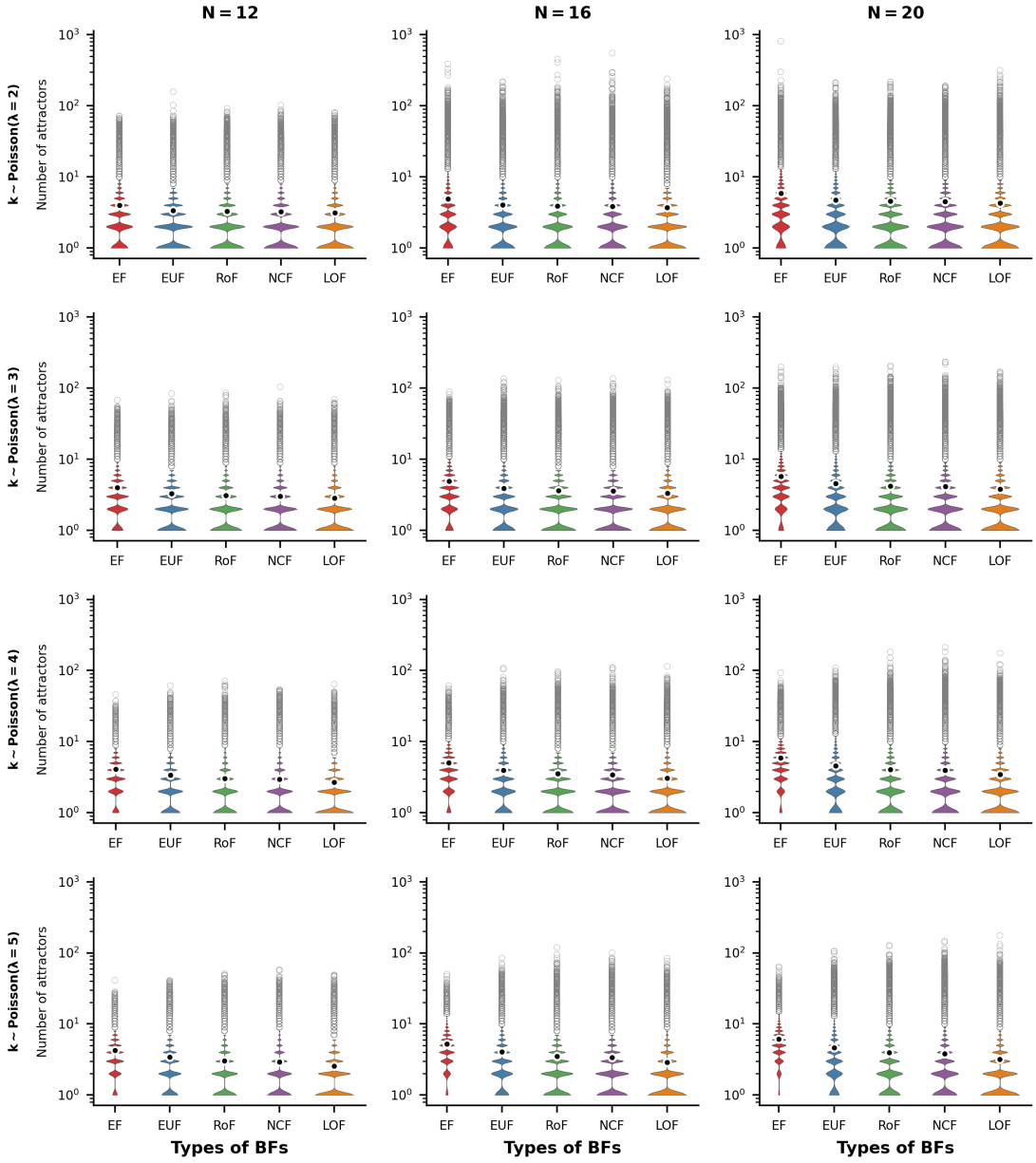


Figure A.36: Distributions of number of attractors for P-P networks. The violin plots display the distribution of the number of attractors for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the number of attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

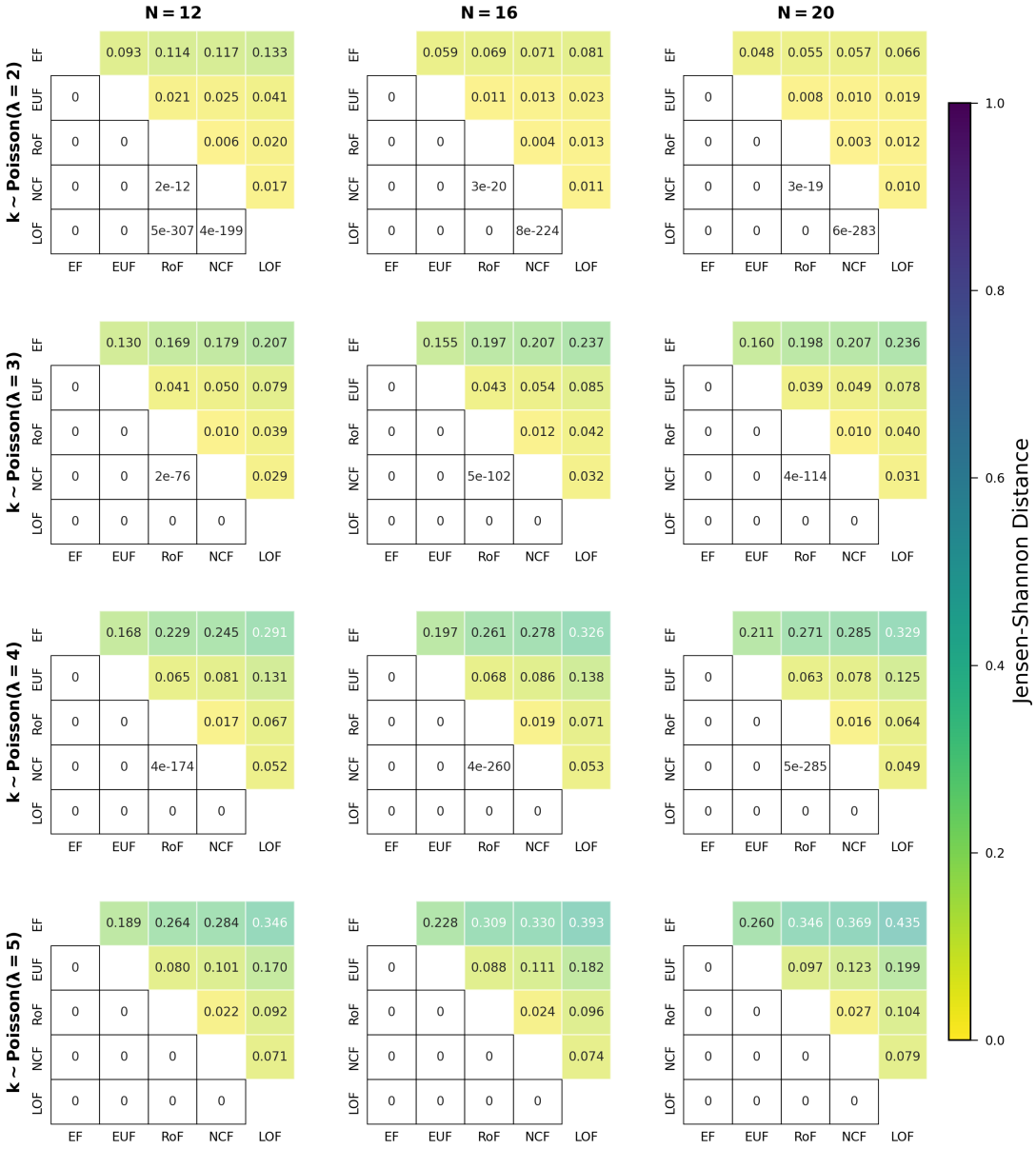


Figure A.37: Comparing distributions for the number of attractors when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean = λ. At a given λ and N, we determined the distribution of the number of attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

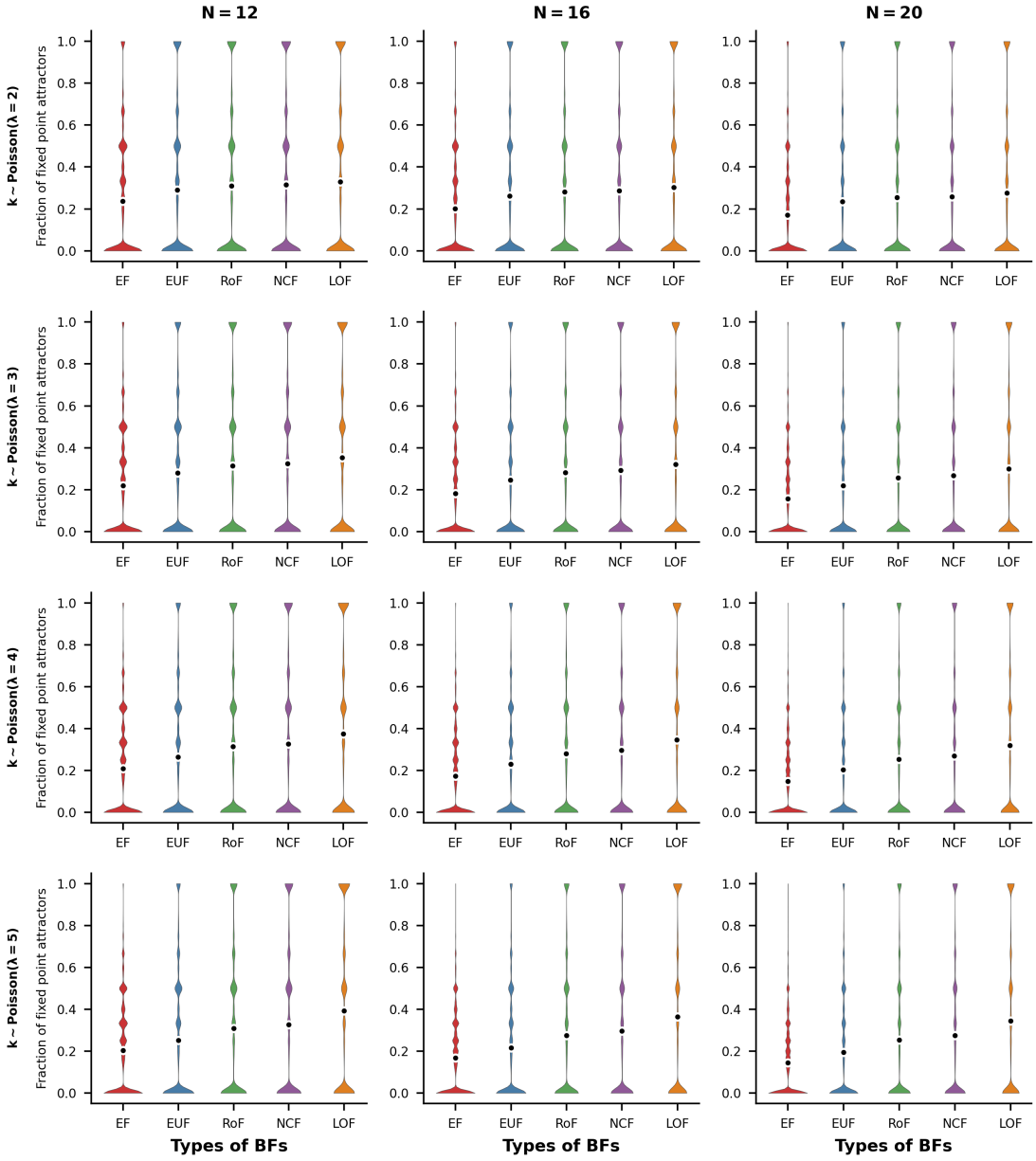


Figure A.38: Distributions of fraction of fixed point attractors for P-P networks. The violin plots display the distribution of the fraction of fixed point attractors for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the fraction of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

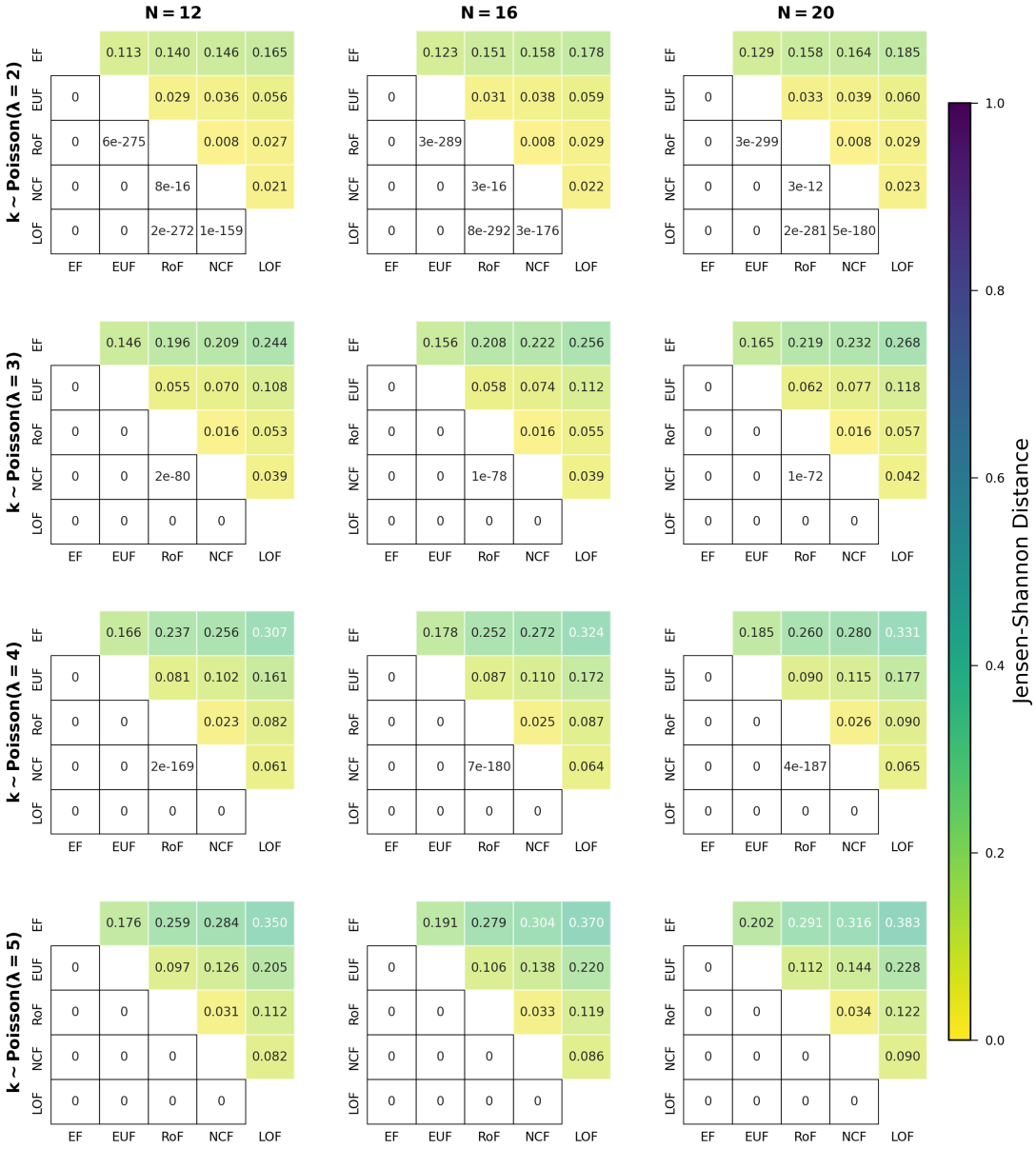


Figure A.39: Comparing distributions for the fraction of fixed point attractors when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean $= \lambda$. At a given λ and N, we determined the distribution of the fraction of fixed point attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

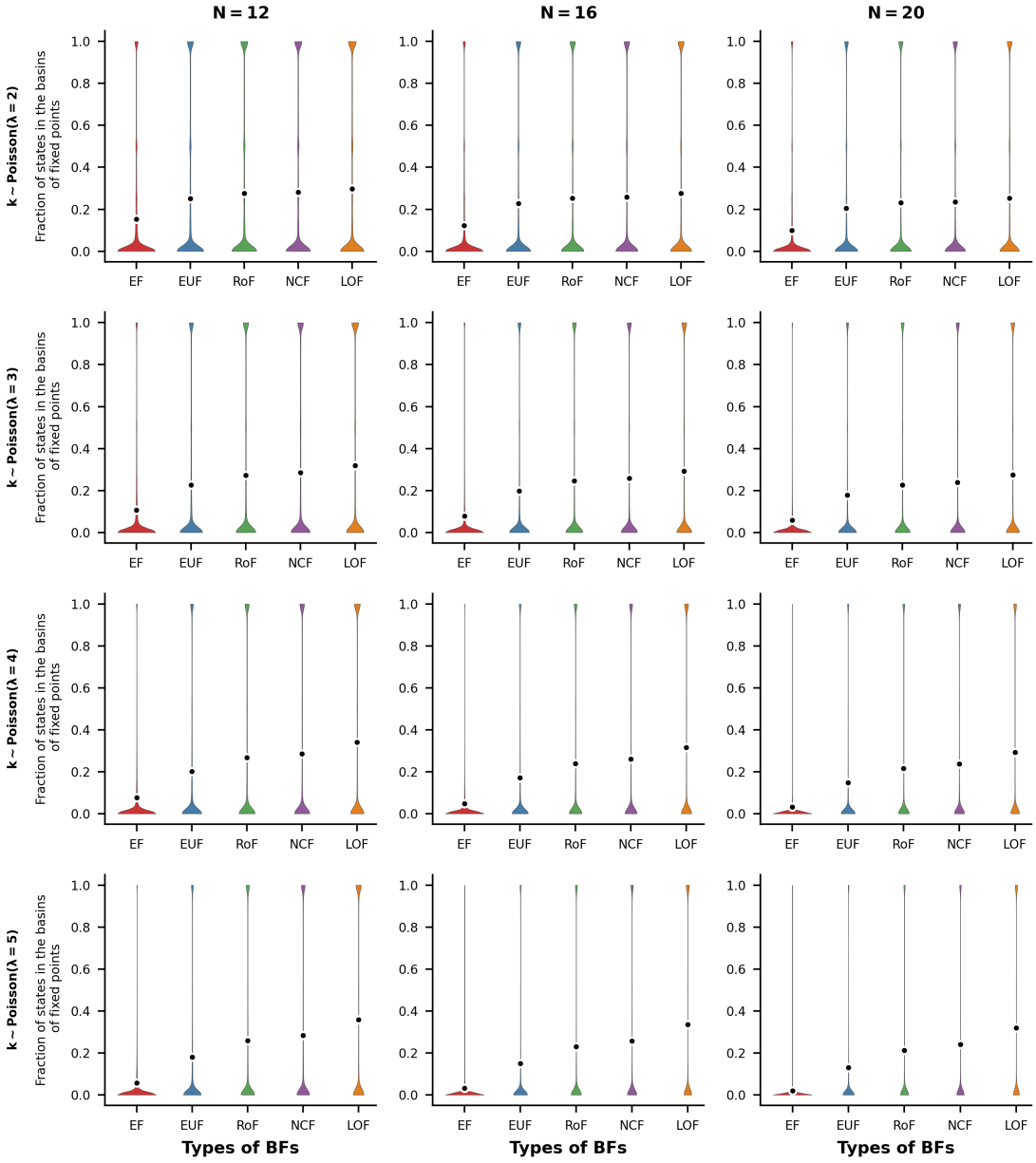


Figure A.40: Distributions of fraction of states in the basins of fixed points for P-P networks. The violin plots display the distribution of the fraction of states in the basins of fixed point attractors for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states in the basins of fixed point attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

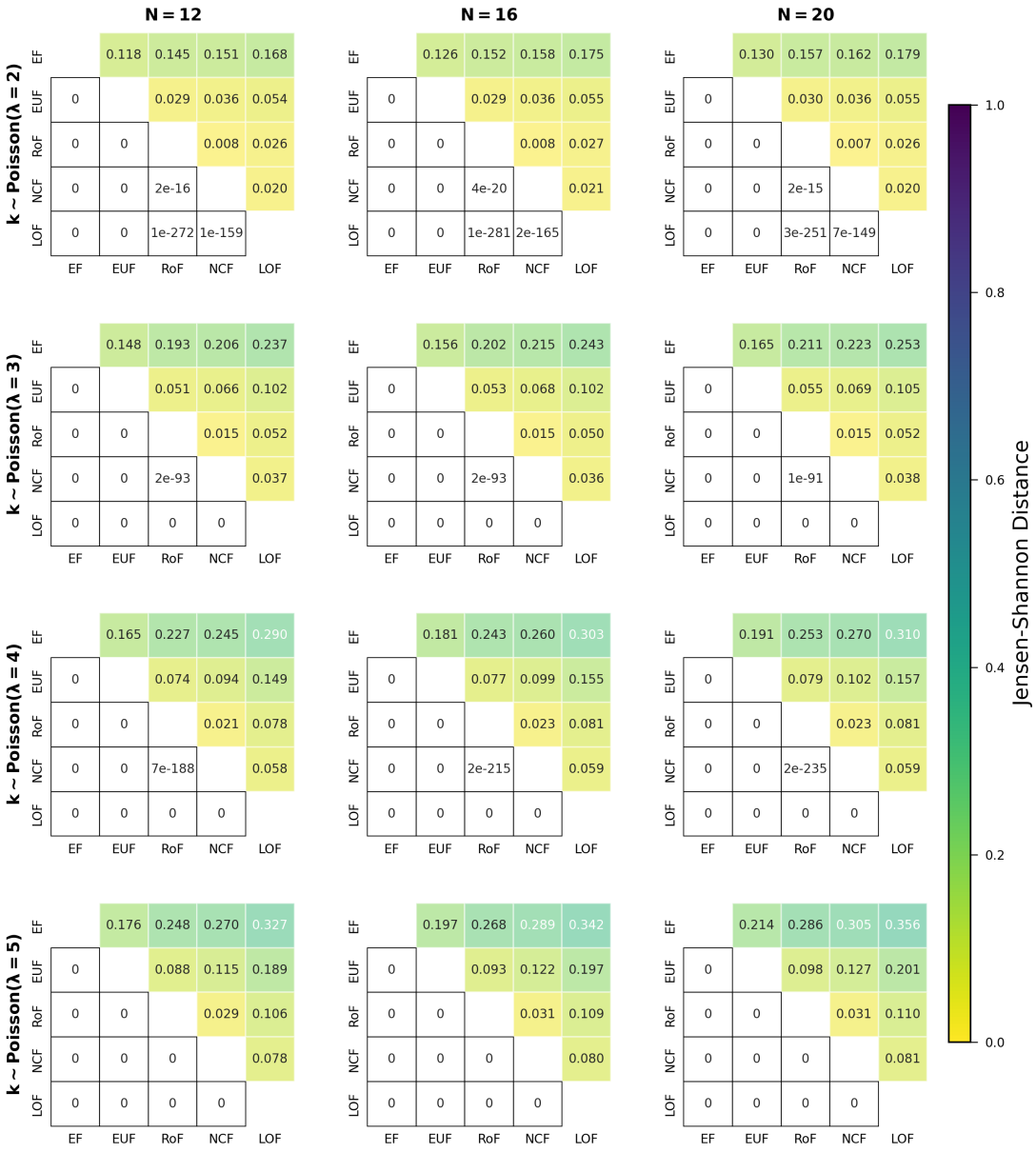


Figure A.41: Comparing distributions for the fraction of states in the basins of fixed points when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean = λ. At a given λ and N, we determined the distribution of the fraction of states in the basins of fixed points for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column greater than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

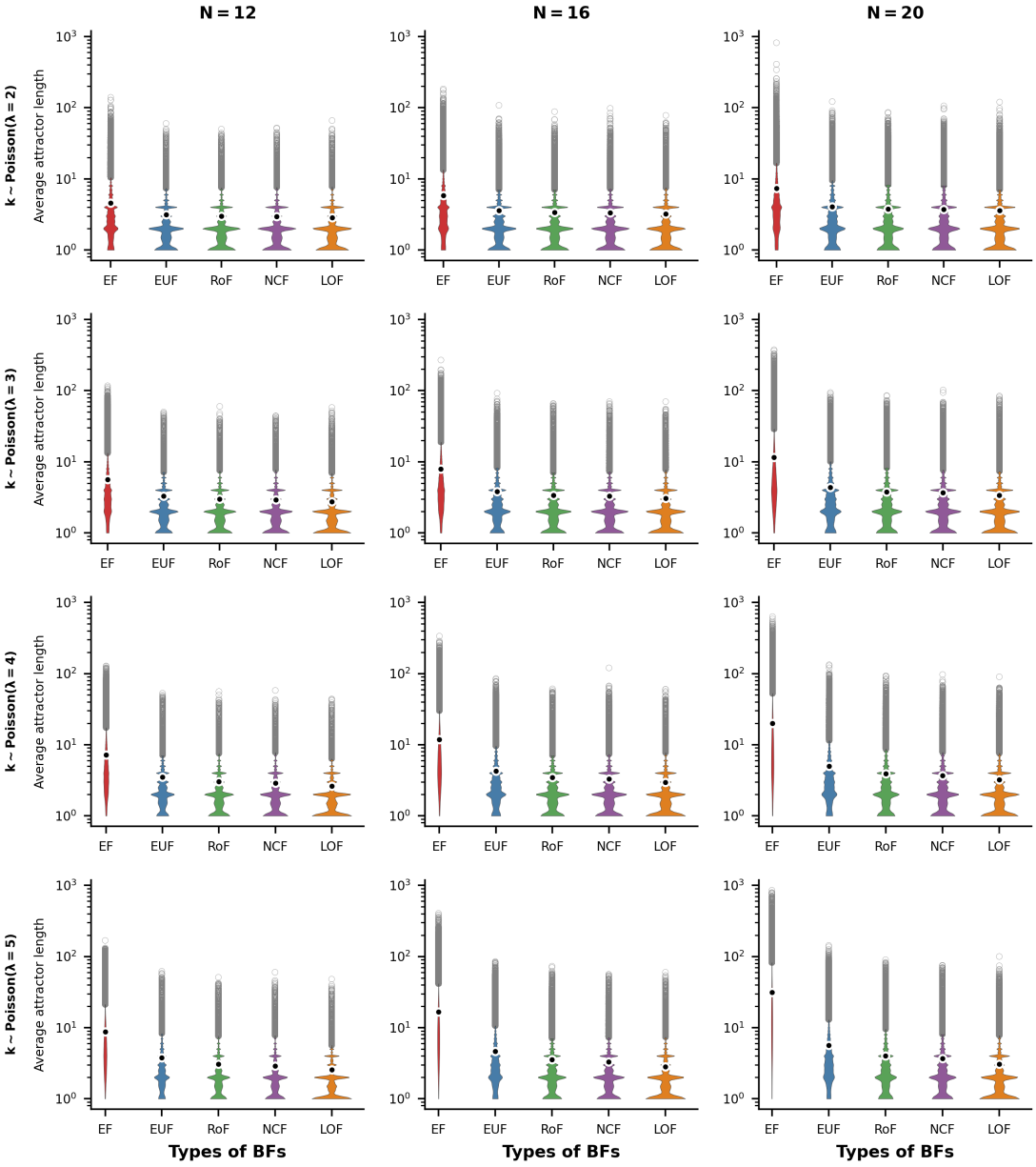


Figure A.42: Distributions of average attractor lengths for P-P networks. The violin plots display the distribution of the average attractor length for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

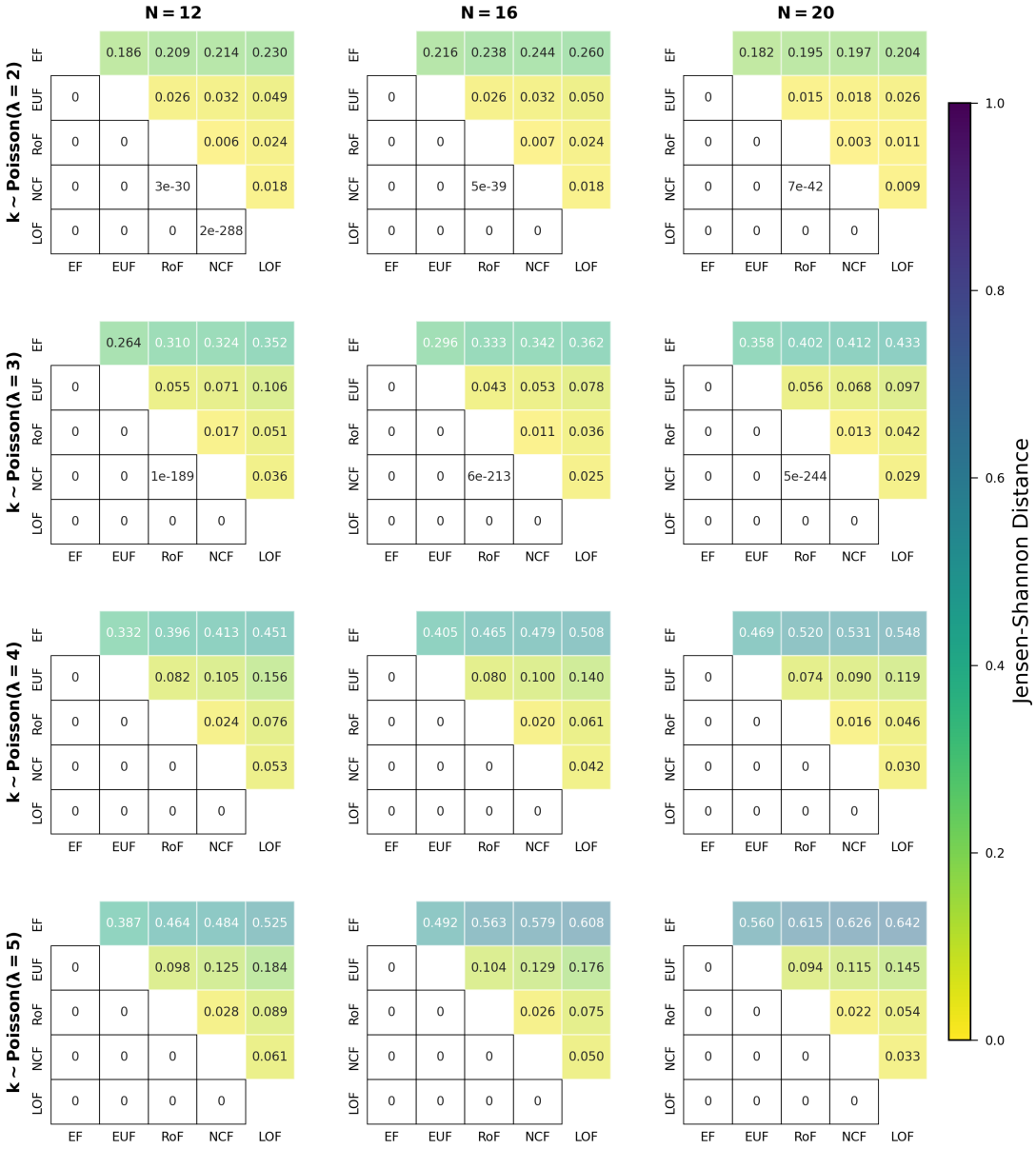


Figure A.43: Comparing distributions for the average attractor lengths when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean = λ. At a given λ and N, we determined the distribution of the average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

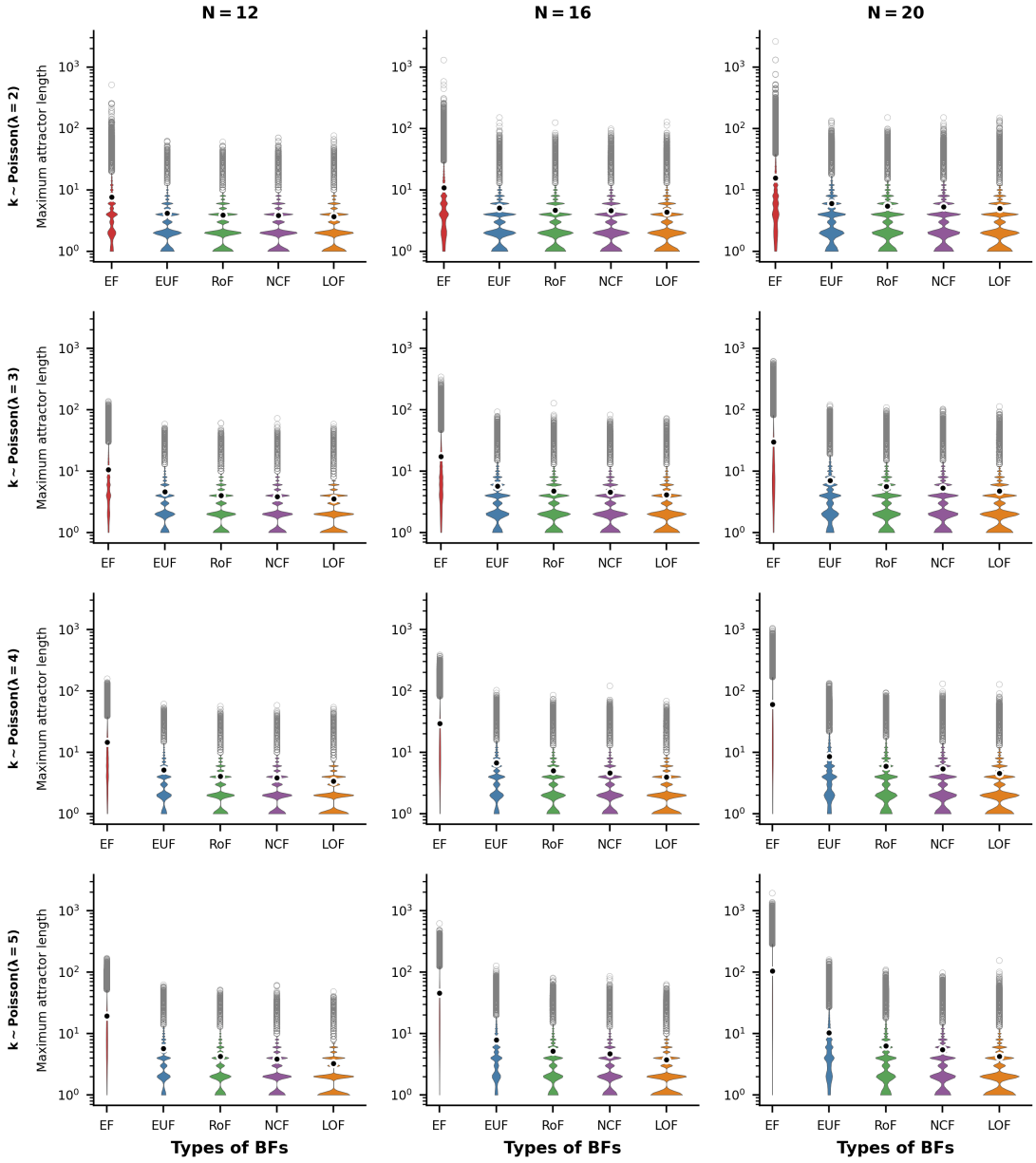


Figure A.44: Distributions of maximum attractor lengths for P-P networks. The violin plots display the distribution of the maximum attractor length for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the maximum attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

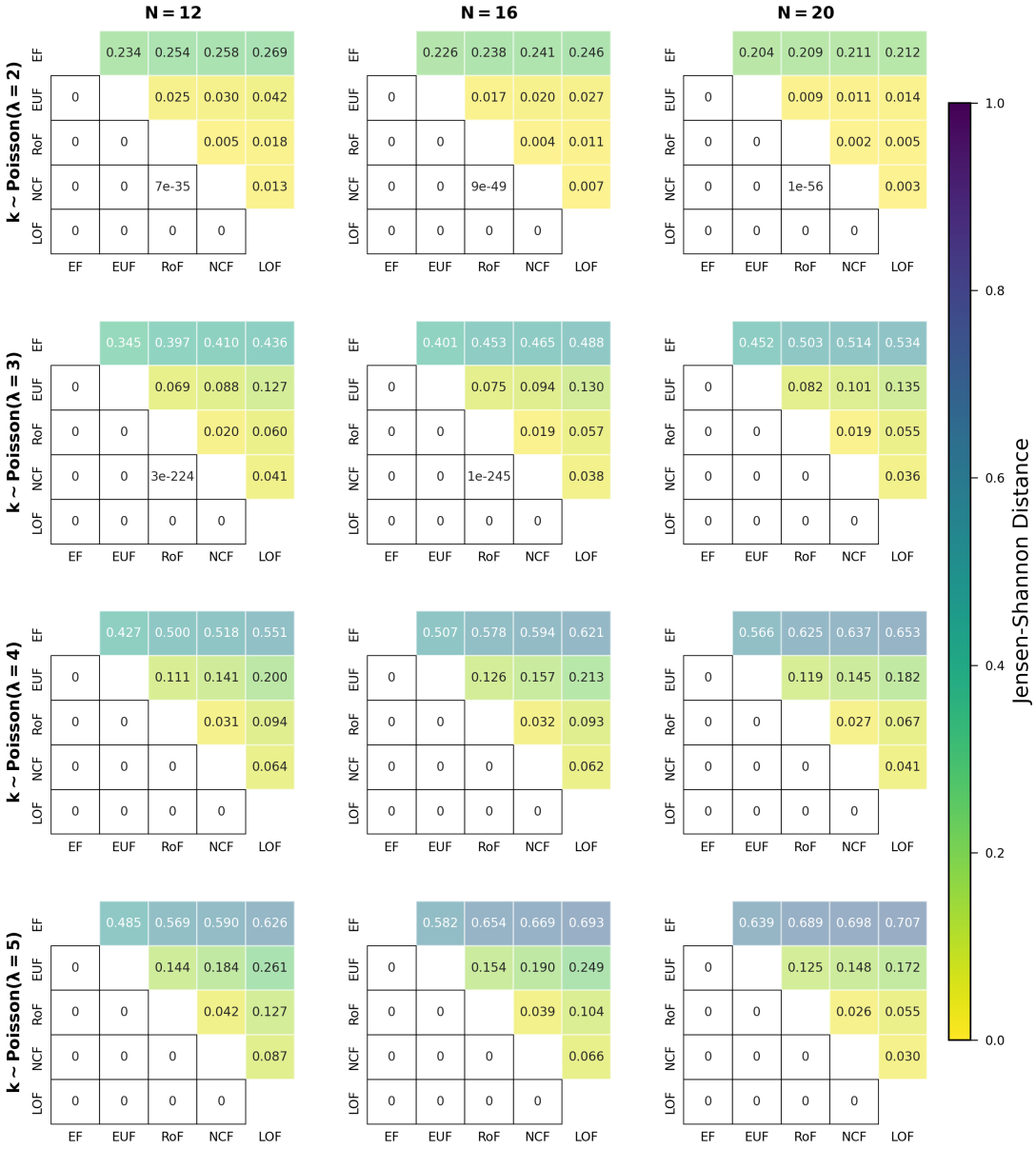


Figure A.45: Comparing distributions for the maximum attractor lengths when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean = λ. At a given λ and N, we determined the distribution of the maximum attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

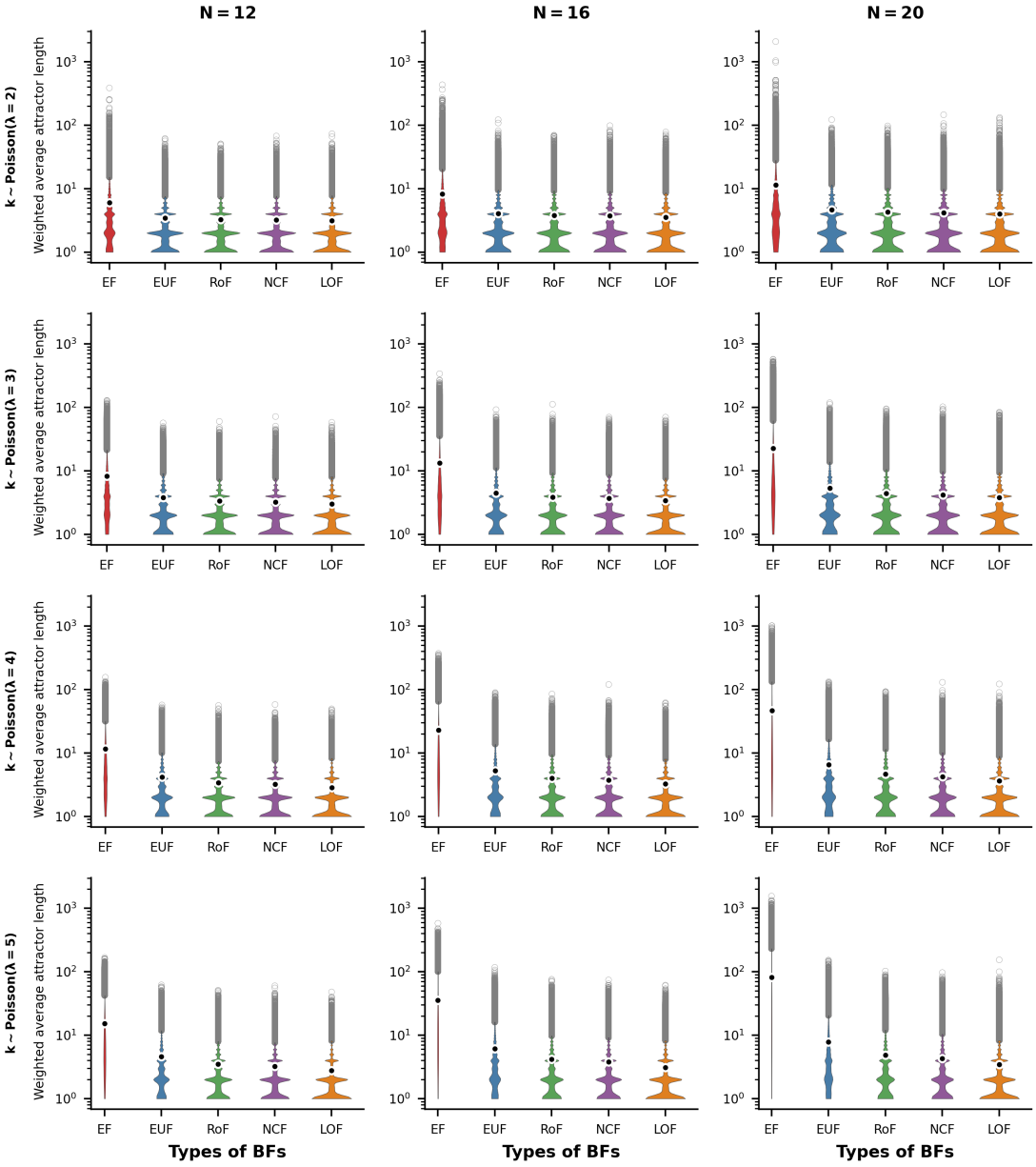


Figure A.46: Distributions of weighted average attractor lengths for P-P networks. The violin plots display the distribution of the weighted average attractor length for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the weighted average attractor length for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

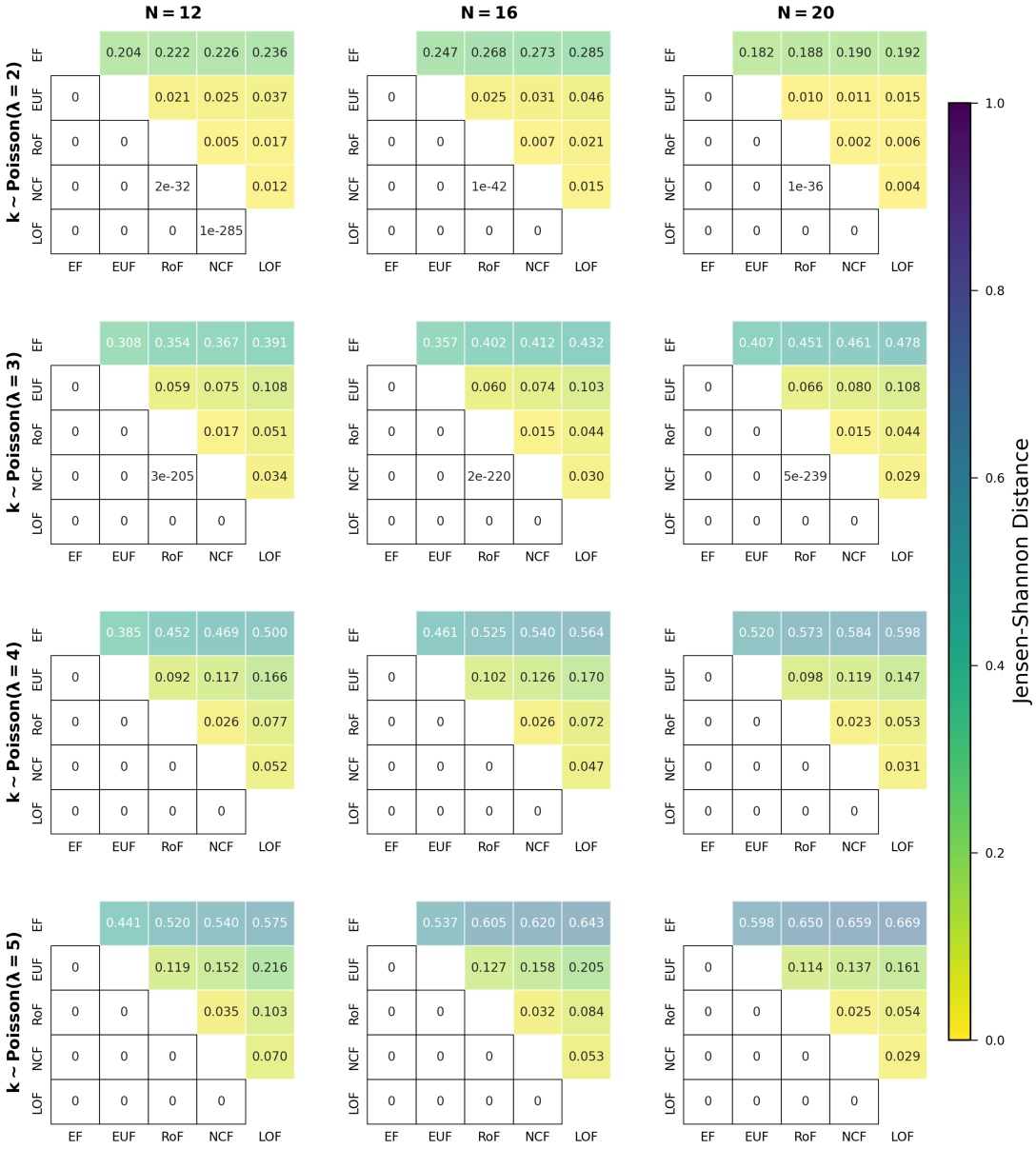


Figure A.47: Comparing distributions for the weighted average attractor lengths when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean $= \lambda$. At a given λ and N, we determined the distribution of the weighted average attractor lengths for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

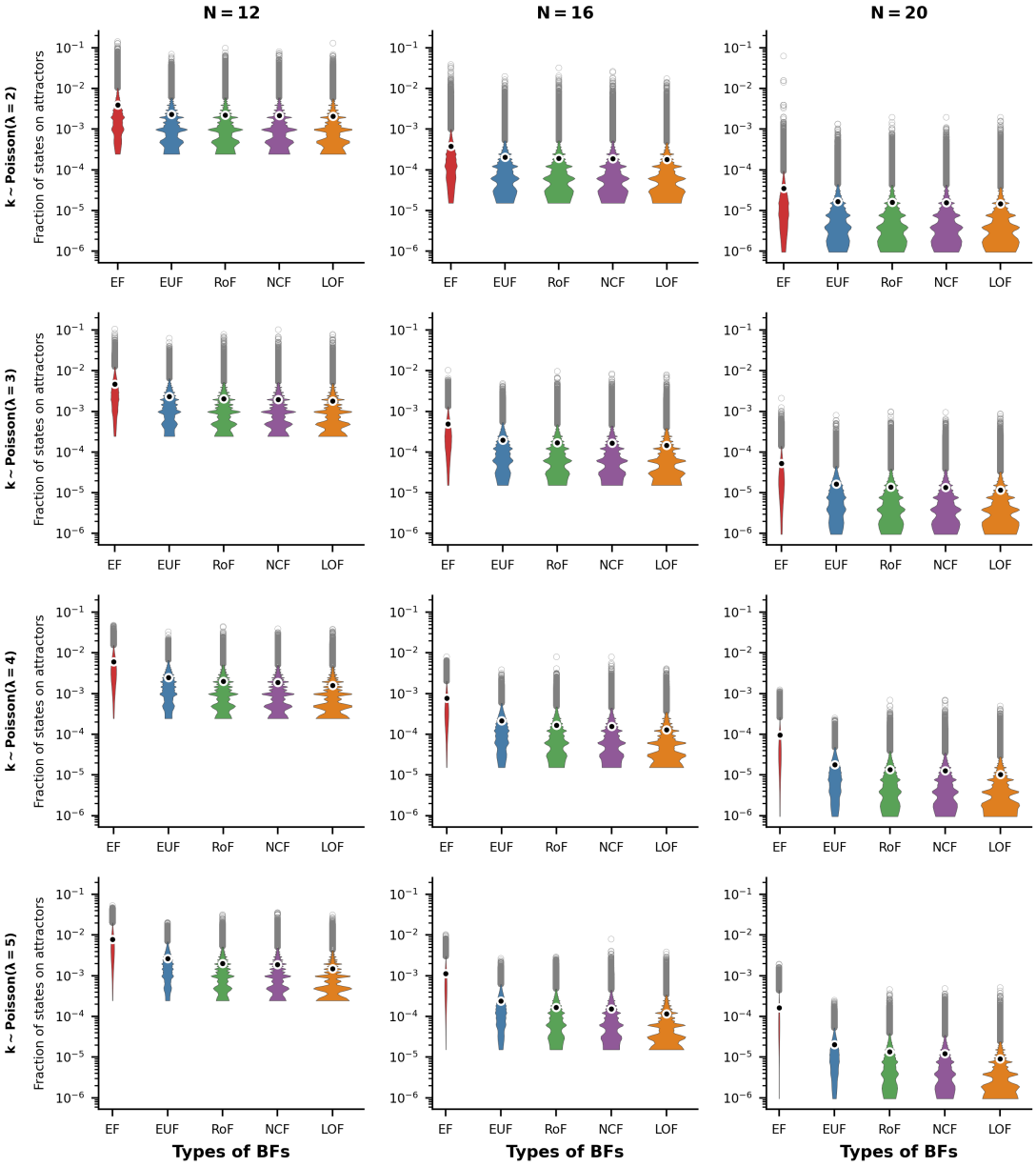


Figure A.48: Distributions of fraction of states on attractors for P-P networks. The violin plots display the distribution of the fraction of states on attractors for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes $N = 12, 16$ and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means $2, 3, 4$ and 5. Within each sub-figure, there are five violins representing the distributions of the fraction of states on attractors for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF. Each violin distribution is based on 10^6 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings. Note the logarithmic scale for the y-axis.

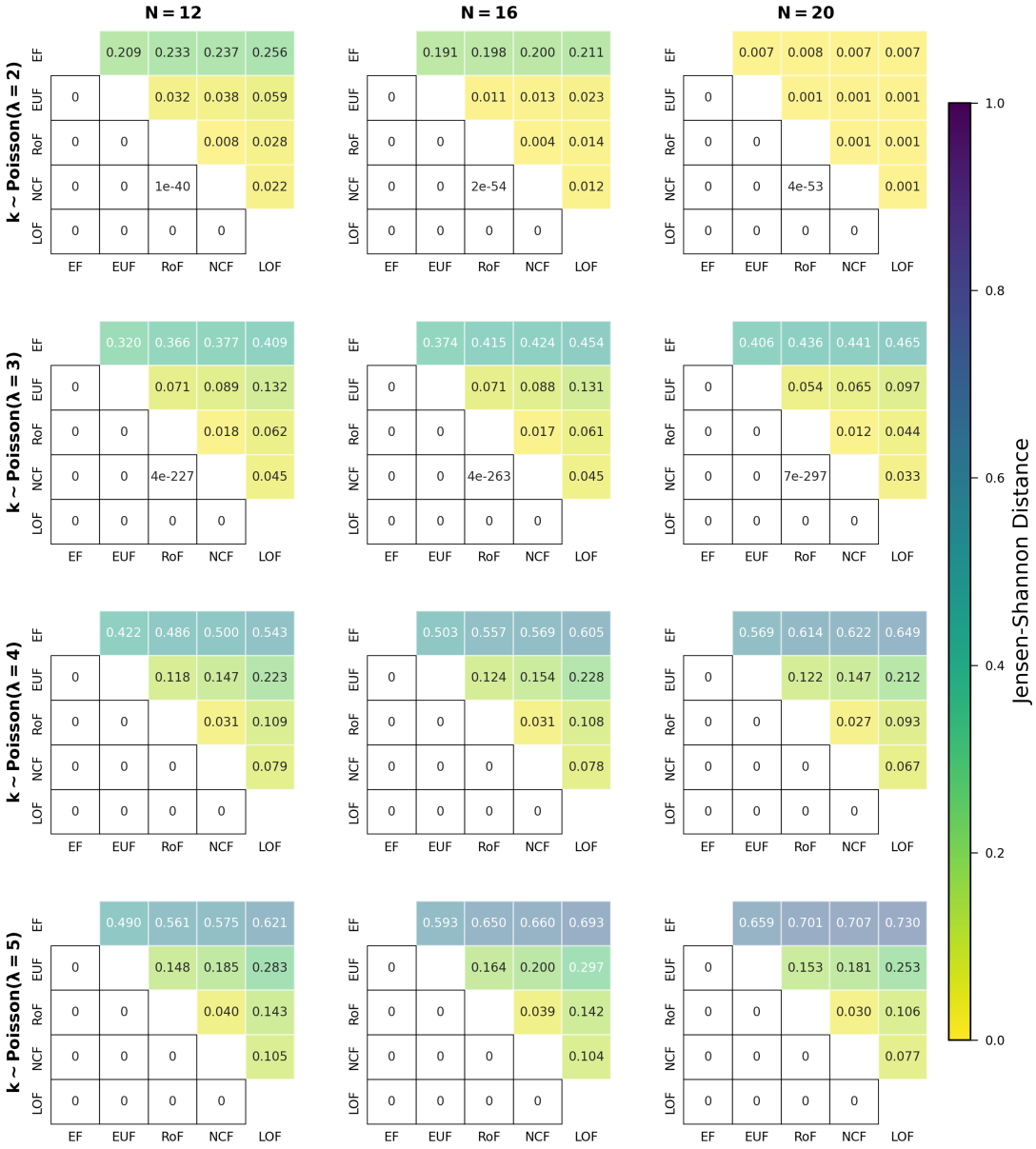


Figure A.49: Comparing distributions for the fraction of states on attractors when using different classes of BFs within the P-P networks. The different sub-figures are organised according to the values of N (the number of nodes in the network) and λ ($\lambda = 2, 3, 4$ and 5). The degrees of each node of the networks are drawn randomly from a Poisson distributions with mean $= \lambda$. At a given λ and N, we determined the distribution of the fraction of states on attractors for each of the 5 classes of BFs. In each sub-figure, the upper triangular part displays the pairwise Jensen-Shannon distances of those distributions. The lower triangular part represents the p-values of the one-sided K-S test (cumulative distribution of the label of column less than that of the row). Entries shown as 0 are actually very small, with values less than 10^{-307}.

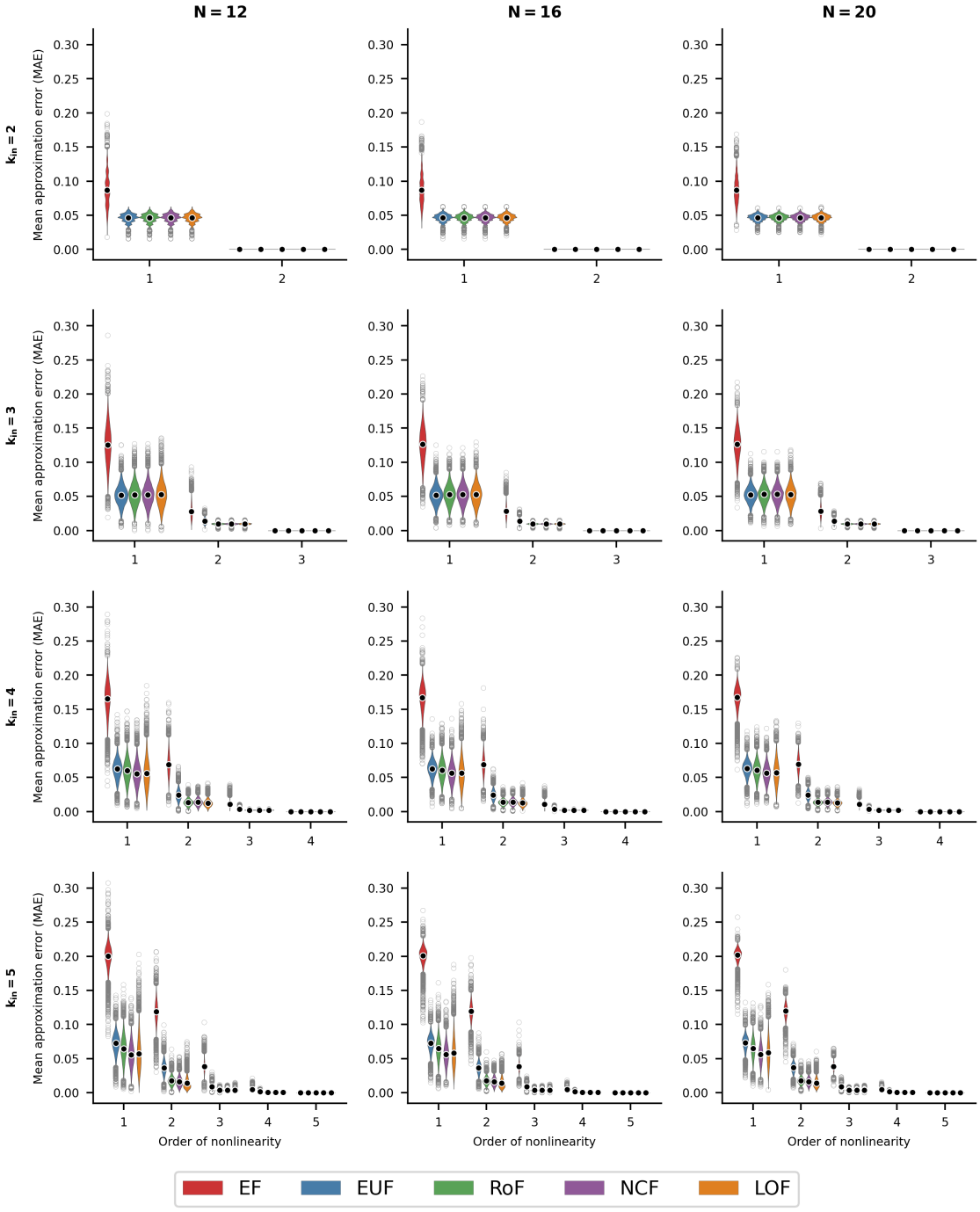


Figure A.50: Distributions of mean approximation error (MAE) for R-P networks. The violin plots display the distribution of the MAE for the R-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes N=12, 16 and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are grouped violins that show the distributions for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF for different orders of nonlinearity. Each violin distribution is based on 10^4 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

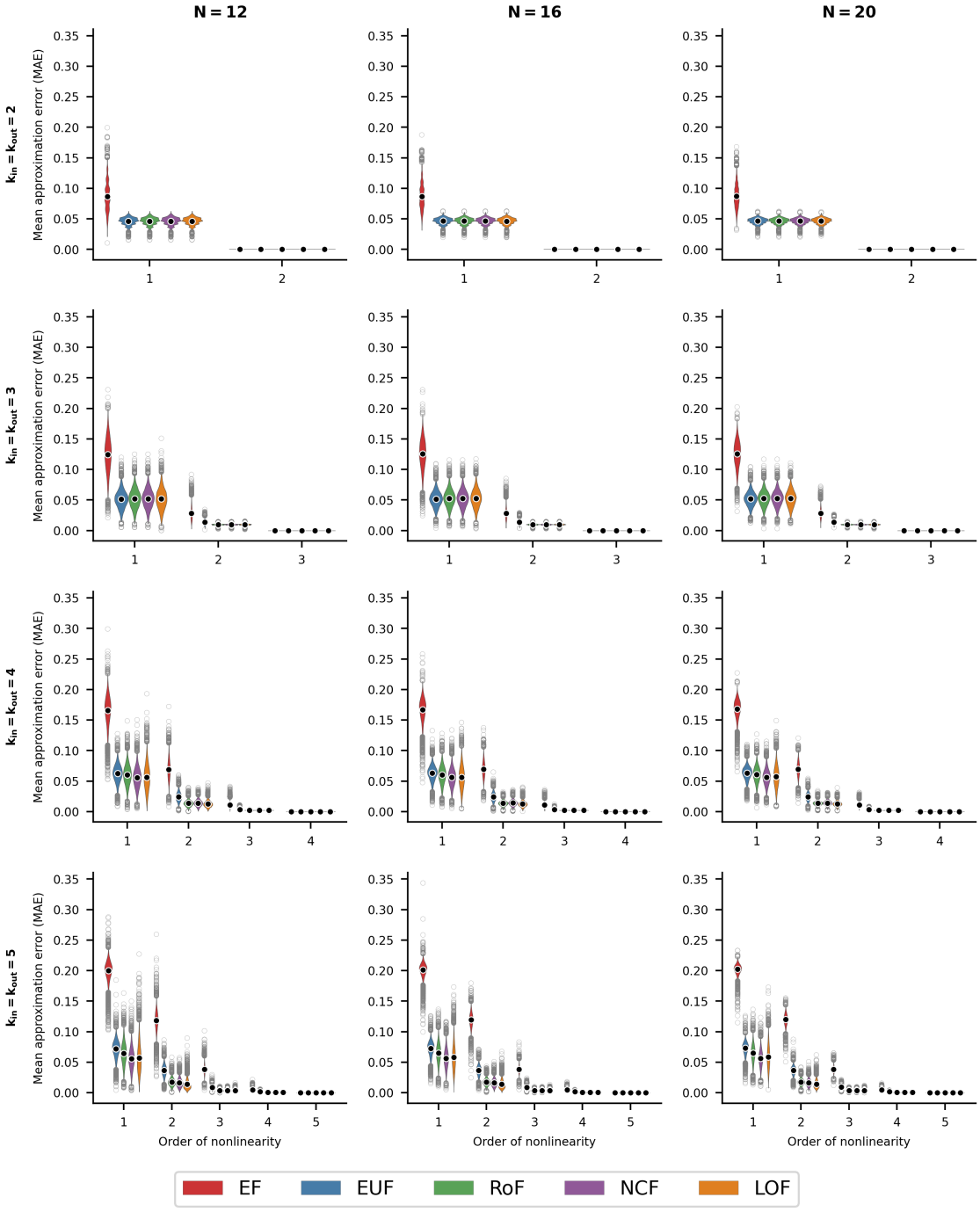


Figure A.51: Distributions of mean approximation error (MAE) for R-R networks. The violin plots display the distribution of the MAE for the R-R network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes N=12, 16 and 20 respectively. Each row represents a different number of inputs per node, ranging from 2 to 5. Within each sub-figure, there are grouped violins that show the distributions for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF for different orders of nonlinearity. Each violin distribution is based on 10^4 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

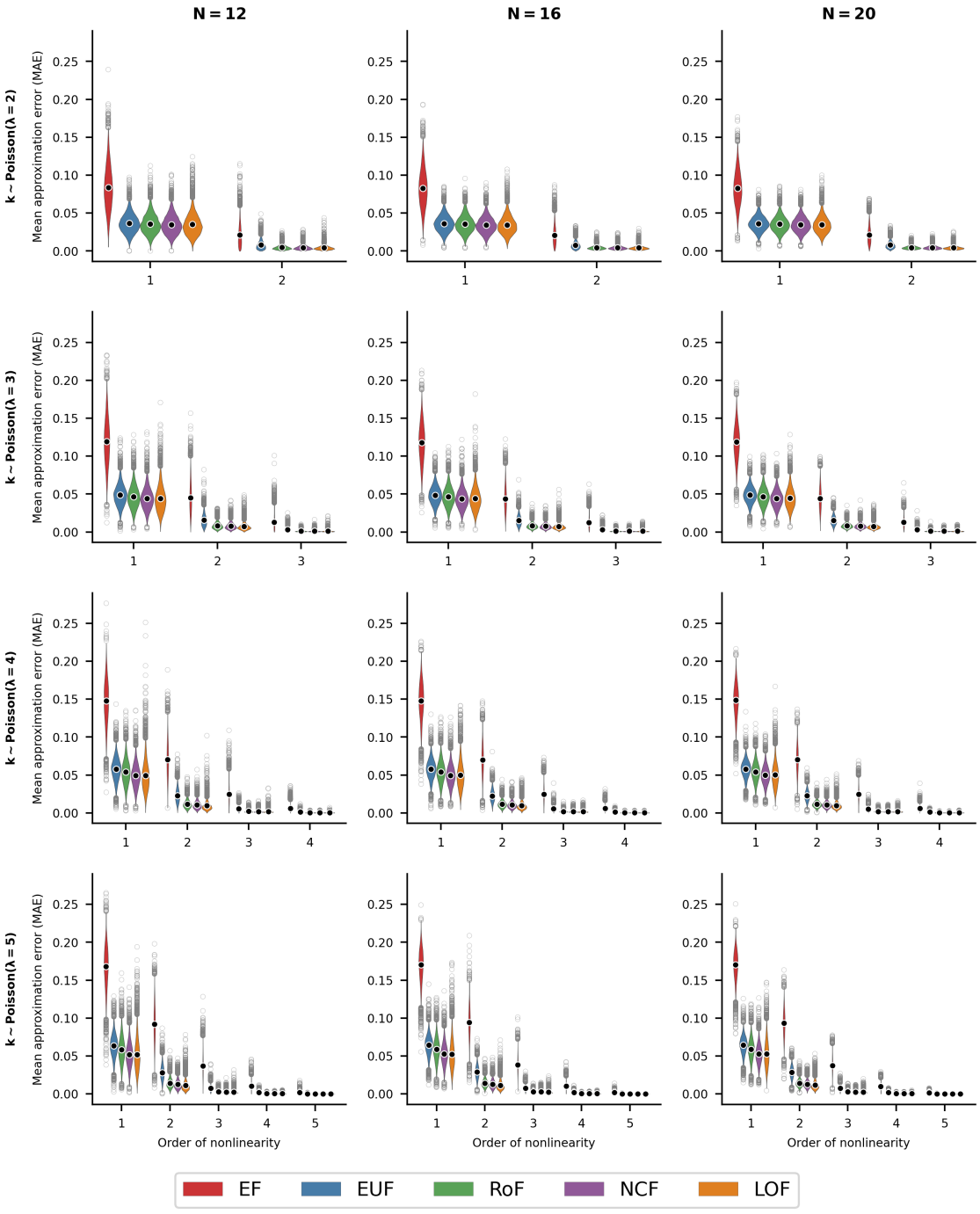


Figure A.52: Distributions of mean approximation error (MAE) for P-P networks. The violin plots display the distribution of the MAE for the P-P network topology. The sub-figures are organized into 3 columns, corresponding to the network sizes N=12, 16 and 20 respectively. Each row represents a different average number of inputs per node, with the degrees drawn from Poisson distributions with means 2, 3, 4 and 5. Within each sub-figure, there are grouped violins that show the distributions for the different ensembles of models, namely EF, EUF, RoF, NCF and LOF for different orders of nonlinearity. Each violin distribution is based on 10^4 data points. Outliers of each distribution are shown as gray rings while the means are shown as black dots encircled with white rings.

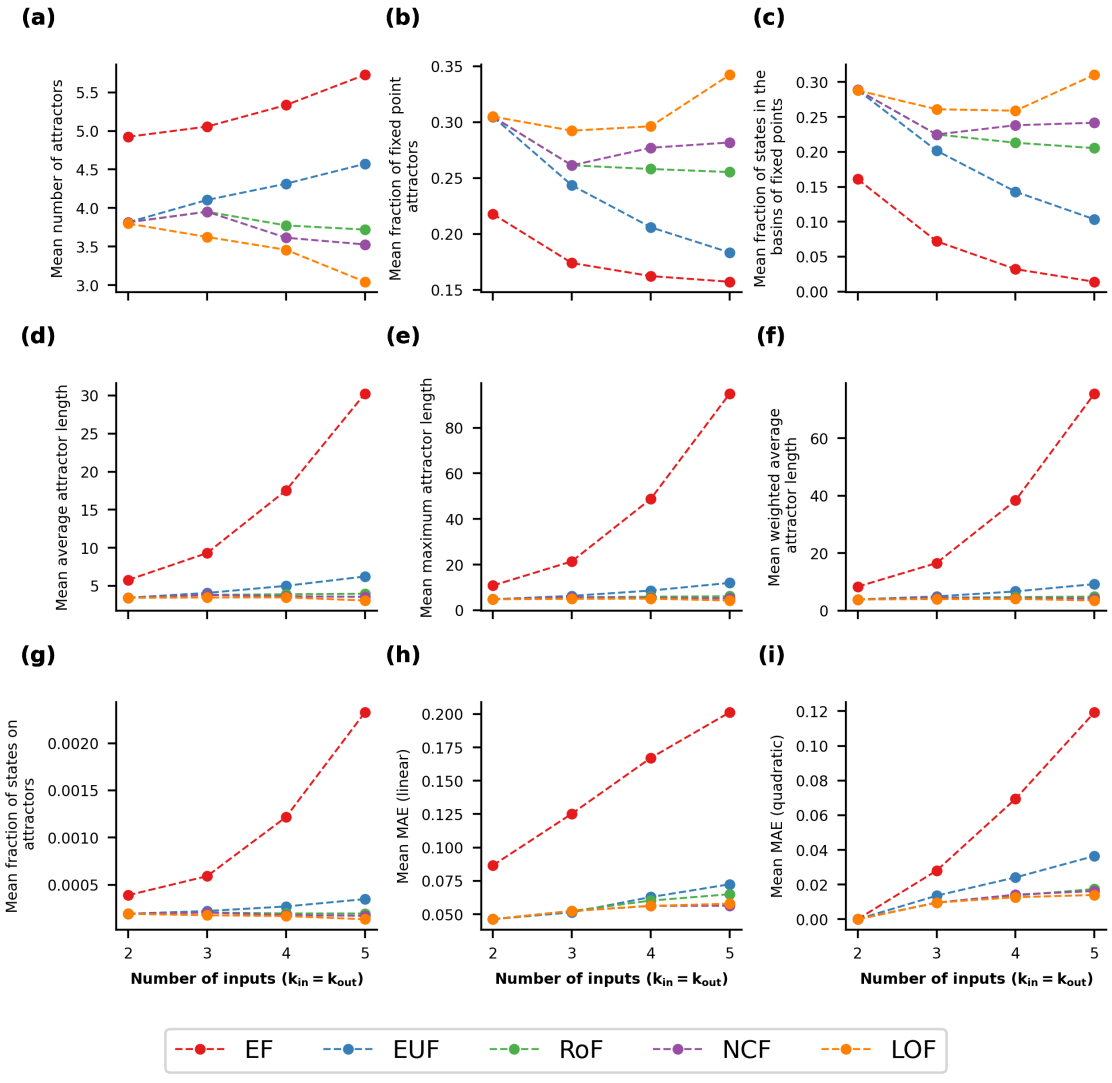


Figure A.53: Mean values of 9 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 16$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for R-R network topology.

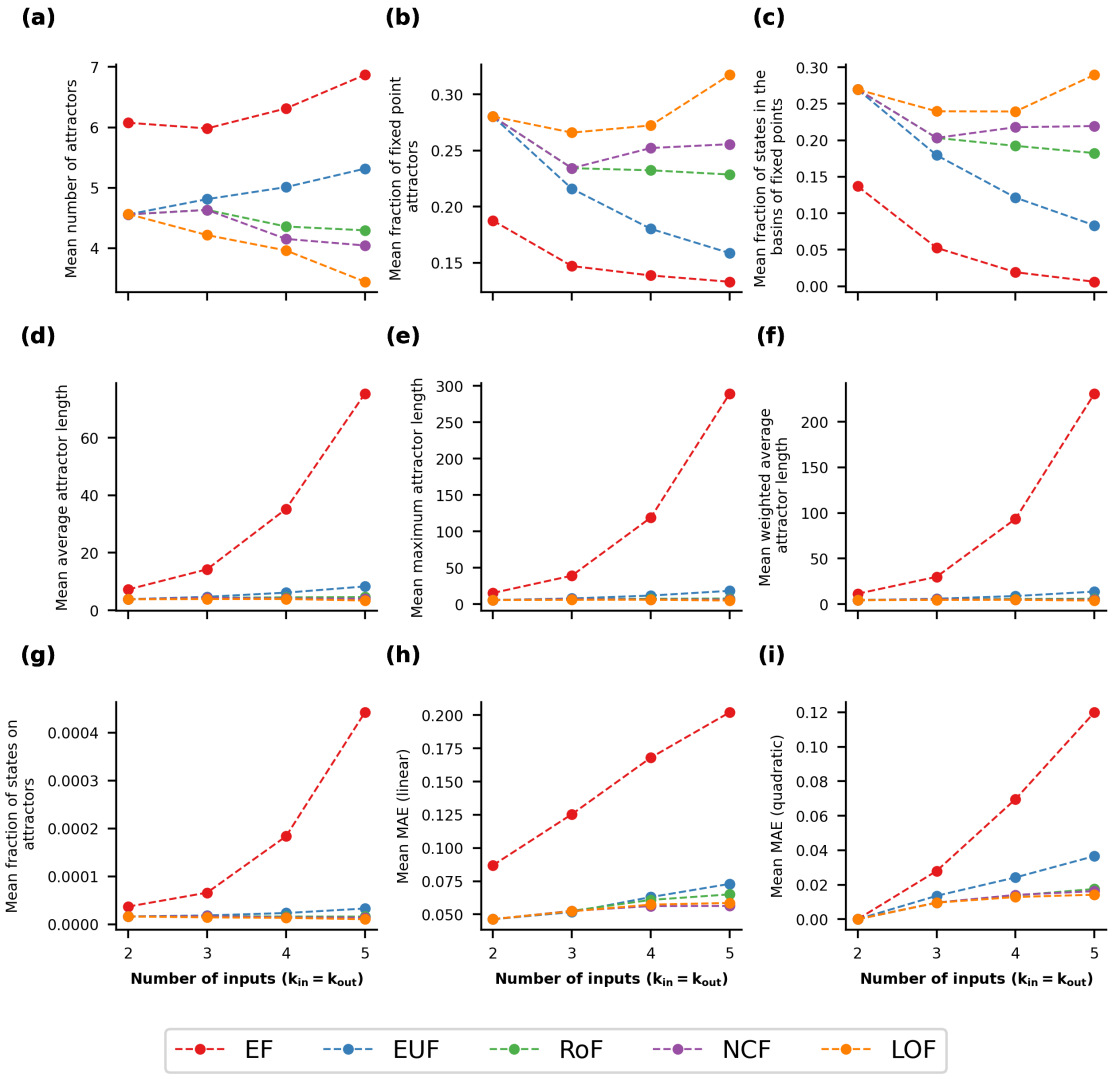


Figure A.54: Mean values of 9 stability observables as a function of $k_{in}(= k_{out})$ for different classes of BFs, in the case of R-R networks at $N = 20$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for R-R network topology.

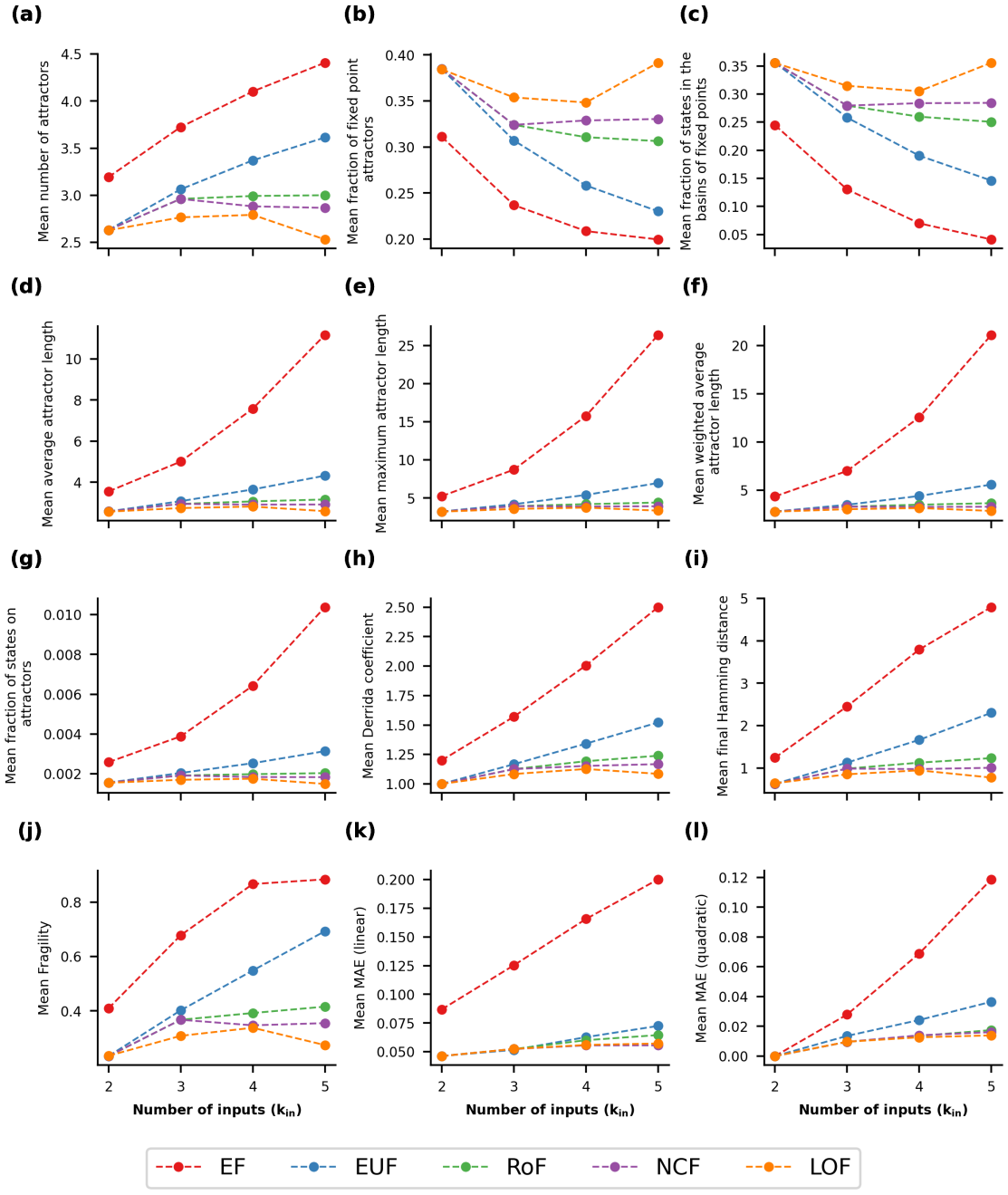


Figure A.55: Mean values of our 12 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 12$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for R-P network topology.

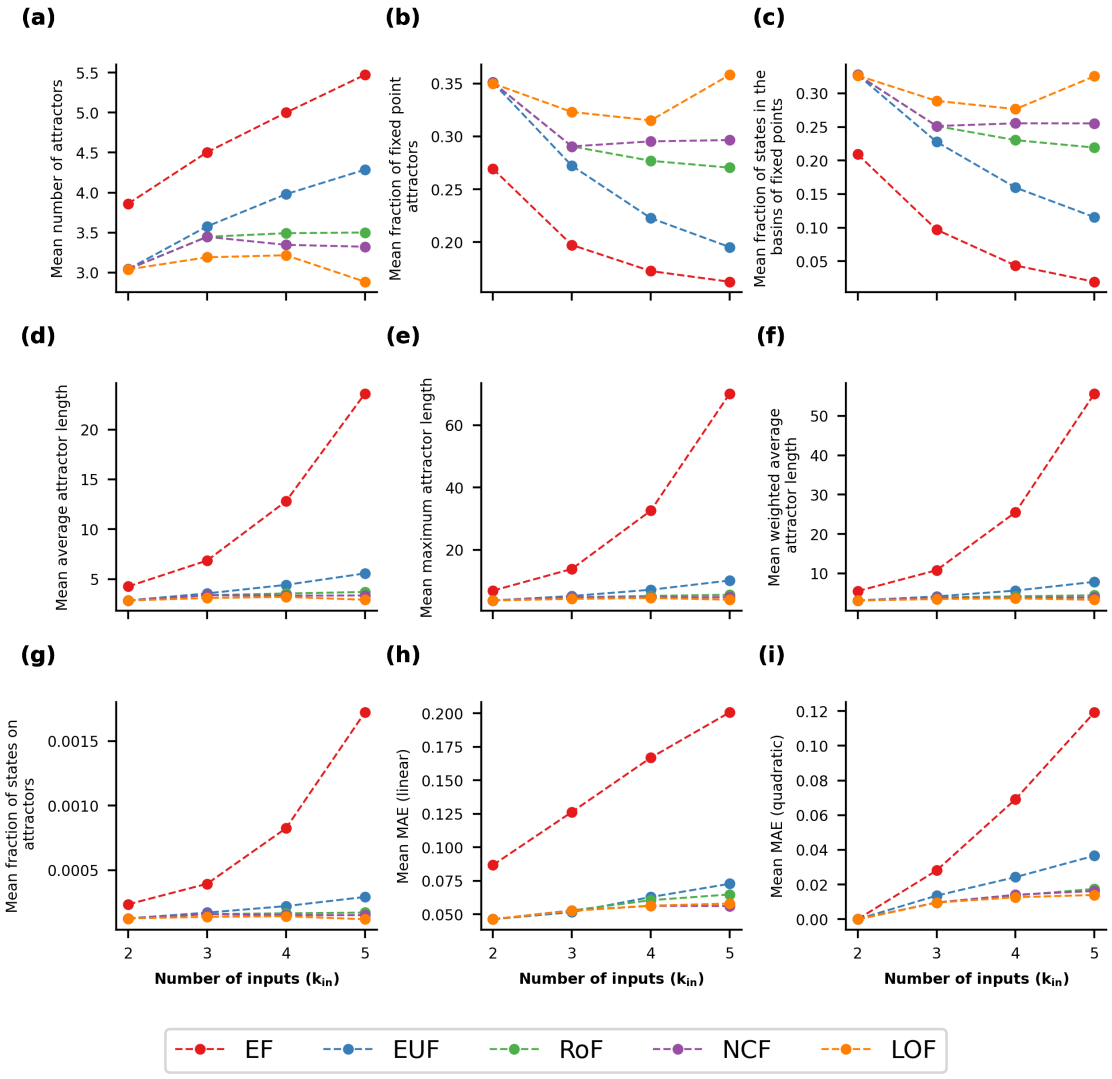


Figure A.56: Mean values of 9 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 16$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for R-P network topology.

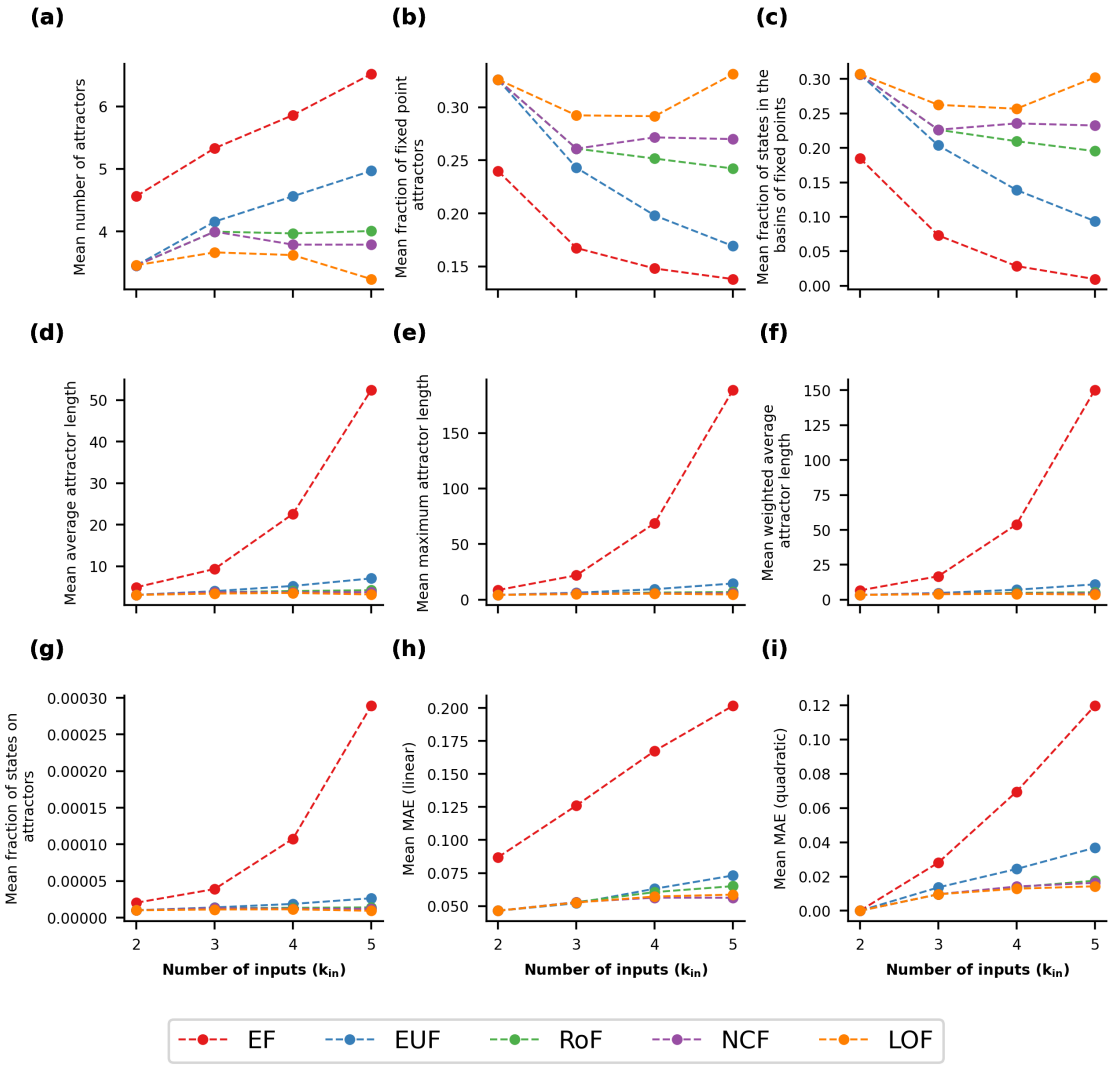


Figure A.57: Mean values of 9 stability observables as a function of k_{in} for different classes of BFs, in the case of R-P networks at $N = 20$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of k_{in} for R-P network topology.

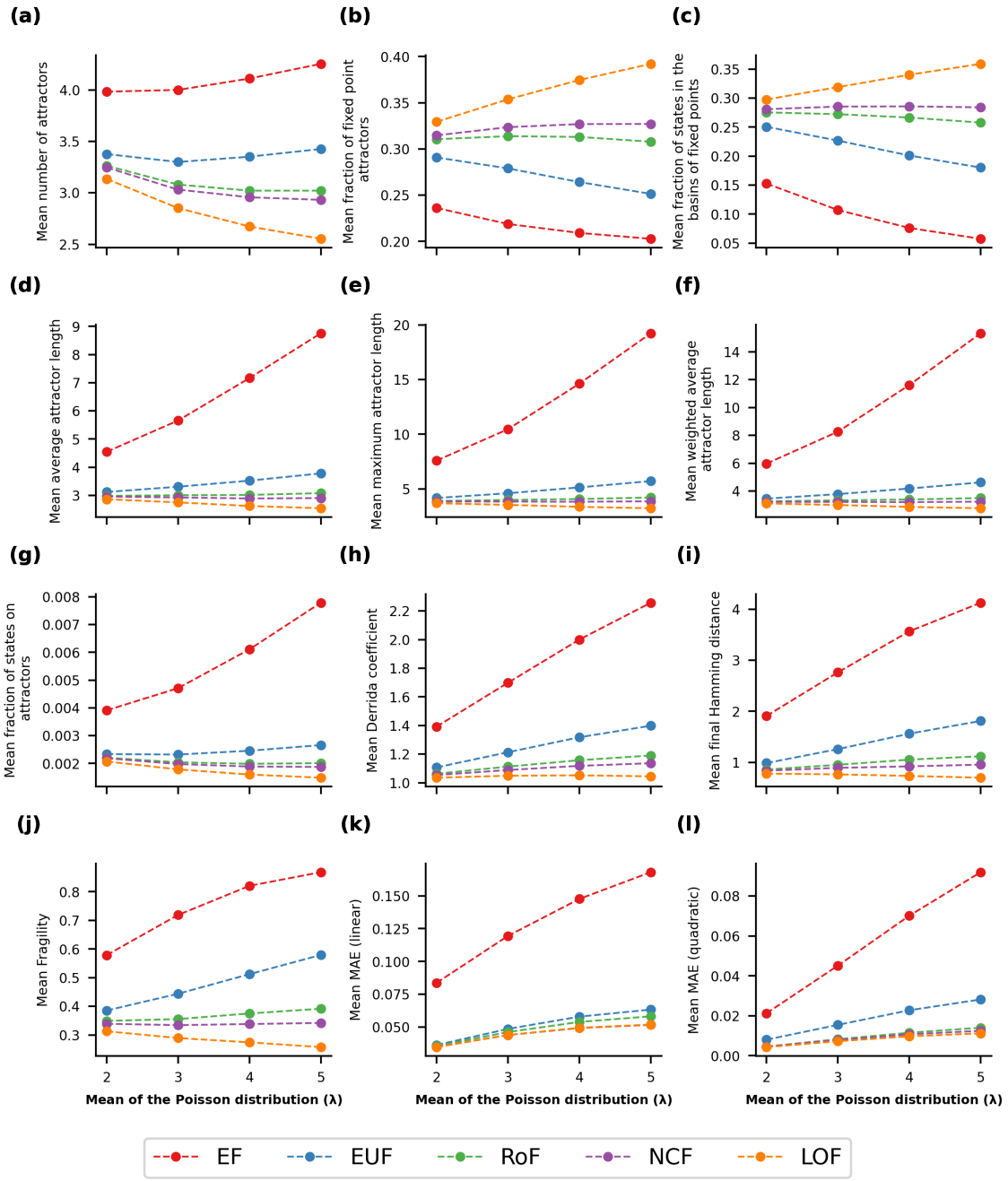


Figure A.58: Mean values of our 12 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 12$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of λ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean = λ) for P-P network topology.

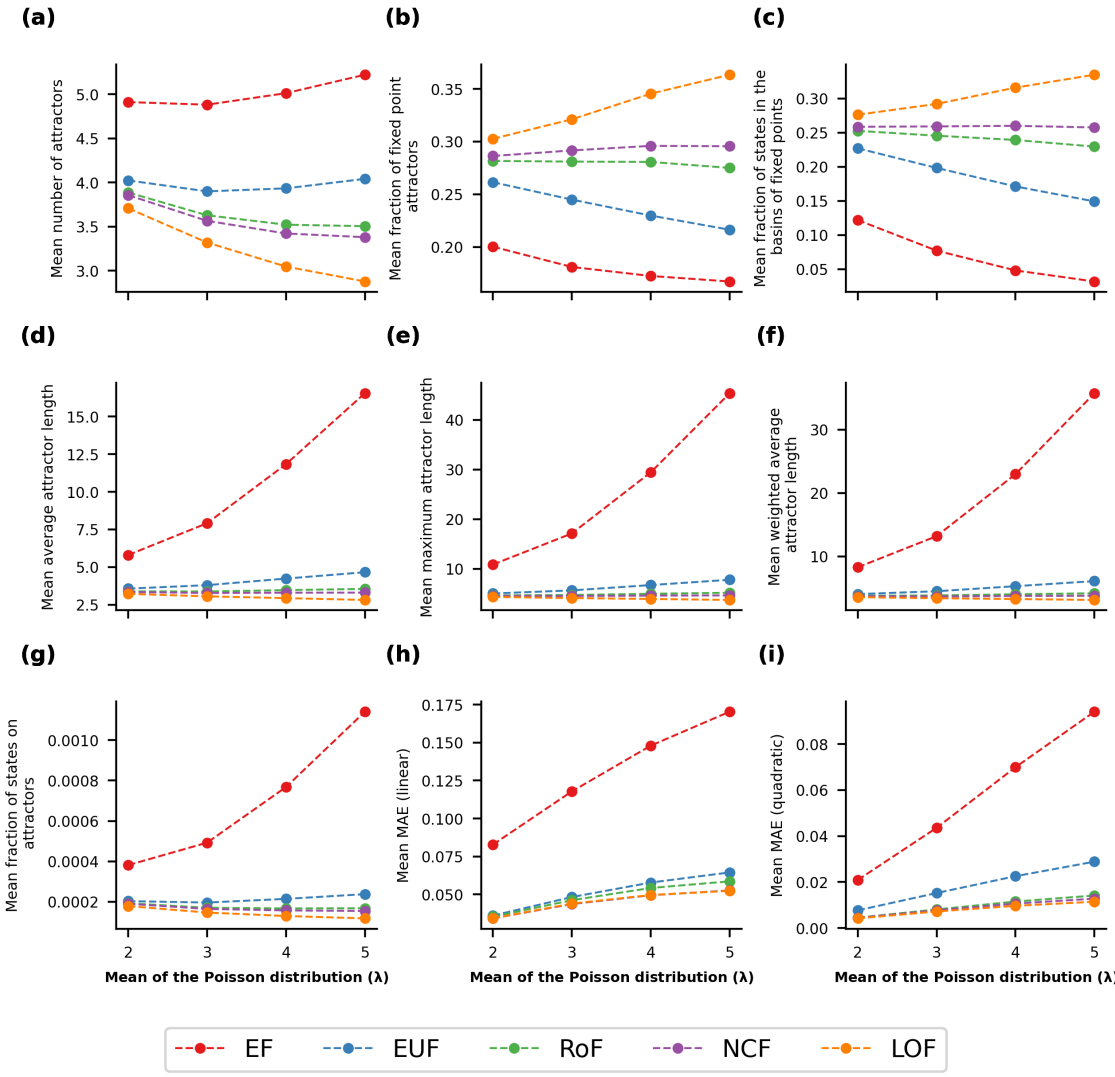


Figure A.59: Mean values of 9 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 16$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of λ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean = λ) for P-P network topology.

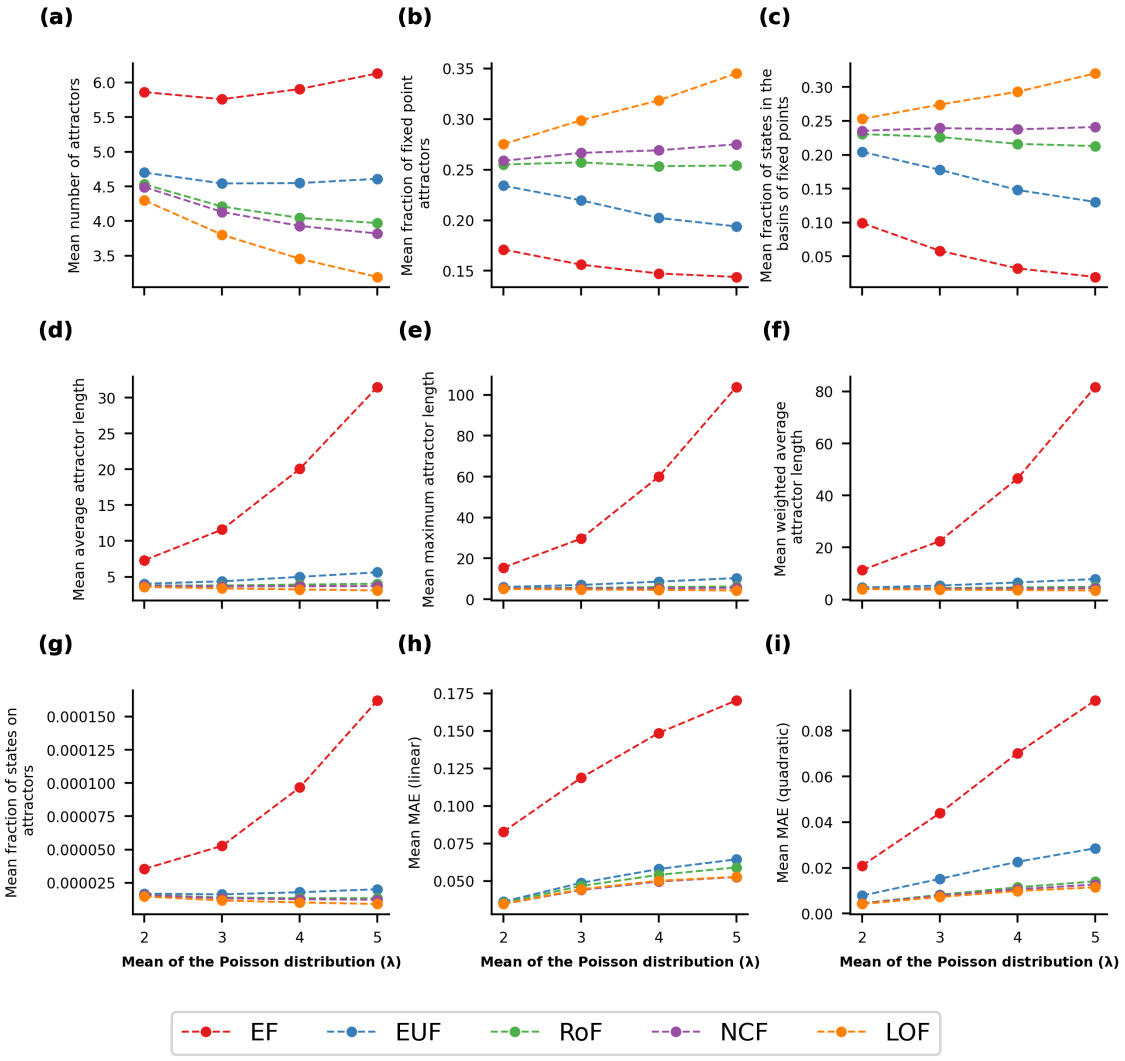


Figure A.60: Mean values of 9 stability observables as a function of λ for different classes of BFs, in the case of P-P networks at $N = 20$. Within each sub-figure, the 5 curves correspond to the five different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of λ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean = λ) for P-P network topology.

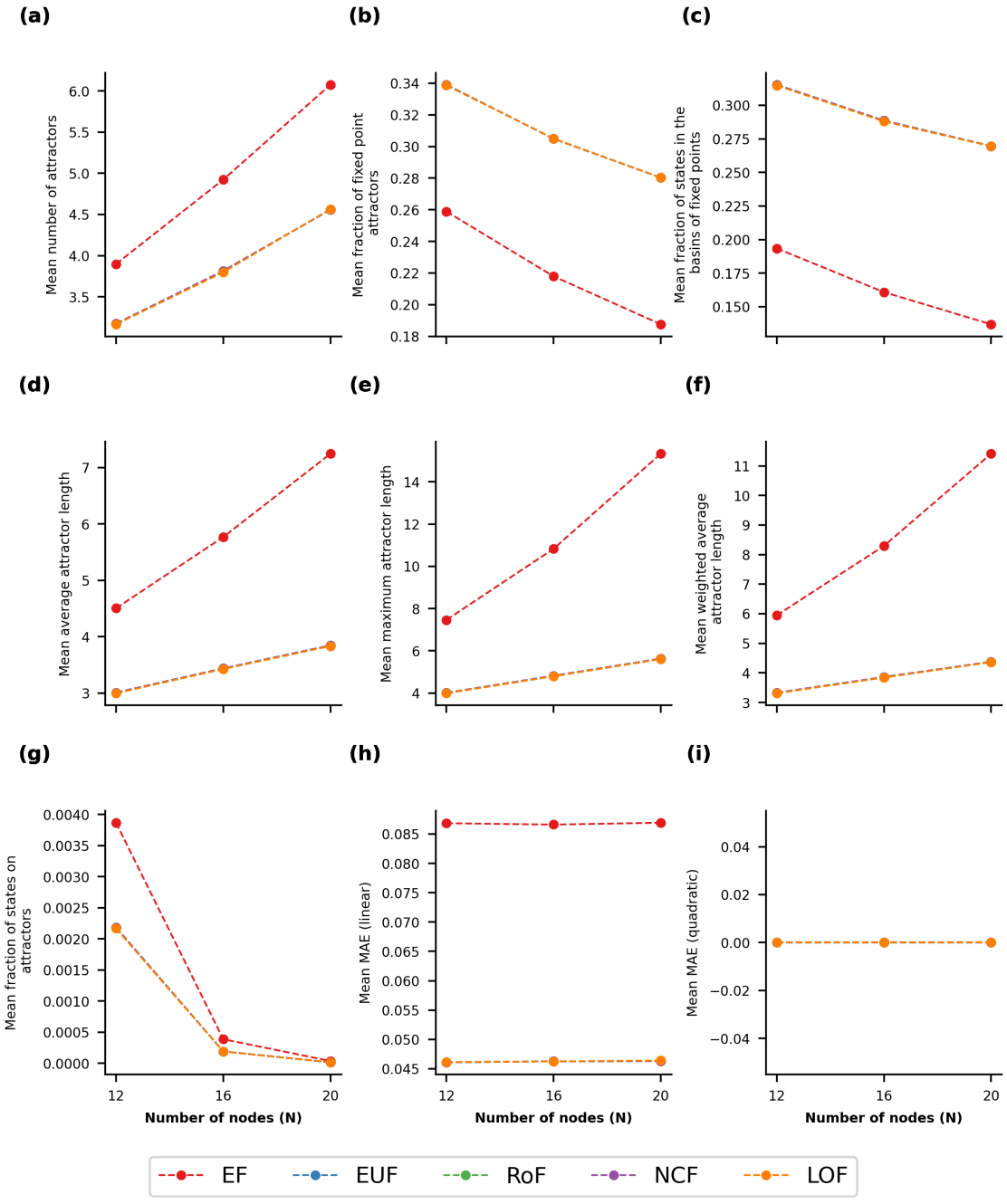


Figure A.61: Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 2$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-R network topology with $k_{in} = k_{out} = 2$.

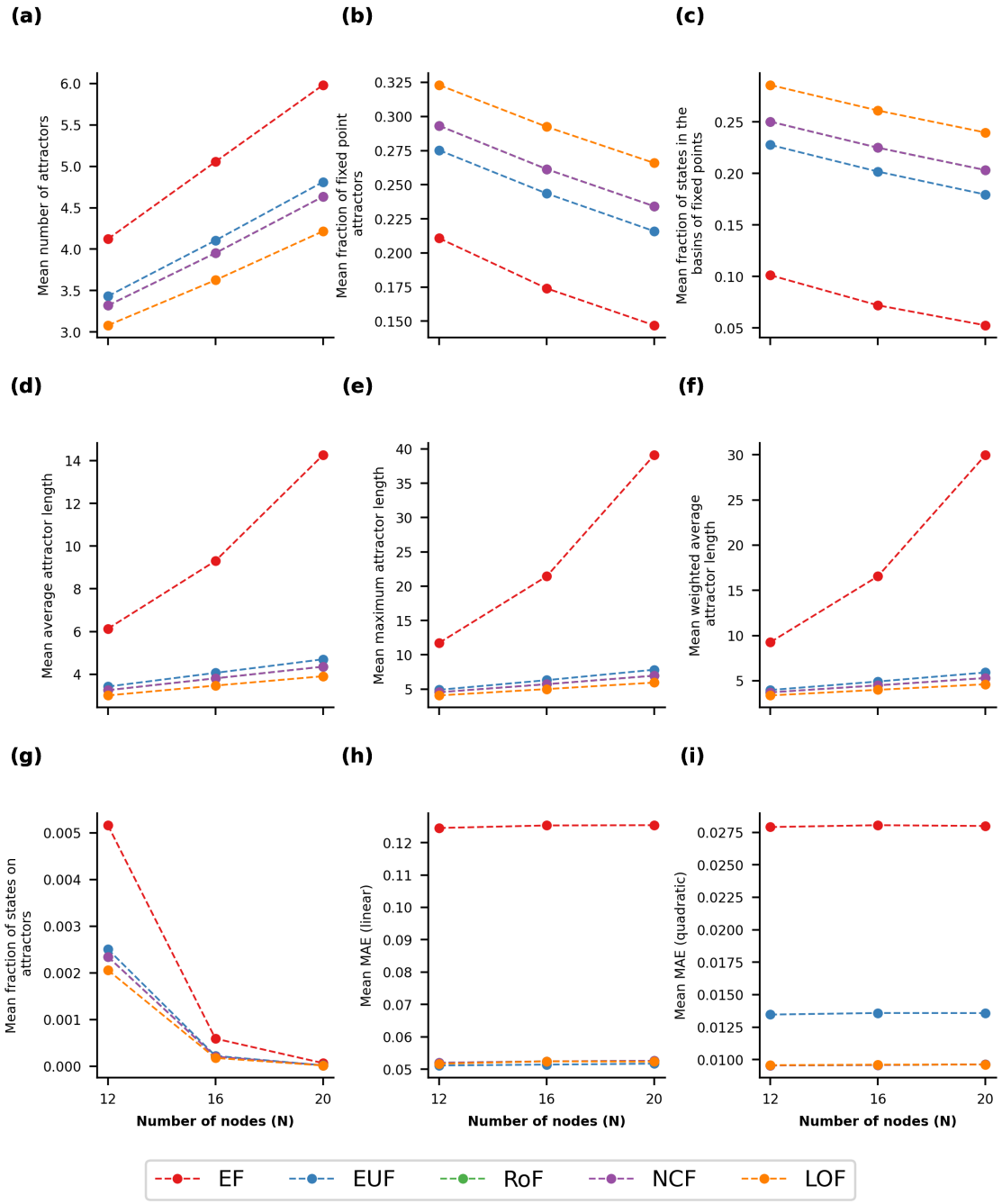


Figure A.62: Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 3$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-R network topology with $k_{in} = k_{out} = 3$.

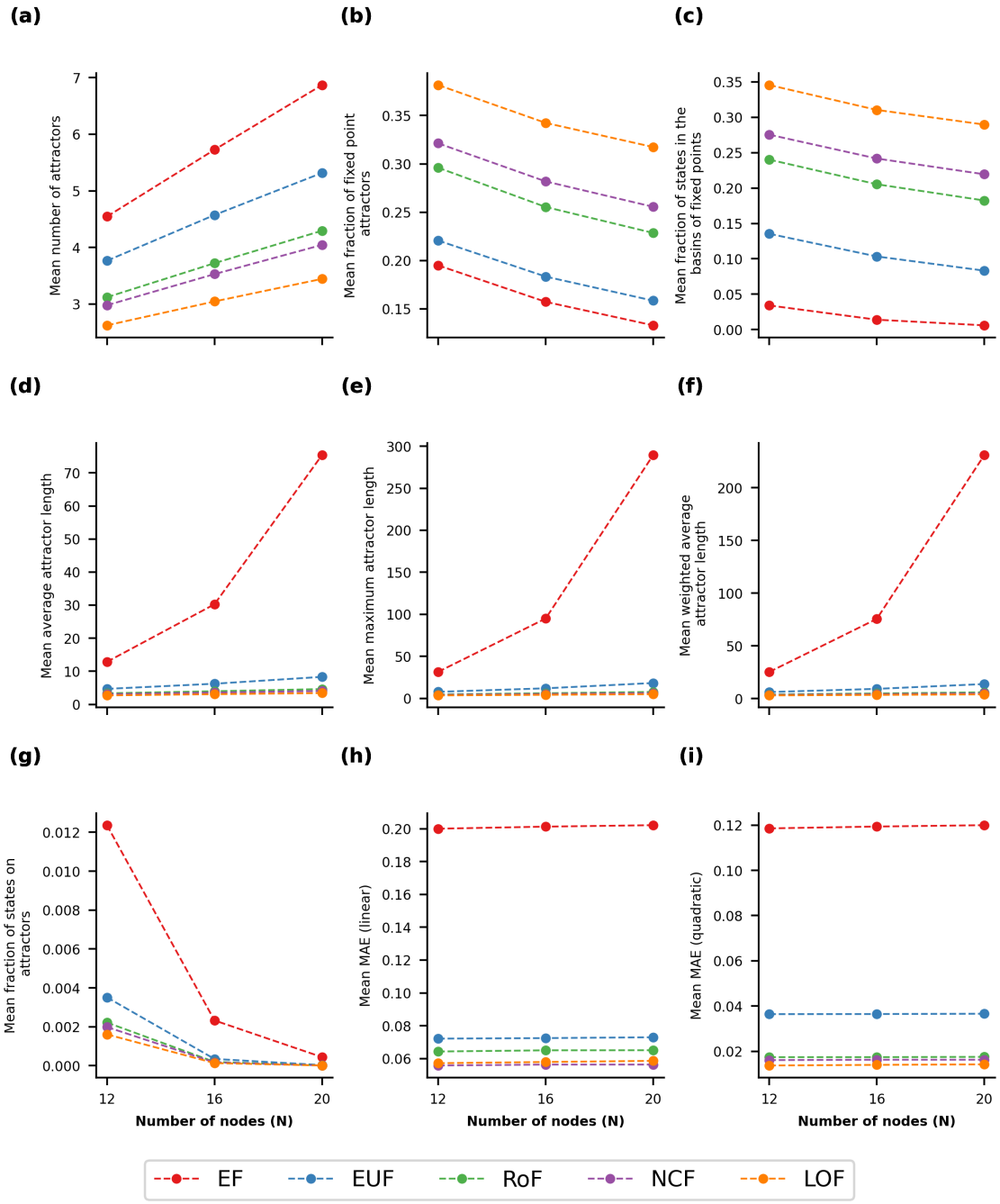


Figure A.63: Mean of the stability observables for different values of N in the case of the R-R topology with $k_{in} = k_{out} = 5$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-R network topology with $k_{in} = k_{out} = 5$.

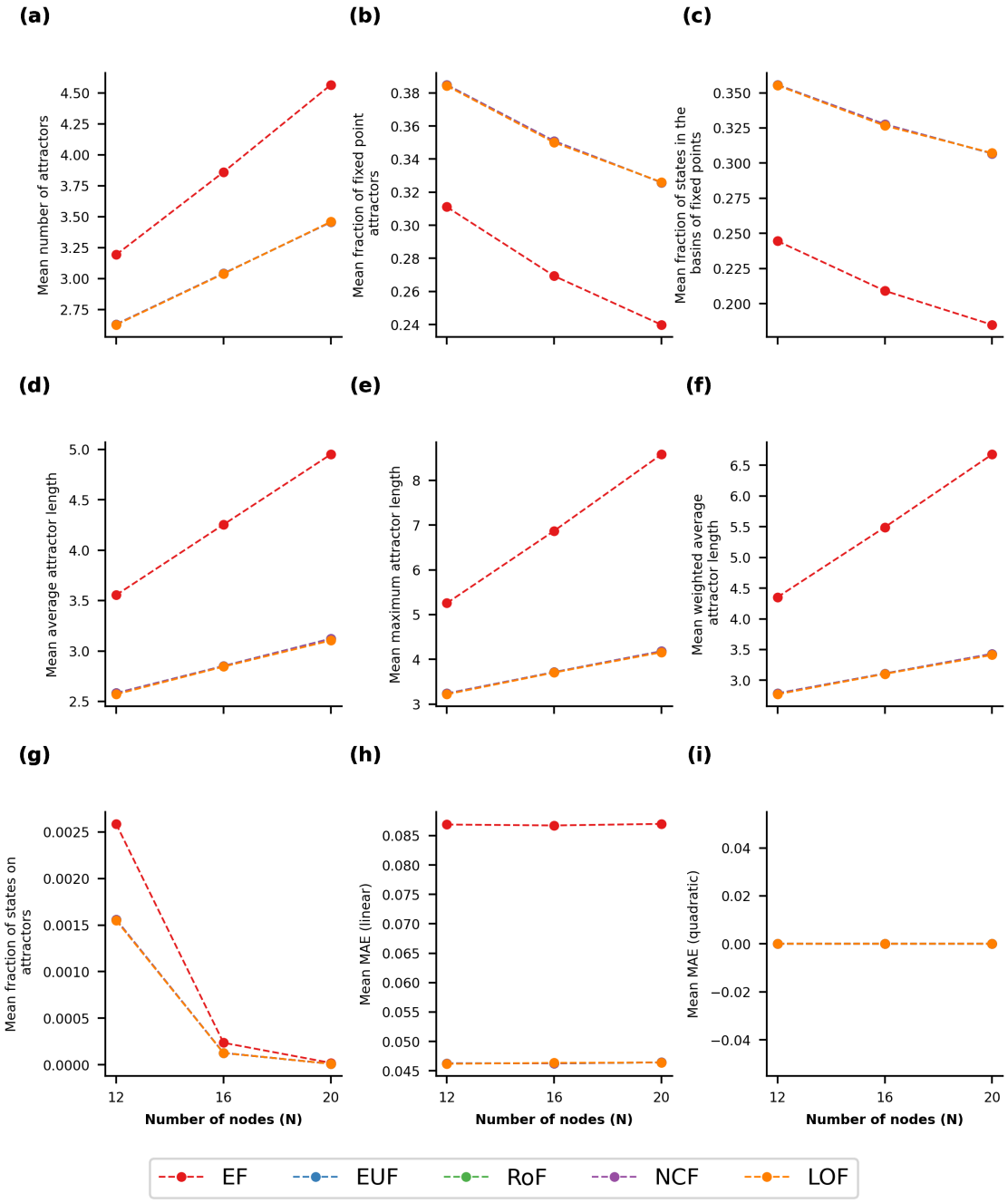


Figure A.64: Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 2$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-P network topology with $k_{in} = 2$.

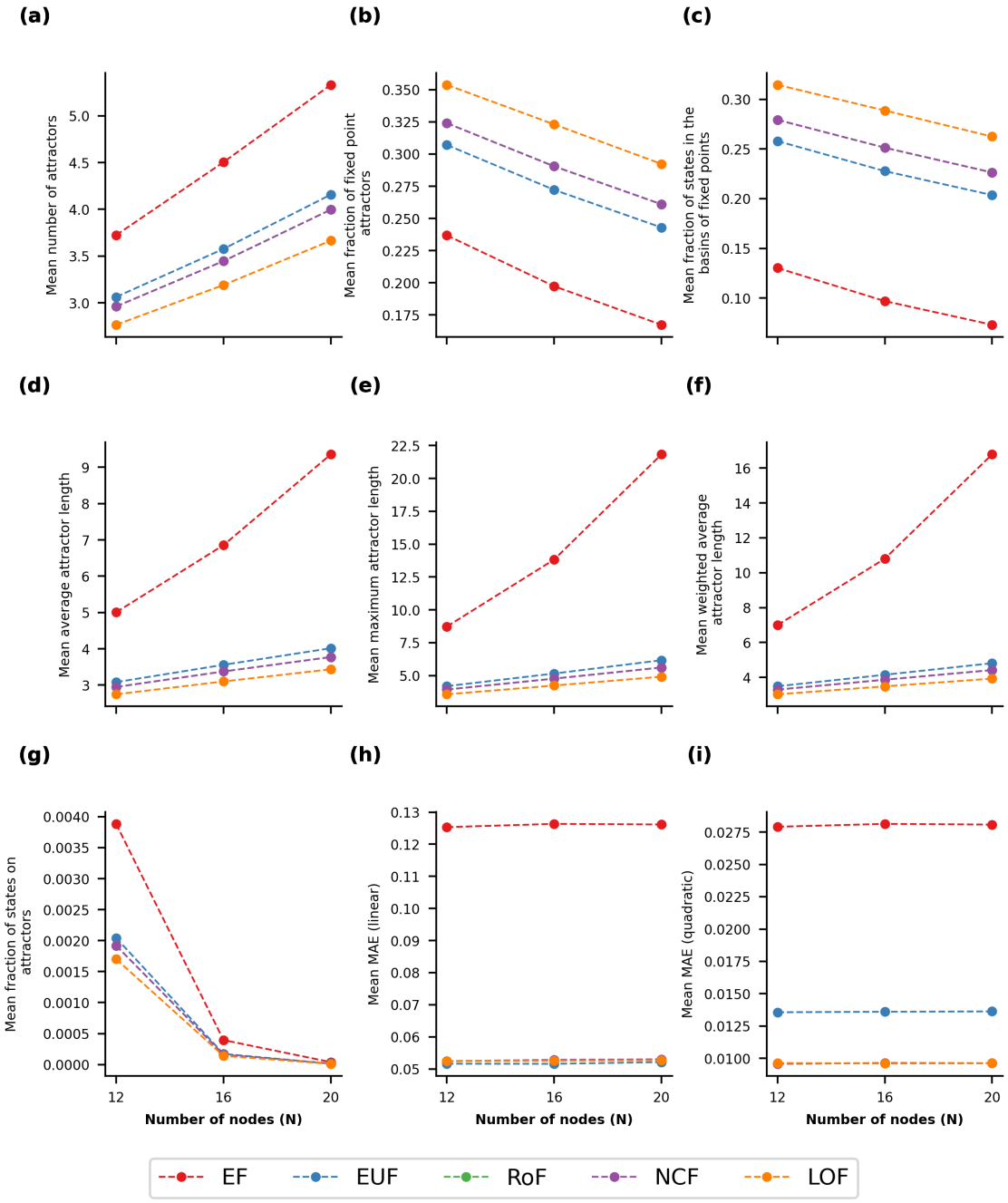


Figure A.65: Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 3$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-P network topology with $k_{in} = 3$.

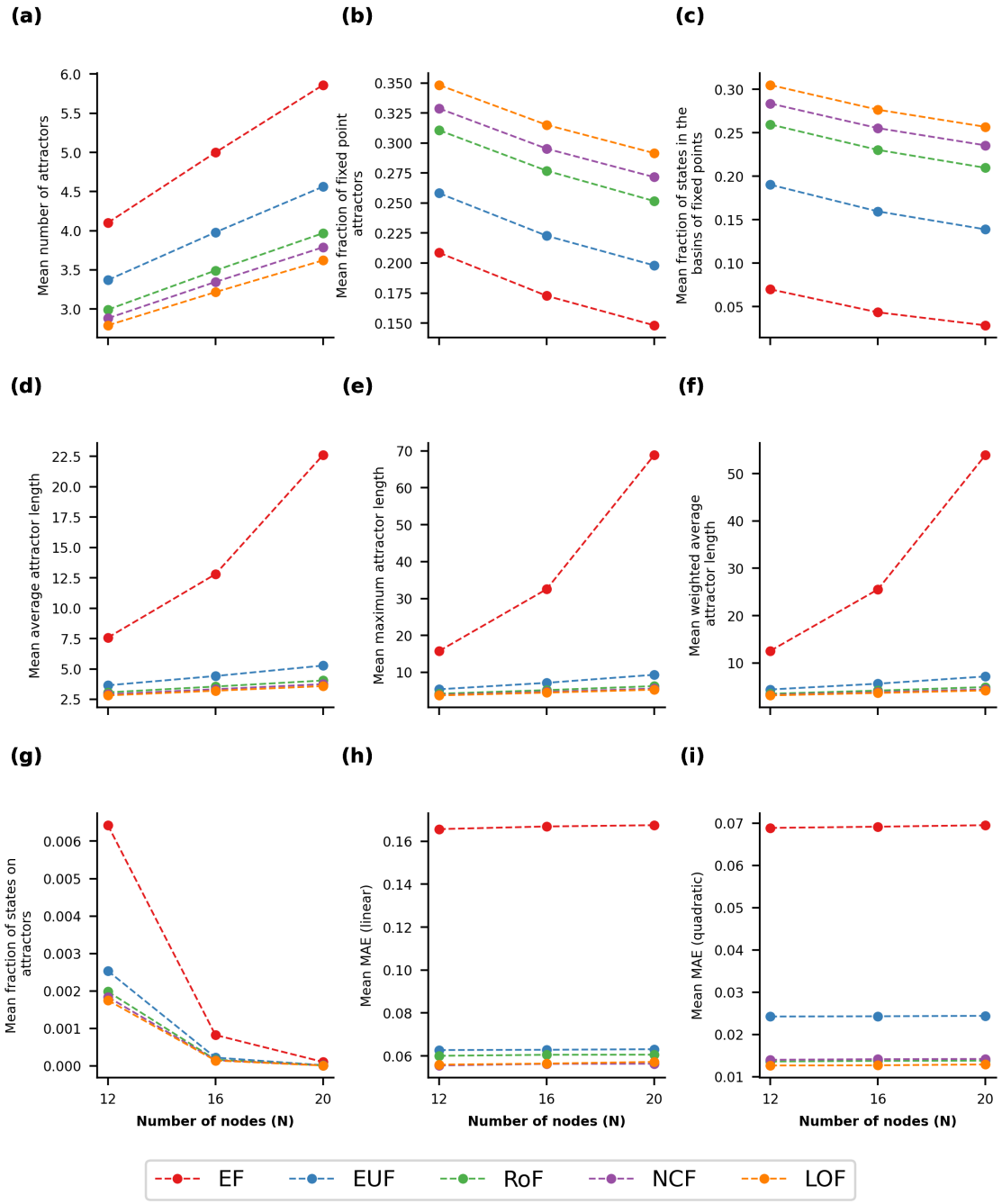


Figure A.66: Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 4$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-P network topology with $k_{in} = 4$.

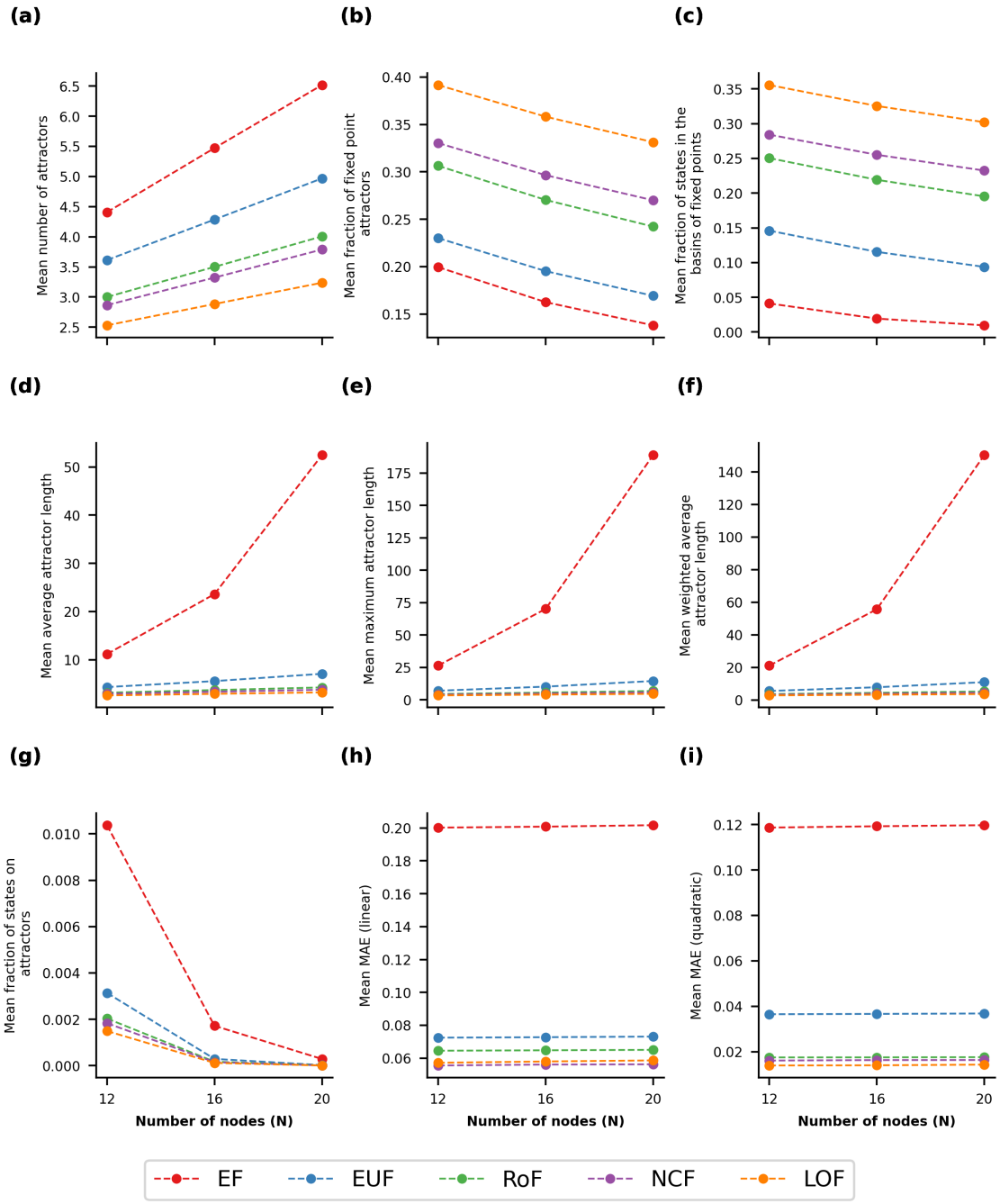


Figure A.67: Mean of the stability observables for different values of N in the case of the R-P topology with $k_{in} = 5$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the R-P network topology with $k_{in} = 5$.

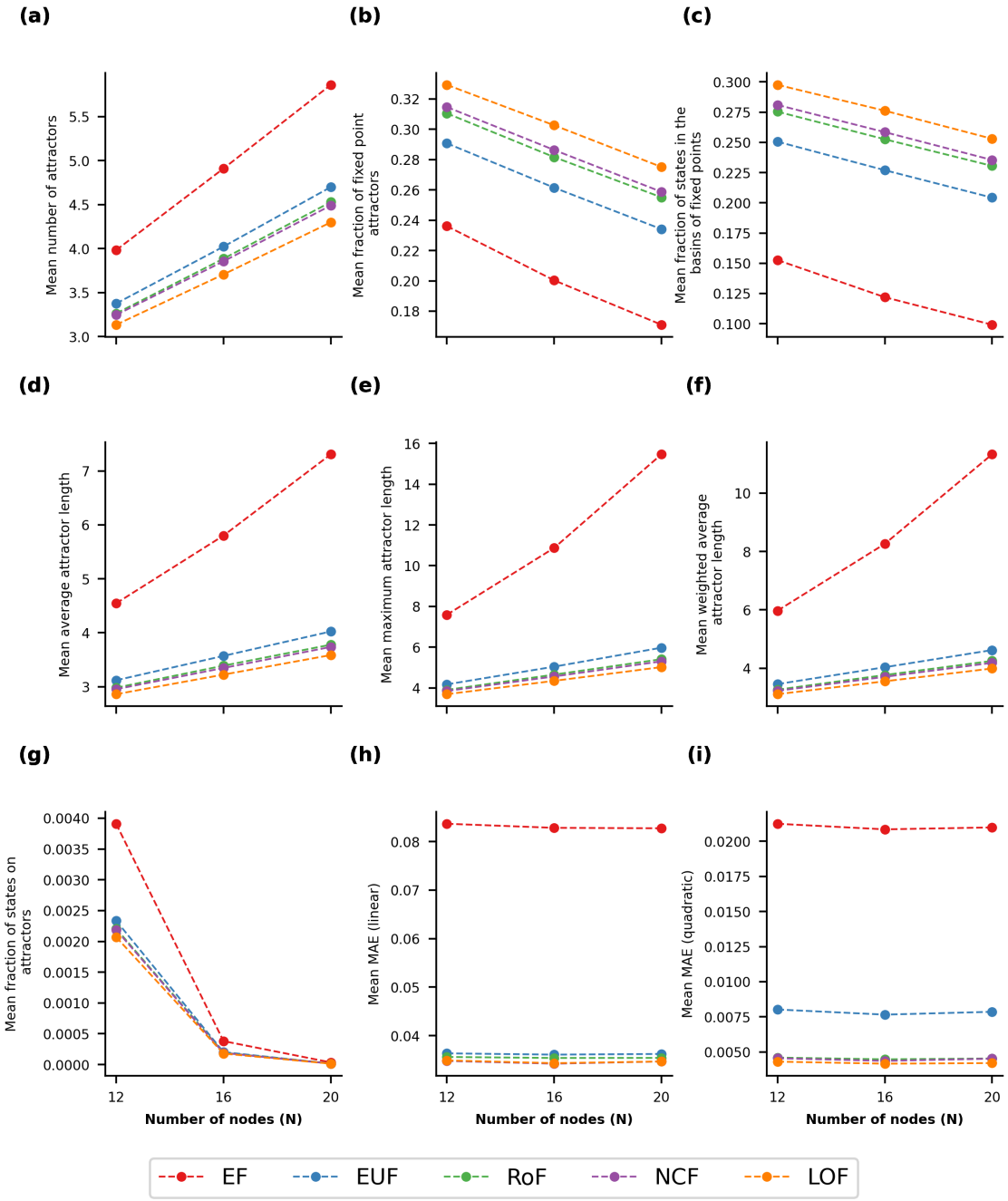


Figure A.68: Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 2$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the P-P network topology with $\lambda = 2$ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean $\lambda = 2$).

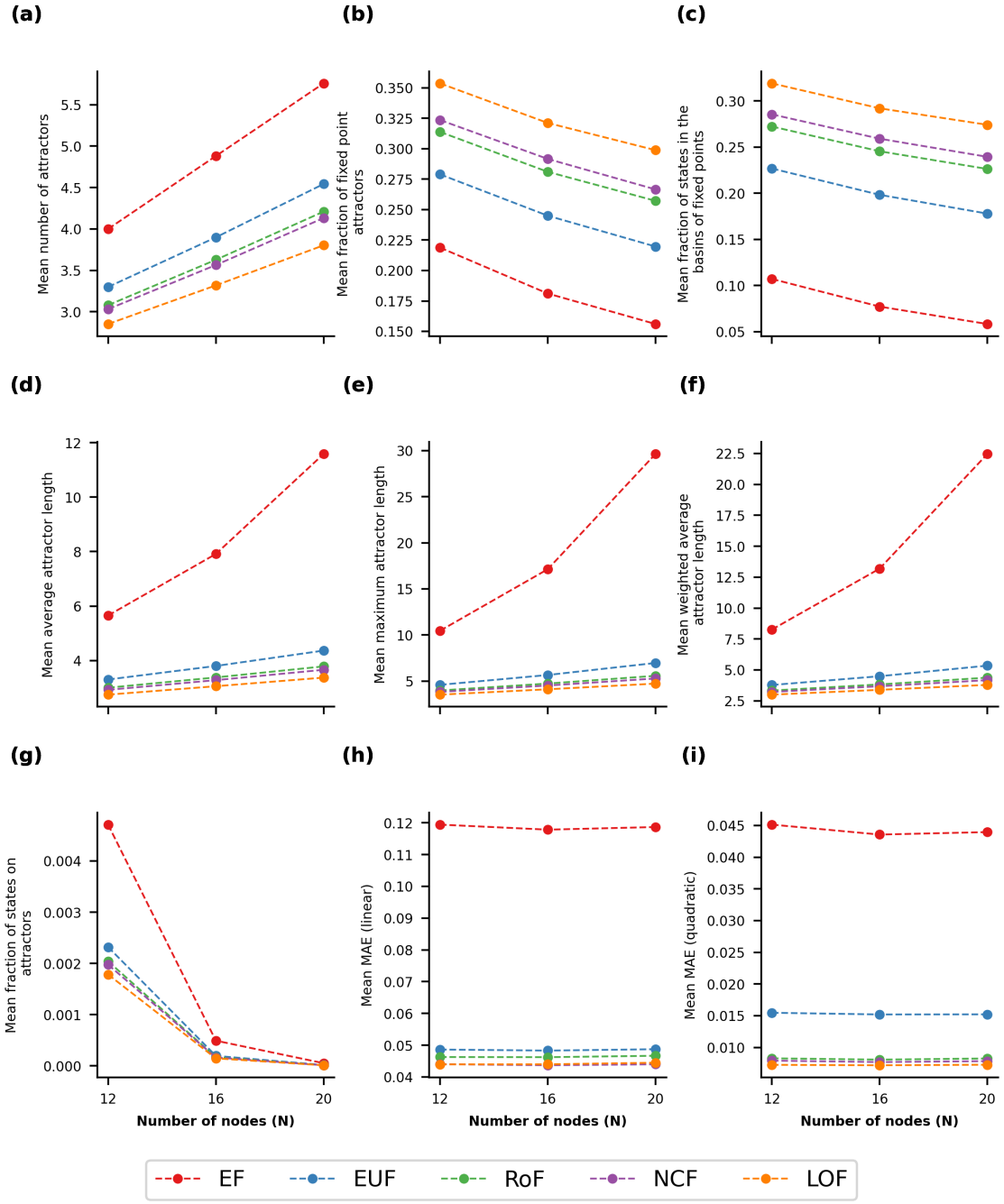


Figure A.69: Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 3$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the P-P network topology with $\lambda = 3$ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean $\lambda = 3$).

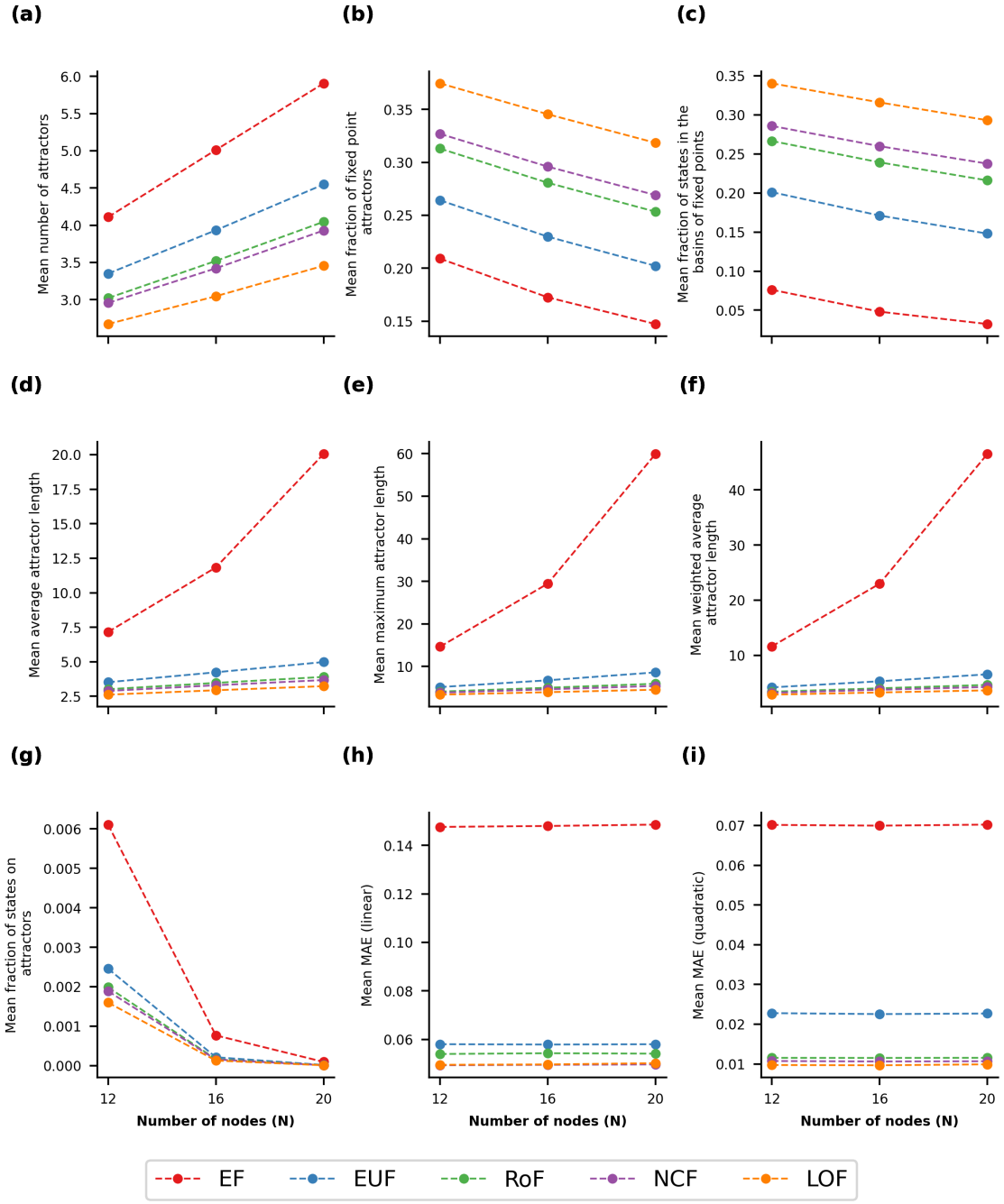


Figure A.70: Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 4$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the P-P network topology with $\lambda = 4$ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean $\lambda = 4$).

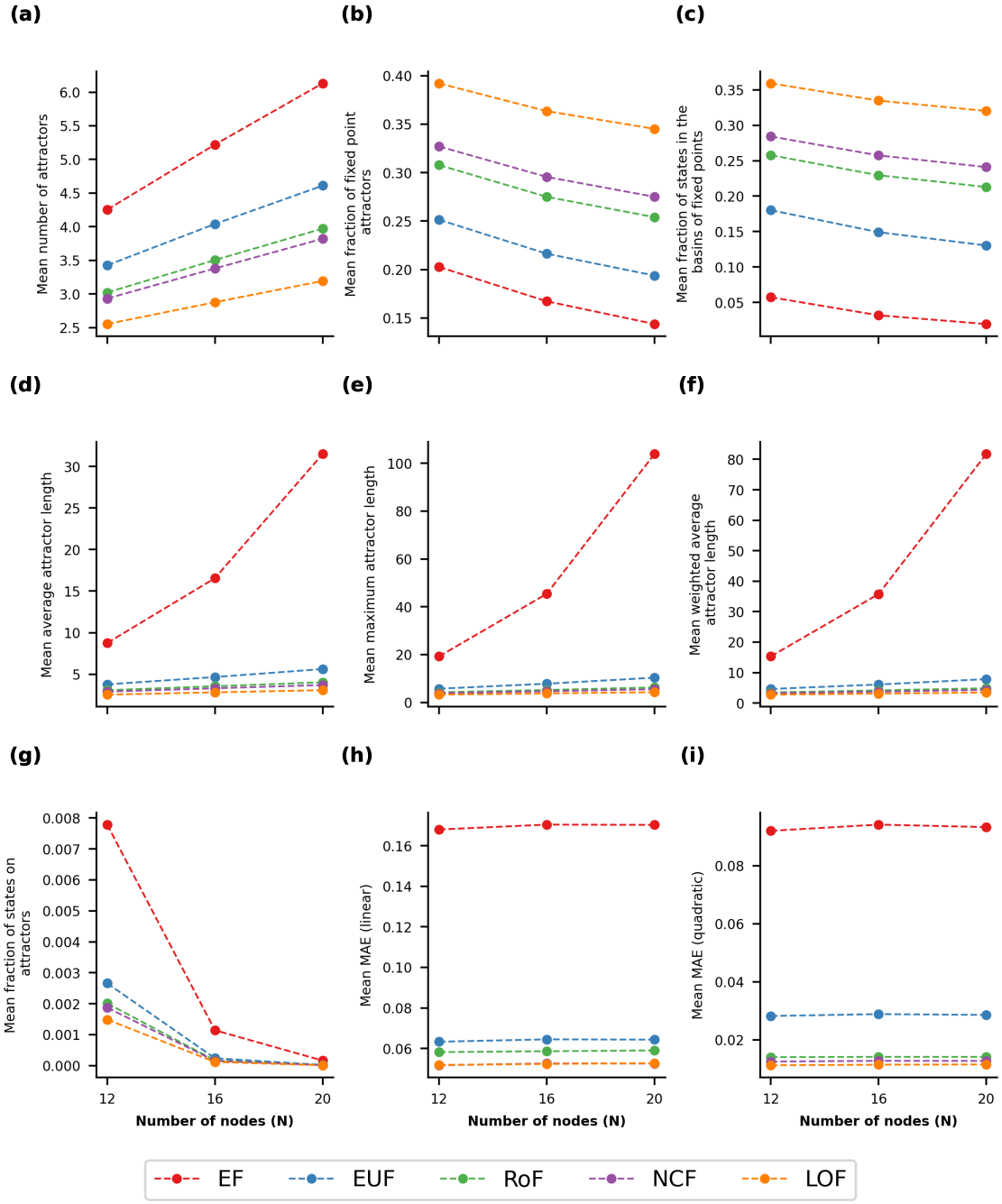


Figure A.71: Mean of the stability observables for different values of N in the case of the P-P topology with $\lambda = 5$. Within each sub-figure, the 5 curves correspond to the 5 different classes of BFs. The data displayed represent the mean value of the corresponding observable at a given value of N for the P-P network topology with $\lambda = 5$ (the degrees of each node in the networks are drawn randomly from a Poisson distribution with mean $\lambda = 5$).

Appendix B

Supplementary Material for Chapter 5

B.1 A brief overview of the ten published Boolean network models used in this study

In Chapter 5, we study the bushiness and convergence of the STG using ten published Boolean network models. These ten Boolean models are listed in Table 5.1 in Chapter 5 and are succinctly described below.

B.1.1 Root Stem Cell Niche GRN (RSCN-GRN)

The RSCN Boolean model (*model A* in [149]) is a reconstructed GRN that controls the differentiation of cells located in the root apex of the plant *Arabidopsis thaliana*, a region referred to as the root stem cell niche. The BFs of this model are provided in Table B.1 and the type of BF for each gene is provided in Table B.21. The nodes (genes), their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.1). This Boolean model has four fixed points corresponding to these cell types: Quiescent center (QC), Vascular initials (VI), Cortex-Endodermis initials (CEI) and Columella epidermis initials (CEpI) (see Table B.11). On applying the various biologically meaningful constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the RSCN-GRN is given in Table B.31. The BF for the AUX node for all models in all four ensembles were chosen to be the same as the one provided in [195].

B.1.2 Epithelial-Mesenchymal Transition GRN (EMT-GRN)

The EMT-GRN is a sub-network of a 150-node reconstructed GRN that controls EMT via mechanosensing–mitogen crosstalk [150]. The BFs of this model are provided in Table B.2 and the type of BF for each gene is provided in Table B.22. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.2). This Boolean model has three fixed points corresponding to these cell types: Epithelial, Mesenchymal and Hybrid (see Table B.12). On applying the various biologically meaningful constraints (see Chapter 2, Section 2.7), the number of BFs allowed at each node of the EMT-GRN is given in Table B.32.

B.1.3 NeoCortex Developmental GRN (NCD-GRN)

The NCD-GRN is a network that models cellular differentiation in the earliest stages of cortical arealisation [151]. The BFs of this model are provided in Table B.3 and the type of BF for each gene is provided in Table B.23. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.3). This Boolean model has two fixed points corresponding to Anterior and Posterior compartments of the developing neocortex (see Table B.13). On applying the various biologically meaningful constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the NCD-GRN is given in Table B.33.

B.1.4 Early Heart Development GRN (EHD-GRN)

The EHD-GRN models two areas of differential gene expression that arise from a common cardiovascular progenitor cell population [152]. The BFs of this model are provided in Table B.4 and the type of BF for each gene is provided in Table B.24. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.4). This Boolean model has two fixed points corresponding to two different areas with differential expression of genes, namely, First Heart Field (FHF) and Second Heart Field (SHF) (see Table B.14). On applying the various biologically meaningful constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the EHD-GRN is given in Table B.34. Exogen BMP2 (exogen_BMP2_II) is an external signal that is 'fixed' to an output value of '1'. Since the number of nodes in this model is 15 but there is one fixed node, the size of the state space is 2^{14}.

B.1.5 Myeloid Differentiation GRN (MD-GRN)

The MD-GRN models hematopoietic stem cell differentiation. In particular, it models myeloid differentiation from common myeloid progenitors to various cell types [37]. The BFs of this model are provided in Table B.5 and the type of BF for each gene is provided in Table B.25. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.5). This Boolean model has four fixed points corresponding to these cell types: granulocytes, erythrocytes, megakaryocytes and monocytes (see Table B.15). On applying the various biologically meaningful constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the MD-GRN is given in Table B.35.

B.1.6 T helper cell differentiation GRN (THC-GRN)

The THC-GRN models the T helper cell network that governs the differentiation of precursor T helper cells (Th0 cells) to effector T helper cells (Th1 and Th2 cells) [110]. The BFs of this model are provided in Table B.6 and the type of BF for each gene is provided in Table B.26. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.6). This Boolean model has three fixed points corresponding to the cell types: Th0, Th1 and Th2 (see Table B.16). On applying the various constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the THC-GRN is given in Table B.36. Of the 23 nodes in this network, four do not have any inputs and are fixed to '0', namely IFN-β, IL-12, IL-18 and TCR. Thus the size of the state space is 2^{19}.

B.1.7 EGFR signalling pathway (EGFR-GRN)

The EGFR-GRN models the regulation of cell division and apoptosis by the EGFR signalling pathway and a BRCA1/TP53 DNA damage response module [153]. The BFs of this model are provided in Table B.7 and the type of BF for each gene is provided in Table B.27. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.7). This Boolean model has two fixed points corresponding to these cell states: apoptosis and cell division (see Table B.17). On applying the various constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the EGFR-GRN is given in Table B.37.

B.1.8 Epithelial-Mesenchymal Transition (with Senescence) (EMT-Senescence-GRN)

The EMT-Senescence-GRN models epithelial to mesenchymal transition in the context of epithelial cancer [154]. The BFs of this model are provided in Table B.8 and the type of BF for each gene is provided in Table B.28. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.8). This Boolean model has three fixed points corresponding to these cell types: epithelial cells, senescent epithelial cells and stem-like mesenchymal cells with a tumorigenic potential (see Table B.18). On applying the various constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the EMT-Senescence-GRN is given in Table B.38.

B.1.9 Flower Organ Specification GRN (FOS-GRN)

The FOS-GRN controls early flower development in *Arabidopsis thaliana*, in particular, the primordial cells of inflorescence, sepals, petals, stamens and carpels. The BFs of this model are provided in Table B.9 and the type of BF for each gene is provided in Table B.29. The nodes, their inputs and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.9). This Boolean model has ten fixed points corresponding to these cell types: inflorescence (4 cell types), sepal (1 cell type), petal (2 cell types), stamen (2 cell types) and carpel (1 cell type) (see Table B.19). On applying the various constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the FOS-GRN is given in Table B.39. Since the computation of the Z-parameters was computationally expensive for seven and eight input BFs, such as AP3 and AG respectively, we sampled 100,000 models for each of the four ensembles.

B.1.10 Gonadal Sex Determination GRN (GSD-GRN)

The Gonadal sex determination involves the decision taken by bipotential gonadal primordium (BGP) to differentiate to testis or ovaries. In particular, the GSD-GRN governs cellular differentiation from Sertoli progenitor cells (SPC) and granulosa progenitor cells (GPC) towards Sertoli (in testis) and granulosa (in ovaries) cells respectively [156]. The BFs of this model are provided in Table B.10 and the type of BF for each gene is provided in Table B.30. The nodes, their inputs

and the sign of the regulatory interactions in the network are all derived from the BoolNet file (see Table B.10). This Boolean model has two fixed points corresponding to these cell types: sertoli and granulosa (see Table B.20). On applying the various constraints (see Chapter 2, Section 2.7) the number of BFs allowed at each node of the GSD-GRN is given in Table B.40. Since the computation of the Z-parameters was computationally expensive for seven and eight input BFs, namely, SOX9 and SRY respectively, we sampled 100,000 models for each of the four ensembles.

B.2 Properties of Z-parameters

Let f be a k-input BF f. We then have the following properties.

Property B.1. Z_{left} of f is invariant under the negation of any of its inputs.

Proof: In the truth table representation of f, negation of the i^{th} input x_i will swap the outputs belonging to the lower and upper half sub-blocks of each 2^{k-i+1}-block (see Chapter 2, Section 2.4.1). Since Z_{left} is completely determined by the n_i's and k (see Chapter 5, Eq. (5.4)), the negation of the i^{th} input will swap the output values of the associated sub-blocks, keeping the value of n_i unchanged, and consequently leaving Z_{left} unchanged. The negation of multiple inputs will also keep the value of Z_{left} unchanged since the negation of multiple inputs can be achieved by successively negating single inputs.

Property B.2. Z_{left} of f is invariant under complementation.

Proof: Let $g = \overline{f}$ be the complement of f. Then for any input combination, if f is T, g is $\overline{T}$. Consider

$$\begin{aligned} f(x_1, x_2, \ldots, x_{k-i}, 0, *, \ldots, *, *) &= T \\ f(x_1, x_2, \ldots, x_{k-i}, 1, *, \ldots, *, *) &= \overline{T} \end{aligned}$$

where $T \in \{0, 1\}$ and $*$ may be 0 or 1. Then,

$$\begin{aligned} g(x_1, x_2, \ldots, x_{k-i}, 0, *, \ldots, *, *) &= \overline{T} \\ g(x_1, x_2, \ldots, x_{k-i}, 1, *, \ldots, *, *) &= \overline{\overline{T}} = T \end{aligned}$$

As a result, n_i's are the same for f and g for any $i \in \{1, 2, ..., k\}$ and consequently so are the R_i values. Thus Z_{left} values of f and g are equal.

Property B.3. The maximum value of Z_{left} over all k-input BFs with bias P is:

$$Z_{left} = \begin{cases} \dfrac{P}{2^{k-1}} & \text{if } P \leq 2^{k-1} \\ 2 - \dfrac{P}{2^{k-1}} & \text{if } P > 2^{k-1} \end{cases} \tag{B.1}$$

Proof: For a k-input BF f with bias $P \leq 2^{k-1}$, consider the vector $\mathbf{n} = (n_1, n_2, \ldots, n_{k-1}, n_k)$ as defined in Section 5.3.1 in Chapter 5. Recall that n_i counts the number of lines in the truth table that arise in CC-2^{k-i+1} blocks, each of which consists of lines where half of the output values are 0s and the other half are 1s. Then, using the fact that all CC-blocks (of all sizes) are non overlapping, the total number of these CC-block associated lines is twice the number of 1s in those same lines. This immediately leads to

$$\sum_{i=1}^{i=k} n_i \leq 2P$$

i.e., the sum of the elements of the vector $\mathbf{n}$ must be at most $2P$.

Consider now the expression of Z_{left},

$$\begin{aligned} Z_{left} &= R_k + R_{k-1}(1 - R_k) + R_{k-2}(1 - R_{k-1})(1 - R_k) + \\ &\qquad R_{k-3}(1 - R_{k-2})(1 - R_{k-1})(1 - R_k) + ... \\ &\leq R_k + R_{k-1} + R_{k-2} + \ldots + R_1 \\ &= \frac{n_k}{2^k} + \frac{n_{k-1}}{2^k} + \ldots + \frac{n_1}{2^k} \\ &= \frac{n_k + n_{k-1} + \ldots + n_1}{2^k} \\ &\leq \frac{2P}{2^k} \\ &= \frac{P}{2^{k-1}} \end{aligned}$$

showing that Z_{left} is bounded from above by $P/2^{k-1}$ (if $P \leq 2^{k-1}$). In addition, for each $P \leq 2^{k-1}$ there are functions f that reach this maximum value. For instance for any P, consider a function

of the type

$$f = 000\ldots0\overbrace{0101\ldots01}^{P\text{ '01' pairs}}$$

Clearly for such a choice, $n_k = 2P$ and $n_i = 0\ \forall\ i \neq k$. This results in $Z_{left} = P/2^{k-1}$.

So, for all bias values $P \leq 2^{k-1}$, the upper bound of Z_{left} is $P/2^{k-1}$. From Property (B.2) it follows that for any function f with bias value $P > 2^{k-1}$, it would have the same value of Z_{left} as its complement function. Now, consider a BF f with bias $P > 2^{k-1}$. Then the maximum possible value of Z_{left} is the same for f and $\overline{f}$ (complement of f). However, the bias of $\overline{f}$ is $2^k - P$ and since $P > 2^{k-1}$, $2^k - P < 2^{k-1}$. Hence, the maximum value of Z_{left} for $\overline{f}$ is $\frac{2^k-P}{2^{k-1}} = 2 - \frac{P}{2^{k-1}}$. Finally, the maximum possible value of Z_{left} for f is then $2 - \frac{P}{2^{k-1}}$. This proves that the maximum value of Z_{left} for any k-input BF with bias P is:

$$Z_{left} = \begin{cases} \dfrac{P}{2^{k-1}} & \text{if } P \leq 2^{k-1} \\ 2 - \dfrac{P}{2^{k-1}} & \text{if } P > 2^{k-1} \end{cases} \tag{B.2}$$

Property B.4. NCFs have the maximum Z_{max} for any k-input BF with odd bias P.

Proof: Consider the output column of the truth table for the following k-input BF with all inputs activatory and having a bias P such that $P < 2^{k-1}$:

$$f = 000\ldots0\overbrace{0101\ldots01}^{P\text{ '01' pairs}}$$

It is not difficult to see that this BF is an NCF with

$$\begin{aligned} & n_k = 2P \text{ and } n_i = 0\ \forall\ i \in \{1, 2, \ldots, k-1\} \\ \implies \quad & R_k = \frac{2P}{2^k} = \frac{P}{2^{k-1}} \text{ and } R_i = 0\ \forall\ i \in \{1, 2, \ldots, k-1\} \\ \implies \quad & Z_{left} = \frac{P}{2^{k-1}} + 0\left(1 - \frac{P}{2^{k-1}}\right) + 0\left(1 - \frac{P}{2^{k-1}}\right)(1-0) + \ldots \\ \implies \quad & Z_{left} = \frac{P}{2^{k-1}} \end{aligned}$$

We thus see that this NCF has reached the maximum possible value (given the results above) and thus its Z_{max} value is precisely Z_{left}. Interestingly, this same value is that of all NCFs with that

same bias P because the set of all NCFs of a given bias are obtained by permuting and negating variables of this particular NCF, a result that was shown in [51]. The desired result then follows from the fact that Z_{left} is invariant under negations. Consider now the case of a function g with bias $P > 2^{k-1}$. The bias of the complement of our function is then $2^k - P$; this complementation does not change Z_{left}. Hence, the value of Z_{max} of our complemented function is $\frac{2^k-P}{2^{k-1}} = 2 - \frac{P}{2^{k-1}}$ (since $2^k - P < 2^{k-1}$). Consequently, the value of Z_{max} of g is $2 - \frac{P}{2^{k-1}}$. In summary, the Z_{max} of a k-input NCF with a bias P (which is necessarily odd) is given by:

$$Z_{max} = \begin{cases} \frac{P}{2^{k-1}} \text{ if } P < 2^{k-1}, \\ 2 - \frac{P}{2^{k-1}} \text{ if } P > 2^{k-1} \end{cases}$$

Property B.5. The minimum value for Z_{left} for any non-constant k-input BF with bias P is $1/2^{k-1}$.

Proof: The minimum possible value of Z_{left} across all k-input BFs is 0 (in particular for 'constant' BFs with bias 0 and bias 2^k). For these BFs, $R_i = 0 \; \forall \; i \in \{1, 2, \ldots k\}$. The minimum possible value of Z_{left} across all non-constant k-input BFs is achieved by BFs with a single CC-2^1-block. For these BFs, Z_{left} is $2/2^k = 1/2^{k-1}$ since $n_k = 2$ and $n_i = 0 \; \forall \; i \in \{1, 2, \ldots, k-1\}$. There are functions f that reach this minimum value (for instance bias 1 and bias $2^k - 1$ functions).

Property B.6. NCFs have the minimum Z_{min} for any k-input BF with odd bias P.

Proof: Consider the output column of the truth table of a k-input NCF with bias P where all inputs are 'activatory':

$$f = \underbrace{000\ldots0}_{(2^k-P)\text{ 0's}} \overbrace{11\ldots1}^{P\text{ 1's}}$$

Since P is odd for NCFs [51, 196] we have,

$$\begin{aligned} & n_k = 2 \text{ and } n_i = 0 \; \forall \; i \in \{1, 2, \ldots, k-1\} \\ \implies \quad & R_k = \frac{2}{2^k} = \frac{1}{2^{k-1}} \text{ and } R_i = 0 \; \forall \; i \in \{1, 2, \ldots, k-1\} \\ \implies \quad & Z_{left} = \frac{1}{2^{k-1}} \end{aligned}$$

Z_{min} is the minimum of the Z_{left} values over all permutations of a BF and Z_{left} is invariant under negation of any of its inputs. Again, there is only one equivalence class per bias for NCFs (equiv-

alence under both permutations and negations), so Z_{min} for any k-input NCF is given by $1/2^{k-1}$. Hence, NCFs have the minimum Z_{min} for any k-input BF with odd bias P. The value of Z_{min} for any k-input NCF is independent of bias (P) and its value is $1/2^{k-1}$.

Property B.7. The value of Z_{mid} for any k-input and P-bias NCF is:

$$Z_{mid} = \begin{cases} \dfrac{1+P}{2^k} & \text{if } P < 2^{k-1} \\ 1 - \dfrac{P-1}{2^k} & \text{if } P > 2^{k-1} \end{cases} \tag{B.3}$$

Proof: Z_{mid} of a BF is defined as the average of its Z_{max} and Z_{min} values for any k-input NCF with bias P (see Properties B.4 and B.6) is given by:

$$\begin{aligned} Z_{min} &= \frac{1}{2^{k-1}} \\ Z_{max} &= \begin{cases} \frac{P}{2^{k-1}} \text{ if } P < 2^{k-1}, \\ 2 - \frac{P}{2^{k-1}} \text{ if } P > 2^{k-1} \end{cases} \end{aligned}$$

Hence,

$$\begin{aligned} Z_{mid} &= \frac{Z_{max} + Z_{min}}{2} \\ &= \begin{cases} \frac{1}{2}\left(\frac{1}{2^{k-1}} + \frac{P}{2^{k-1}}\right) \text{ if } P < 2^{k-1}, \\ \frac{1}{2}\left(\frac{1}{2^{k-1}} + 2 - \frac{P}{2^{k-1}}\right) \text{ if } P > 2^{k-1} \end{cases} \\ &= \begin{cases} \frac{1+P}{2^k} \text{ if } P < 2^{k-1}, \\ 1 - \frac{P-1}{2^k} \text{ if } P > 2^{k-1} \end{cases} \end{aligned}$$

B.3 Results on non-equivalent permutations of the $k!$ permutations of Boolean functions

B.3.1 Preliminaries

Definition 1 (Left coset)**.** Let G be a group and H be a subgroup of G. Then for any element $a \in G$ the set $\{ah \mid h \in H\}$ is called the *left coset* of H in G containing a and is denoted by aH. In this case, a is called the coset representative of aH.

Note: The cardinality of any left coset of H is equal to the cardinality of H.

Definition 2 (Left group action)**.** Let G be a group with e as its identity element and X be a set, then the (left) *group action* α of G on X is the function

$$\alpha : G \times X \to X$$

that obeys the two axioms:

- $\alpha(e, x) = e.x = x \; \forall \; x \in X$
- $\alpha(g_1, \alpha(g_2, x)) = \alpha(g_1 g_2, x)$ i.e. $g_1.(g_2.x) = (g_1 g_2).x \; \forall \; x \in X$ and $g_1, g_2 \in G$

Definition 3 (Orbits)**.** Suppose that G is a group that acts on a set X on the left. Let x be an element of X. Then, the *orbit* of x under G is defined by

$$Orb_G(x) = \{g.x \mid g \in G\} = \{y \in X \mid y = g.x \text{ for some } g \in G\}$$

Definition 4 (Stabilizers)**.** Suppose that G is a group that acts on a set X on the left. Then, for an element $x \in X$, the *stabilizer* of x is defined to be the set

$$Stab_G(x) = \{g \in G \mid g.x = x\}$$

Note: Stabilizers are subgroups of G.

Theorem 1 (Orbit-Stabilizer theorem [197])**.** Suppose that G is a finite group acting (on the left) on a set X. Then, for any $x \in X$,

$$|Orb_G(x)||Stab_G(x)| = |G|$$

B.3.2 Two theorems on non-equivalent permutations of the $k!$ permutations of BFs and their proofs

Theorem 2. For any k-input BF, the total number of non-equivalent permutations divides $k!$.

Proof. We approach the problem from a group theoretic perspective. First we define a map

$$\alpha : G \times T \rightarrow T$$

where G is the symmetric group S_k, *i.e.*, the set of all permutations of k elements (k is the number of inputs to the BF).

We define T as the set of sets by:

$$T = \{t \mid t = \{(X_i, b_i) \mid b_i = 0 \text{ or } 1 \ \ \forall i = 1, 2, ..., 2^k\}\}$$

where, $X_i = (x_1, x_2, ..., x_k)_i$ corresponds to the input values of the i^{th} row of the truth table. Clearly, each element $t \in T$ can be associated with the truth table of a BF of k-inputs. Each element of a set t is an ordered pair, the first component denoting some assignment of input variables of the truth table and second component denoting the associated output value. Now, the cardinality of T is $|T| = 2^{2^k}$. For a better understanding of the set T let us take one example. Consider the case of $k = 2$. Then, the set T would be:

$$T = \{t \mid t = \{((x_1, x_2)_i, b_i) \mid b_i = 0 \text{ or } 1 \ \ \forall i = 1, 2, 3, 4\}\}$$

Let us take one element t (for $b_1 = 0,\ b_2 = 1,\ b_3 = 0,\ b_4 = 0$) from the set T as

$$t = \{((0,0),0),((0,1),1),((1,0),0),((1,1),0)\}$$

Clearly, one can observe that t is a set that corresponds to the truth table for a specific BF. Each element of t is 2-tuple whose first component is an input assignment and the second component is the associated output. The set T consists of all such possible t's. So, in this case the cardinality of the set T is $2^{2^2} = 16$.

Now we define the map $\alpha : S_k \times T \to T$ by

$$\begin{aligned} \sigma.t &= t' \\ \text{or, } \sigma.\{((x_1, x_2, ..., x_k)_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} &= \{((x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(k)})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \end{aligned}$$

where, $\sigma \in S_k$ and $t = \{((x_1, x_2, ..., x_k)_i, b_i) \text{ for } i = 1, 2, ..., 2^k\}$, $t' = \{((x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(k)})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \in T$.

Now, we intend to show that the defined map α is a left group action of S_k on T. For that, we need to show that the map α obeys the axioms stated in definition 2.

Let, $e \in S_k$ be the identity permutation and $t \in T$. Then,

$$\begin{aligned} & e.t \\ = \ & e.\{((x_1, x_2, ..., x_k)_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\ = \ & \{((x_{e(1)}, x_{e(2)}, ..., x_{e(k)})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\ = \ & \{((x_1, x_2, ..., x_k)_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\ = \ & t \end{aligned}$$

Again, for $\sigma_1, \sigma_2 \in S_k$ and $t \in T$,

$$\begin{aligned}
& \sigma_1.(\sigma_2.t) \\
= \; & \sigma_1.\{((x_{\sigma_2(1)}, x_{\sigma_2(2)}, ..., x_{\sigma_2(k)})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\
= \; & \{((x_{\sigma_1(\sigma_2(1))}, x_{\sigma_1(\sigma_2(2))}, ..., x_{\sigma_1(\sigma_2(k))})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\
= \; & \{((x_{(\sigma_1\sigma_2)(1)}, x_{(\sigma_1\sigma_2)(2)}, ..., x_{(\sigma_1\sigma_2)(k)})_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\
= \; & (\sigma_1\sigma_2).\{((x_1, x_2, ..., x_k)_i, b_i) \text{ for } i = 1, 2, ..., 2^k\} \\
= \; & (\sigma_1\sigma_2).t
\end{aligned}$$

Hence, the defined map α satisfies both of the conditions for being a group action (left).

Now, the orbit of an element $t \in T$ under S_k is given by (see definition 3)

$$Orb_{S_k}(t) = \{\sigma.t \mid \sigma \in S_k\}$$

So, the set $Orb_{S_k}(t)$ provides all possible elements that correspond to non-equivalent permutations of the BF associated with t. In other words, the cardinality of the set $Orb_{S_k}(t)$, $|Orb_{S_k}(t)|$, is the number of non-equivalent permutations of the BF associated with t.

Now, by the Orbit-stabilizer theorem (see theorem 1) we can say that,

$$\begin{aligned}
|Orb_{S_k}(t)||Stab_{S_k}(t)| &= |S_k| \\
\text{or, } |Orb_{S_k}(t)||Stab_{S_k}(t)| &= k!
\end{aligned}$$

Hence, $|Orb_{S_k}(t)|$ divides $k!$. In other words, the number of non-equivalent permutations of any k-input BF is always a divisor of k!. This proves the theorem. □

Theorem 3. In $k!$ permutations of a k-input BF, each of the m non-equivalent permutations occurs an equal number of times.

Proof. We consider all the quantities defined for Theorem 2. We will prove that in the $k!$ possible permutations of a k-input BF, each non-equivalent permutation appears $|Stab_{S_k}(t)|$ times where, the BF is associated with the element $t \in T$.

The stabilizer of an element $t \in T$ under S_k is given by (see definition 4)

$$Stab_{S_k}(t) = \{\sigma \in S_k \mid \sigma.t = t\}$$

Evidently, $Stab_{S_k}(t)$ corresponds to the *symmetry* (or *invariance* or *automorphism*) [198] group of the BF. Assume that there are m non-equivalent permutations among the $k!$ permutations of the BF associated with t. Then, we can consider the orbit of t,

$$Orb_{S_k}(t) = \{t_1, t_2, ..., t_m\}$$

We will show that there are $Stab_{S_k}(t)$ many σ_j's $\forall\, j = 1, 2, ..., m$ such that the following conditions hold,

$$\begin{aligned} \sigma_1.t &= t_1 \\ \sigma_2.t &= t_2 \\ &\vdots \\ \sigma_m.t &= t_m \end{aligned}$$

The element t always belongs to the set $Orb_{S_k}(t)$ since, $e.t = t$ where e is the identity permutation. So, without loss of generality, let us consider $t_1 = t$. We need to find the number of σ_1's such that $\sigma_1.t = t$. From the definition of stabilizer of an element it follows that we have $|Stab_{S_k}(t)|$ many σ_1's.

Now, consider an element from $Orb_{S_k}(t)$ different from t (if it exists), say t_2. Then, we want to know how many σ_2's satisfy

$$\sigma_2.t = t_2 \tag{B.4}$$

Since $Stab_{S_k}(t)$ is a subgroup of S_k, we can consider the following left coset of $Stab_{S_k}(t)$ in S_k with the coset representative σ_2:

$$\sigma_2 Stab_{S_k}(t) = \{\sigma_2\sigma \mid \sigma \in Stab_{S_k}(t)\}$$

Let us take an element $\sigma_2\sigma \in \sigma_2 S\,tab_{S_k}(t)$. Then,

$$\begin{aligned}(\sigma_2\sigma).t &= \sigma_2.(\sigma.t)\\ &= \sigma_2.t \text{ [since, } \sigma \in S\,tab_{S_k}(t)]\\ &= t_2 \text{ (from } B.4)\end{aligned}$$

Hence, all the elements of $\sigma_2 S\,tab_{S_k}(t)$ satisfy equation B.4.

Now, we will show that all elements $\sigma' \in S_k$ for which equation B.4 is satisfied (i.e. $\sigma'.t = t_2$), $\sigma' \in \sigma_2 S\,tab_{S_k}(t)$. We have,

$$\begin{aligned}\sigma'.t &= t_2\\ \text{or, } \sigma'.t &= \sigma_2.t \text{ (from } B.4)\\ \text{or, } \sigma'^{-1}.(\sigma'.t) &= \sigma'^{-1}.(\sigma_2.t)\\ \text{or, } (\sigma'^{-1}\sigma').t &= (\sigma'^{-1}\sigma_2).t\\ \text{or, } e.t &= (\sigma'^{-1}\sigma_2).t\\ \text{or, } t &= (\sigma'^{-1}\sigma_2).t\end{aligned}$$

From the definition of $S\,tab_{S_k}(t)$ it follows that, $\sigma'^{-1}\sigma_2 \in S\,tab_{S_k}(t)$. Since $S\,tab_{S_k}(t)$ is a subgroup of S_k, we also have $(\sigma'^{-1}\sigma_2)^{-1} = \sigma_2^{-1}\sigma' \in S\,tab_{S_k}(t)$. Taking $\sigma_2^{-1}\sigma' = b$, we then have:

$$\begin{aligned}\sigma' &= \sigma_2 b \text{ where, } b \in S\,tab_{S_k}(t)\\ \text{or, } \sigma' &\in \sigma_2 S\,tab_{S_k}(t)\end{aligned}$$

Hence, the number of σ_2's for which equation (B.4) is satisfied is $|\sigma_2 S\,tab_{S_k}(t)|$. Now, since $\sigma_2 S\,tab_{S_k}(t)$ is a left coset of $S\,tab_{S_k}(t)$, their cardinalities are identical, $|\sigma_2 S\,tab_{S_k}(t)| = |S\,tab_{S_k}(t)|$. So, the number of σ_2's for which equation (B.4) is satisfied is precisely $|S\,tab_{S_k}(t)|$.

In a similar manner, we can show that there are $|S\,tab_{S_k}(t)|$ many σ_3's, σ_4's, ..., σ_m's such that the conditions $\sigma_3.t = t_3$, $\sigma_4.t = t_4$, ..., $\sigma_m.t = t_m$ respectively are satisfied. Hence, all non-equivalent permutations of a k-input BF associated with t, appears $|S\,tab_{S_k}(t)|$ times in the $k!$ permutations. This proves the theorem. □

B.4 Supplementary tables for Chapter 5

Table B.1: Boolean functions of the given RSCN-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	SCR	(SHR & SCR & !JKD & !MGP) | (SHR & SCR & JKD & !MGP) | (SHR & SCR & JKD & MGP)
2	PLT	ARF
3	ARF	!AUXIAA
4	AUXIAA	!AUX
5	AUX	!AUX | AUX
6	SHR	SHR
7	JKD	SHR & SCR
8	MGP	SHR & SCR & !WOX5
9	WOX5	(ARF & SHR & SCR & !MGP & !WOX5) | (ARF & SHR & SCR & !MGP & WOX5) | (ARF & SHR & SCR & MGP & WOX5)

Table B.2: Boolean functions of the given EMT-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	SNAI1	((ZEB1_H | ZEB1) & !miR_34) | (ZEB1_H & ZEB1)
2	LEF1	((ZEB1 & (!miR_200)) | ZEB1_H)
3	Twist	SNAI1 & !miR_34
4	SNAI2	Twist & (SNAI2 | (N_bcatenin_H & LEF1))
5	ZEB1	SNAI2 | (b_catenin_TCF4 & (!miR_200))
6	N_bcatenin_H	!miR_34 & !miR_200
7	ZEB1_H	ZEB1 & ((N_bcatenin_H & LEF1) & (SNAI2 | (!miR_200)))
8	b_catenin_TCF4	N_bcatenin_H & SNAI1 & SNAI2
9	miR_34	(!SNAI1) | (!(ZEB1 | ZEB1_H))
10	miR_200	!(Twist & ZEB1_H & SNAI1) & !(ZEB1 & (!miR_200))
11	Ecadherin_mRNA	!((((ZEB1_H & ZEB1) & SNAI1) & SNAI2) & Twist)
12	TGFb_secr	b_catenin_TCF4 & (!miR_200)

Table B.3: Boolean functions of the given NCD-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	**Target node name**	**Regulatory logic rule**
1	Fgf8_g	Fgf8_p & !Emx2_p & Sp8_p
2	Fgf8_p	Fgf8_g
3	Emx2_g	!Fgf8_p & !Pax6_p & Coup_tfi_p & !Sp8_p
4	Emx2_p	Emx2_g
5	Pax6_g	!Emx2_p & !Coup_tfi_p & Sp8_p
6	Pax6_p	Pax6_g
7	Coup_tfi_g	!Fgf8_p & !Sp8_p
8	Coup_tfi_p	Coup_tfi_g
9	Sp8_g	Fgf8_p & !Emx2_p
10	Sp8_p	Sp8_g

Table B.4: Boolean functions of the given EHD-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	Bmp2	(!canWnt & exogen_BMP2_II)
2	canWnt	exogen_canWnt_II
3	Dkk1	(Mesp1 | (canWnt & !exogen_BMP2_II))
4	Fgf8	(!Mesp1 & (Foxc1/2 | Tbx1))
5	Foxc1/2	(canWnt & exogen_canWnt_II)
6	GATAs	(Nkx2_5 | Mesp1 | Tbx5)
7	IsI1	(Tbx1 | Mesp1 | Fgf8 | (canWnt & exogen_canWnt_II))
8	Mesp1	(canWnt & !exogen_BMP2_II)
9	Nkx2_5	((IsI1 & GATAs) | Tbx1 | (Mesp1 & Dkk1) | (Bmp2 & GATAs) | Tbx5)
10	Tbx1	Foxc1/2
11	Tbx5	(!(Tbx1 | canWnt) & (Nkx2_5 | Tbx5 | Mesp1) & !(Dkk1 & !(Mesp1 | Tbx5)))
12	exogen_BMP2_I	1
13	exogen_BMP2_II	exogen_BMP2_I
14	exogen_canWnt_I	exogen_canWnt_I
15	exogen_canWnt_II	exogen_canWnt_I

Table B.5: Boolean functions of the given MD-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	GATA_2	GATA_2 & !(GATA_1 & FOG_1) & !PU_1
2	GATA_1	(GATA_1 | GATA_2 | Fli_1) & !PU_1
3	FOG_1	GATA_1
4	EKLF	GATA_1 & !Fli_1
5	Fli_1	GATA_1 & !EKLF
6	SCL	GATA_1 & !PU_1
7	C/EBP_alpha	C/EBP_alpha & !(GATA_1 & FOG_1 & SCL)
8	PU_1	(C/EBP_alpha | PU_1) & !(GATA_1 | GATA_2)
9	cJun	PU_1 & !Gfi_1
10	EgrNab	(PU_1 & cJun) & !Gfi_1
11	Gfi_1	C/EBP_alpha & !EgrNab

Table B.6: Boolean functions of the given THC-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	NFAT	TCR
2	IRAK	IL18R
3	STAT1	IFNbR | JAK1
4	IL10R	IL10
5	IL12R	IL12
6	IL18R	IL18 & !(STAT6)
7	IFNgR	IFNg
8	IFNg	(IRAK | NFAT | STAT4 | Tbet) & !(STAT3)
9	STAT4	IL12R & !(GATA3)
10	IL10	GATA3
11	STAT3	IL10R
12	IFNbR	IFNb
13	IL4	GATA3 & !(STAT1)
14	IL4R	IL4 & !(SOCS1)
15	GATA3	(GATA3 | STAT6) & !Tbet)
16	STAT6	IL4R
17	SOCS1	STAT1 | Tbet
18	JAK1	IFNgR & !(SOCS1)
19	Tbet	(STAT1 | Tbet) & !(GATA3)

Table B.7: Boolean functions of the given EGFR-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	EGFR	!BRCA1
2	ERK1/2	EGFR
3	PI3K	!PTEN & EGFR
4	AKT	PI3K
5	GSK3_beta	!AKT
6	MDM2	AKT & TP53
7	TP53	!MDM2 & (BRCA1 | !PARP1)
8	PTEN	TP53
9	PARP1	ERK1/2
10	BRCA1	!CYCD1
11	BCL_2	AKT
12	BAX	!BCL_2 & TP53
13	CYCD1	(!GSK3_beta & ERK1/2)| (!BRCA1 & PARP1)

Table B.8: Boolean functions of the given EMT-Senescence-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	Snai2	(Snai2 & NFkB) | !ESE2
2	ESE2	!Snai2 | (ESE2 & !NFkB)
3	p16	(E2F & p16) | (p16 & !Snai2) | (p53 & !Snai2) | (p53 & !TELasa)
4	E2F	!p53 & !Snai2 & !Rb
5	Cyclin	(!Snai2 & !p16) | (!ESE2 & !p16 & NFkB)
6	TELasa	!ESE2
7	NFkB	Snai2 | ESE2 | p16 | NFkB
8	Rb	p53 | (p16 & Cyclin)
9	p53	(!Snai2 & p16 & !TELasa) | (!Snai2 & !TELasa & !NFkB)

Table B.9: Boolean functions of the given FOS-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	AG	(!EMF1 & !AP2 & !TFL1) | (!EMF1 & !AP1 & LFY) | (!EMF1 & !AP2 & LFY) | (!EMF1 & !TFL1 & LFY & (AG & SEP)) | (!EMF1 & (LFY & WUS))
2	AP1	(!AG & !TFL1) | (FT & LFY & !AG) | (FT & !AG & !PI) | (LFY & !AG & !PI) | (FT & !AG & !AP3) | (LFY & !AG & !AP3)
3	AP2	!TFL1
4	AP3	(LFY & UFO) | (PI & SEP & AP3 & (AG | AP1))
5	EMF1	!LFY
6	FT	!EMF1
7	FUL	!AP1 & !TFL1
8	LFY	(!EMF1) | (!TFL1)
9	PI	(LFY & (AG | AP3)) | (PI & SEP & AP3 & (AG | AP1))
10	SEP	LFY
11	TFL1	!AP1 & (EMF1 & !LFY)
12	UFO	UFO
13	WUS	WUS & (!AG | !SEP)

Table B.10: Boolean functions of the given GSD-GRN Boolean model in BoolNet format. The column with header 'Target node name' contains the list of nodes whose regulation is captured by the corresponding row entry in the column 'Regulatory logic rule'. The symbols &, | and ! correspond to the logic operators AND, OR and NOT respectively.

Serial Number	Target node name	Regulatory logic rule
1	UGR	UGR & !(NR5A1 | WNT4)
2	CBX2	UGR & !(NR0B1 & WNT4 & CTNNB1)
3	GATA4	(UGR | WNT4 | NR5A1 | SRY)
4	WT1mKTS	(UGR | GATA4)
5	WT1pKTS	(UGR | GATA4) & !(WNT4 & CTNNB1)
6	NR5A1	(UGR | CBX2 | WT1mKTS | GATA4) & !(NR0B1 & WNT4)
7	NR0B1	(WT1mKTS | (WNT4 & CTNNB1)) & !(NR5A1 & SOX9)
8	SRY	((NR5A1 & WT1mKTS & CBX2) | (GATA4 & WT1pKTS & CBX2 & NR5A1) | (SOX9 | SRY)) & !(CTNNB1)
9	SOX9	((SOX9 & FGF9) | (SRY | PGD2) | (SRY & CBX2) | (GATA4 & NR5A1 & SRY)) & !(WNT4 | CTNNB1 | FOXL2)
10	FGF9	SOX9 & !WNT4
11	PGD2	SOX9
12	DMRT1	(SRY | SOX9) & !(FOXL2)
13	DHH	SOX9
14	DKK1	(SRY | SOX9)
15	AMH	((SOX9 & GATA4 & NR5A1) | (SOX9 & NR5A1 & GATA4 & WT1mKTS)) & !(NR0B1 & CTNNB1)
16	WNT4	(GATA4 | (CTNNB1 | RSPO1 | NR0B1)) & !(FGF9 | DKK1)
17	RSPO1	(WNT4 | CTNNB1) & !(DKK1)
18	FOXL2	(WNT4 & CTNNB1) & !(DMRT1 | SOX9)
19	CTNNB1	(WNT4 | RSPO1) & !(SRY | (SOX9 & AMH))

Table B.11: Biological fixed points recovered by the RSCN-GRN Boolean model. The 4 biological fixed points recovered by the model are the Quiescent center (QC), Vascular initials (VI), Cortex-Endodermis initials (CEI) and Columella epidermis initials (CEpI).

Cell types	Nodes								
	SCR	PLT	ARF	AUXIAA	AUXIN	SHR	JKD	MGP	WOX5
QC	1	1	1	0	1	1	1	0	1
VI	0	1	1	0	1	1	0	0	0
CEI	1	1	1	0	1	1	1	1	0
CEpI	0	1	1	0	1	0	0	0	0

Table B.12: Biological fixed points recovered by the EMT-GRN Boolean model. The 3 biological fixed points recovered by the model are the cell types: Epithelial, Mesenchymal and Hybrid.

Cell types	Nodes											
	SNAI1	LEF1	Twist	SNAI2	ZEB1	N_bcatenin_H	ZEB1_H	b_catenin_TCF4	miR_34	miR_200	Ecadherin_mRNA	TGFb_secr
Epithelial	0	0	0	0	0	0	0	0	1	1	1	0
Hybrid	1	0	1	1	1	0	0	0	0	1	1	0
Mesenchymal	1	1	1	1	1	1	1	1	0	0	0	1

Table B.13: Biological fixed points recovered by the NCD-GRN Boolean model. The 2 biological fixed points recovered by the model are the Anterior and Posterior compartments of the developing neocortex.

Cell types	Nodes									
	Fgf8_g	Fgf8_p	Emx2_g	Emx2_p	Pax6_g	Pax6_p	Coup_tfi_g	Coup_tfi_p	Sp8_g	Sp8_p
Anterior	1	1	0	0	1	1	0	0	1	1
Posterior	0	0	1	1	0	0	1	1	0	0

Table B.14: Biological fixed points recovered by the EHD-GRN Boolean model. The 2 biological fixed points recovered by the model are the cell types in the First Heart Field (FHF) and Second Heart Field (SHF).

Cell types	Nodes														
	Bmp2	canWnt	Dkk1	Fgf8	Foxc1/2	GATAs	Isl1	Mesp1	Nkx2_5	Tbx1	Tbx5	exogen_BMP2_I	exogen_BMP2_II	exogen_canWnt_I	exogen_canWnt_II
FHF	1	0	0	0	0	1	0	0	1	0	1	1	1	0	0
SHF	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1

Table B.15: Biological fixed points recovered by the MD-GRN Boolean model. The 4 biological fixed points recovered by the model correspond to the cell types: erythrocytes, megakaryocytes, monocytes and granulocytes.

Cell types	Nodes										
	GATA_2	GATA_1	FOG_1	EKLF	Fli_1	SCL	C/EBP_alpha	PU_1	cJun	EgrNab	Gfi_1
Erythrocyte	0	1	1	1	0	1	0	0	0	0	0
Megakaryocyte	0	1	1	0	1	1	0	0	0	0	0
Monocyte	0	0	0	0	0	0	1	1	1	1	0
Granulocyte	0	0	0	0	0	0	1	1	0	0	1

Table B.16: Biological fixed points recovered by the THC-GRN Boolean model. The 3 biological fixed points recovered by the model correspond to different types of T-helper cells, namely, Th0, Th1 and Th2.

Cell types	Nodes																						
	NFAT	IRAK	STAT1	IL10R	IL12R	IL18R	IFNgR	IFNg	STAT4	IL10	STAT3	IFNbR	IL4	IL4R	GATA3	STAT6	SOCS1	JAK1	Tbet	IFNb	TCR	IL18	IL12
Th0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Th1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Th2	0	0	0	1	0	0	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0

Table B.17: Biological fixed points recovered by the EGFR-GRN Boolean model. The 2 biological fixed points recovered by the model correspond to the cell states: cell division and apoptosis.

Cell types	Nodes												
	EGFR	ERK1/2	PI3K	AKT	GSK3_beta	MDM2	TP53	PTEN	PARP1	BRCA1	BCL_2	BAX	CYCD1
Division	1	1	1	1	0	0	0	0	1	0	1	0	1
Apoptosis	0	0	0	0	1	0	1	1	0	1	0	1	0

Table B.18: Biological fixed points recovered by the EMT-Senescence-GRN Boolean model. The 3 biological fixed points recovered by the model correspond to epithelial cells, senescent epithelial cells and stem-like mesenchymal cells.

Cell types	Nodes								
	Snai2	ESE2	p16	E2F	Cyclin	TELasa	NFkB	Rb	p53
Epithelial	0	1	0	1	1	0	1	0	0
Mesenchymal	1	0	0	0	1	1	1	0	0
Senescent	0	1	1	0	0	0	1	1	1

Table B.19: Biological fixed points recovered by the FOS-GRN Boolean model. The 10 biological fixed points recovered by the model correspond to inflorescence (4 cell types), sepal (1 cell type), petal (2 cell types), stamen (2 cell types) and carpel (1 cell type).

Cell types	Nodes												
	AG	AP1	AP2	AP3	EMF1	FT	FUL	LFY	PI	SEP	TFL1	UFO	WUS
INF1	0	0	0	0	1	0	0	0	0	0	1	0	0
INF2	0	0	0	0	1	0	0	0	0	0	1	1	0
INF3	0	0	0	0	1	0	0	0	0	0	1	0	1
INF4	0	0	0	0	1	0	0	0	0	0	1	1	1
SEP	0	1	1	0	0	1	0	1	0	1	0	0	0
PET1	0	1	1	1	0	1	0	1	1	1	0	1	0
PET2	0	1	1	1	0	1	0	1	1	1	0	0	0
STM1	1	0	1	1	0	1	1	1	1	1	0	1	0
STM2	1	0	1	1	0	1	1	1	1	1	0	0	0
CAR	1	0	1	0	0	1	1	1	1	1	0	0	0

Table B.20: Biological fixed points recovered by the GSD-GRN Boolean model. The 2 biological fixed points recovered by the model correspond to the Sertoli progenitor cells (SPC) and granulosa progenitor cells (GPC).

Cell types	Nodes																		
	UGR	CBX2	GATA4	WT1mKTS	WT1pKTS	NR5A1	NR0B1	SRY	SOX9	FGF9	PGD2	DMRT1	DHH	DKK1	AMH	WNT4	RSPO1	FOXL2	CTNNB1
Toward_Sertoli	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0
Toward_granulosa	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1

Table B.21: Type of the BF at each node of the given RSCN-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	SCR	Yes	Yes	Yes	Yes
2	PLT	Yes	Yes	Yes	Yes
3	ARF	Yes	Yes	Yes	Yes
4	AUXIAA	Yes	Yes	Yes	Yes
5	AUX	No	Yes	No	No
6	SHR	Yes	Yes	Yes	Yes
7	JKD	Yes	Yes	Yes	Yes
8	MGP	Yes	Yes	Yes	Yes
9	WOX5	Yes	Yes	Yes	Yes

Table B.22: Type of the BF at each node of the given EMT-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	SNAI1	Yes	Yes	No	No
2	LEF1	Yes	Yes	Yes	Yes
3	Twist	Yes	Yes	Yes	Yes
4	SNAI2	Yes	Yes	Yes	Yes
5	ZEB1	Yes	Yes	Yes	Yes
6	N_bcatenin_H	Yes	Yes	Yes	Yes
7	ZEB1_H	Yes	Yes	Yes	Yes
8	b_catenin_TCF4	Yes	Yes	Yes	Yes
9	miR_34	Yes	Yes	Yes	Yes
10	miR_200	Yes	Yes	No	Yes
11	Ecadherin_mRNA	Yes	Yes	Yes	Yes
12	TGFb_secr	Yes	Yes	Yes	Yes

Table B.23: Type of the BF at each node of the given NCD-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	Fgf8_g	Yes	Yes	Yes	Yes
2	Fgf8_p	Yes	Yes	Yes	Yes
3	Emx2_g	Yes	Yes	Yes	Yes
4	Emx2_p	Yes	Yes	Yes	Yes
5	Pax6_g	Yes	Yes	Yes	Yes
6	Pax6_p	Yes	Yes	Yes	Yes
7	Coup_tfi_g	Yes	Yes	Yes	Yes
8	Coup_tfi_p	Yes	Yes	Yes	Yes
9	Sp8_g	Yes	Yes	Yes	Yes
10	Sp8_p	Yes	Yes	Yes	Yes

Table B.24: Type of the BF at each node of the given EHD-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	Bmp2	Yes	Yes	Yes	Yes
2	canWnt	Yes	Yes	Yes	Yes
3	Dkk1	Yes	Yes	Yes	Yes
4	Fgf8	Yes	Yes	Yes	Yes
5	Foxc1/2	Yes	Yes	Yes	Yes
6	GATAs	Yes	Yes	Yes	Yes
7	IsI1	Yes	Yes	Yes	Yes
8	Mesp1	Yes	Yes	Yes	Yes
9	Nkx2_5	Yes	Yes	No	Yes
10	Tbx1	Yes	Yes	Yes	Yes
11	Tbx5	Yes	Yes	Yes	Yes
12	exogen_BMP2_I	-	-	-	-
13	exogen_BMP2_II	Yes	Yes	Yes	Yes
14	exogen_canWnt_I	Yes	Yes	Yes	Yes
15	exogen_canWnt_II	Yes	Yes	Yes	Yes

Table B.25: Type of the BF at each node of the given MD-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	GATA_2	Yes	Yes	Yes	Yes
2	GATA_1	Yes	Yes	Yes	Yes
3	FOG_1	Yes	Yes	Yes	Yes
4	EKLF	Yes	Yes	Yes	Yes
5	Fli_1	Yes	Yes	Yes	Yes
6	SCL	Yes	Yes	Yes	Yes
7	C/EBP_alpha	Yes	Yes	Yes	Yes
8	PU_1	Yes	Yes	Yes	Yes
9	cJun	Yes	Yes	Yes	Yes
10	EgrNab	Yes	Yes	Yes	Yes
11	Gfi_1	Yes	Yes	Yes	Yes

Table B.26: Type of the BF at each node of the given THC-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	NFAT	Yes	Yes	Yes	Yes
2	IRAK	Yes	Yes	Yes	Yes
3	STAT1	Yes	Yes	Yes	Yes
4	IL10R	Yes	Yes	Yes	Yes
5	IL12R	Yes	Yes	Yes	Yes
6	IL18R	Yes	Yes	Yes	Yes
7	IFNgR	Yes	Yes	Yes	Yes
8	IFNg	Yes	Yes	Yes	Yes
9	STAT4	Yes	Yes	Yes	Yes
10	IL10	Yes	Yes	Yes	Yes
11	STAT3	Yes	Yes	Yes	Yes
12	IFNbR	Yes	Yes	Yes	Yes
13	IL4	Yes	Yes	Yes	Yes
14	IL4R	Yes	Yes	Yes	Yes
15	GATA3	Yes	Yes	Yes	Yes
16	STAT6	Yes	Yes	Yes	Yes
17	SOCS1	Yes	Yes	Yes	Yes
18	JAK1	Yes	Yes	Yes	Yes
19	Tbet	Yes	Yes	Yes	Yes

Table B.27: Type of the BF at each node of the given EGFR-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	EGFR	Yes	Yes	Yes	Yes
2	ERK1/2	Yes	Yes	Yes	Yes
3	PI3K	Yes	Yes	Yes	Yes
4	AKT	Yes	Yes	Yes	Yes
5	GSK3_beta	Yes	Yes	Yes	Yes
6	MDM2	Yes	Yes	Yes	Yes
7	TP53	Yes	Yes	Yes	Yes
8	PTEN	Yes	Yes	Yes	Yes
9	PARP1	Yes	Yes	Yes	Yes
10	BRCA1	Yes	Yes	Yes	Yes
11	BCL_2	Yes	Yes	Yes	Yes
12	BAX	Yes	Yes	Yes	Yes
13	CYCD1	Yes	Yes	No	Yes

Table B.28: Type of the BF at each node of the given EMT-Senescence-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	Snai2	Yes	Yes	Yes	Yes
2	ESE2	Yes	Yes	Yes	Yes
3	p16	Yes	Yes	No	No
4	E2F	Yes	Yes	Yes	Yes
5	Cyclin	Yes	Yes	Yes	Yes
6	TELasa	Yes	Yes	Yes	Yes
7	NFkB	Yes	Yes	Yes	Yes
8	Rb	Yes	Yes	Yes	Yes
9	p53	Yes	Yes	Yes	Yes

Table B.29: Type of the BF at each node of the given FOS-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	AG	Yes	Yes	No	No
2	AP1	Yes	Yes	No	No
3	AP2	Yes	Yes	Yes	Yes
4	AP3	Yes	Yes	No	Yes
5	EMF1	Yes	Yes	Yes	Yes
6	FT	Yes	Yes	Yes	Yes
7	FUL	Yes	Yes	Yes	Yes
8	LFY	Yes	Yes	Yes	Yes
9	PI	Yes	Yes	No	No
10	SEP	Yes	Yes	Yes	Yes
11	TFL1	Yes	Yes	Yes	Yes
12	UFO	Yes	Yes	Yes	Yes
13	WUS	Yes	Yes	Yes	Yes

Table B.30: Type of the BF at each node of the given GSD-GRN Boolean model. The 'Node' column contains all the nodes of the network. An entry in the columns corresponding to BF types, namely EF, UF, NCF and RoF, is 'Yes' if the BF in the original model belongs to that function type and is 'No' if it does not belong to that function type.

Serial Number	Node	EF	UF	NCF	RoF
1	UGR	Yes	Yes	Yes	Yes
2	CBX2	Yes	Yes	Yes	Yes
3	GATA4	Yes	Yes	Yes	Yes
4	WT1mKTS	Yes	Yes	Yes	Yes
5	WT1pKTS	Yes	Yes	No	Yes
6	NR5A1	Yes	Yes	No	Yes
7	NR0B1	Yes	Yes	No	Yes
8	SRY	Yes	Yes	Yes	Yes
9	SOX9	No	Yes	No	No
10	FGF9	Yes	Yes	Yes	Yes
11	PGD2	Yes	Yes	Yes	Yes
12	DMRT1	Yes	Yes	Yes	Yes
13	DHH	Yes	Yes	Yes	Yes
14	DKK1	Yes	Yes	Yes	Yes
15	AMH	No	Yes	No	No
16	WNT4	Yes	Yes	Yes	Yes
17	RSPO1	Yes	Yes	Yes	Yes
18	FOXL2	Yes	Yes	Yes	Yes
19	CTNNB1	Yes	Yes	No	Yes

Table B.31: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the RSCN-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	SCR	4	4010	50	20	17
2	PLT	1	1	1	1	1
3	ARF	1	1	1	1	1
4	AUXIAA	1	1	1	1	1
5	AUXIN	1	1	1	1	1
6	SHR	1	1	1	1	1
7	JKD	2	1	1	1	1
8	MGP	3	14	1	1	1
9	WOX5	5	-	732	94	75
Total			-	36600	1880	1275
Sampled			True	False	False	False

Table B.32: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EMT-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	SNAI1	3	26	7	6	6
2	LEF1	3	26	7	6	6
3	Twist	2	2	2	2	2
4	SNAI2	4	8074	57	26	23
5	ZEB1	3	29	2	2	2
6	N_bcatenin_H	2	1	1	1	1
7	ZEB1_H	5	-	5119	310	213
8	b_catenin_TCF4	3	29	2	2	2
9	miR_34	3	26	7	6	6
10	miR_200	5	-	1775	162	119
11	Ecadherin_mRNA	5	-	114	52	46
12	TGFb_secr	2	2	2	2	2
Total			-	324024087794400	234653552640	92679987456
Sampled			True	True	True	True

Table B.33: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the NCD-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	Fgf8_g	3	55	9	8	8
2	Fgf8_p	1	1	1	1	1
3	Emx2_g	4	16148	114	52	46
4	Emx2_p	1	1	1	1	1
5	Pax6_g	3	55	9	8	8
6	Pax6_p	1	1	1	1	1
7	Coup_tfi_g	2	2	2	2	2
8	Coup_tfi_p	1	1	1	1	1
9	Sp8_g	2	2	2	2	2
10	Sp8_p	1	1	1	1	1
Total			195390800	36936	13312	11776
Sampled			True	False	False	False

Table B.34: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EHD-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	Bmp2	2	3	1	1	1
2	canWnt	1	1	1	1	1
3	Dkk1	3	52	7	6	6
4	Fgf8	3	54	7	6	6
5	Foxc1/2	2	2	2	2	2
6	GATAs	3	52	2	2	2
7	IsI1	5	-	6780	420	286
8	Mesp1	2	2	1	1	1
9	Nkx2_5	7	-	-	15976	6363
10	Tbx1	1	1	1	1	1
11	Tbx5	6	-	7190326	3648	1735
12	exogen_BMP2_I	1	1	1	1	1
13	exogen_BMP2_II	1	1	1	1	1
14	exogen_canWnt_I	1	1	1	1	1
15	exogen_canWnt_II	1	1	1	1	1
Total			-	-	3524801495040	454663329120
Sampled			True	True	True	True

Table B.35: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the MD-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	GATA_2	4	16148	55	24	21
2	GATA_1	4	8046	57	26	23
3	FOG_1	1	1	1	1	1
4	EKLF	2	2	1	1	1
5	Fli_1	2	2	1	1	1
6	SCL	2	2	2	2	2
7	C/EBP_alpha	4	16148	114	52	46
8	PU_1	4	16149	105	44	38
9	cJun	2	2	1	1	1
10	EgrNab	3	27	6	5	5
11	Gfi_1	2	2	1	1	1
Total			29273650720346317824	450311400	14277120	8442840
Sampled			True	True	True	True

Table B.36: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the THC-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	NFAT	1	1	1	1	1
2	IRAK	1	1	1	1	1
3	STAT1	2	5	2	2	2
4	IL10R	1	1	1	1	1
5	IL12R	1	1	1	1	1
6	IL18R	2	2	1	1	1
7	IFNgR	1	1	1	1	1
8	IFNg	5	-	1661	110	73
9	STAT4	2	2	1	1	1
10	IL10	1	1	1	1	1
11	STAT3	1	1	1	1	1
12	IFNbR	1	1	1	1	1
13	IL4	2	3	1	1	1
14	IL4R	2	1	1	1	1
15	GATA3	3	26	7	6	6
16	STAT6	1	1	1	1	1
17	SOCS1	2	3	1	1	1
18	JAK1	2	3	1	1	1
19	Tbet	3	27	5	4	4
20	IFNb	1	1	1	1	1
21	TCR	1	1	1	1	1
22	IL18	1	1	1	1	1
23	IL12	1	1	1	1	1
Total			-	116270	5280	3504
Sampled			True	False	False	False

Table B.37: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EGFR-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	EGFR	1	1	1	1	1
2	ERK1/2	1	1	1	1	1
3	PI3K	2	2	2	2	2
4	AKT	1	1	1	1	1
5	GSK3_beta	1	1	1	1	1
6	MDM2	2	3	1	1	1
7	TP53	3	54	7	6	6
8	PTEN	1	1	1	1	1
9	PARP1	1	1	1	1	1
10	BRCA1	1	1	1	1	1
11	BCL_2	1	1	1	1	1
12	BAX	2	2	2	2	2
13	CYCD1	4	16148	114	52	46
Total			10463904	3192	1248	1104
Sampled			True	False	False	False

Table B.38: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the EMT-Senescence-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	Snai2	3	54	7	6	6
2	ESE2	3	54	7	6	6
3	p16	5	-	1718	136	96
4	E2F	3	28	2	2	2
5	Cyclin	4	8103	43	14	11
6	TELasa	1	1	1	1	1
7	NFkB	4	8048	36	18	17
8	Rb	3	55	5	4	4
9	p53	4	8103	48	18	15
Total			-	62550593280	177666048	77552640
Sampled			True	True	True	True

Table B.39: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the FOS-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	AG	8	-	-	324888	66042
2	AP1	6	-	274247	380	178
3	AP2	1	1	1	1	1
4	AP3	7	-	-	9018	2920
5	EMF1	1	1	1	1	1
6	FT	1	1	1	1	1
7	FUL	2	2	1	1	1
8	LFY	2	2	2	2	2
9	PI	6	-	3500117	1436	627
10	SEP	1	1	1	1	1
11	TFL1	3	26	7	6	6
12	UFO	1	1	1	1	1
13	WUS	3	13	2	2	2
Total			-	-	38370121979258880	516537496316160
Sampled			True	True	True	True

Table B.40: Nodewise enumeration of the number of BFs that satisfy various biologically meaningful constraints for the GSD-GRN model. The 'Node' column contains all nodes of the network. The 'k' column is the number of effective inputs to each node. Each of the columns EF, EUF, RoF and NCF provides the numbers of the respective type of BF that satisfy the fixed point constraints for each of the nodes. Additionally, the BFs in EUF, RoF and NCF also satisfy the signs of the inputs in the network architecture. The row labeled 'Total' gives the total number of Boolean models allowed when imposing a given type of BF at each node. Here and in the previous rows, '-' denotes that it was computationally infeasible to obtain the exact values. Finally, the row labeled 'Sampled' specifies whether we had to resort to sampling of these models. When set to False, we considered all possible models by exhaustive enumeration.

Serial Number	**Node**	k	**EF**	**EUF**	**RoF**	**NCF**
1	UGR	3	55	6	5	5
2	CBX2	4	16148	9	8	8
3	GATA4	4	16148	7	6	6
4	WT1mKTS	2	5	1	1	1
5	WT1pKTS	4	16146	96	36	30
6	NR5A1	6	-	6609378	2736	1260
7	NR0B1	5	-	6780	420	286
8	SRY	8	-	-	1032456	239279
9	SOX9	7	-	-	78416	29024
10	FGF9	2	2	2	2	2
11	PGD2	1	1	1	1	1
12	DMRT1	3	55	9	8	8
13	DHH	1	1	1	1	1
14	DKK1	2	2	2	2	2
15	AMH	5	-	6780	420	286
16	WNT4	6	-	7778168	5032	2542
17	RSPO1	3	55	9	8	8
18	FOXL2	4	16148	114	52	46
19	CTNNB1	5	-	6894	472	332
Total			-	-	1.067×10^{34}	5.121×10^{31}
Sampled			True	True	True	True

Table B.41: Performance evaluation for using $1 - \overline{Z}_{mid}$ as a proxy for G-density, based on average AUROC and AUPRC scores. Shown are the AUROC (Area Under the Receiver Operating Characteristic Curve) and AUPRC (Area Under the Precision-Recall Curve) scores for $1 - \overline{Z}_{mid}$ when averaging over ground truths. A ground truth is obtained by thresholding the G-density of the models in the EF ensembles, models generated using network structures from 10 published Boolean GRNs. In detail, we successively consider the percentile thresholds $0.1\%, 0.2\%, \ldots, 99.9\%$. For each threshold, we identify the models having G-density percentiles above that threshold and label them as 'true' while the others are labelled as 'false'. The AUROC and AUPRC scores are then computed for the $1 - \overline{Z}_{mid}$ values for that ground truth. Finally, these scores are average over all thresholds, with the results being given in the present table. (Note that the computations were performed for $1 - \overline{Z}_{mid}$ rather than for $\overline{Z}_{mid}$ since $\overline{Z}_{mid}$ and G-density are negatively correlated.)

Names of Boolean GRN models	Average AUROC Score	Average AUPRC Score
RSCN-GRN	0.795	0.737
EMT-GRN	0.804	0.744
NCD-GRN	0.747	0.675
EHD-GRN	0.747	0.699
MD-GRN	0.857	0.762
THC-GRN	0.722	0.641
EGFR-GRN	0.670	0.653
EMT-Senescence-GRN	0.732	0.661
FOS-GRN	0.886	0.818
GSD-GRN	0.743	0.666

B.5 Supplementary figures for chapter 5

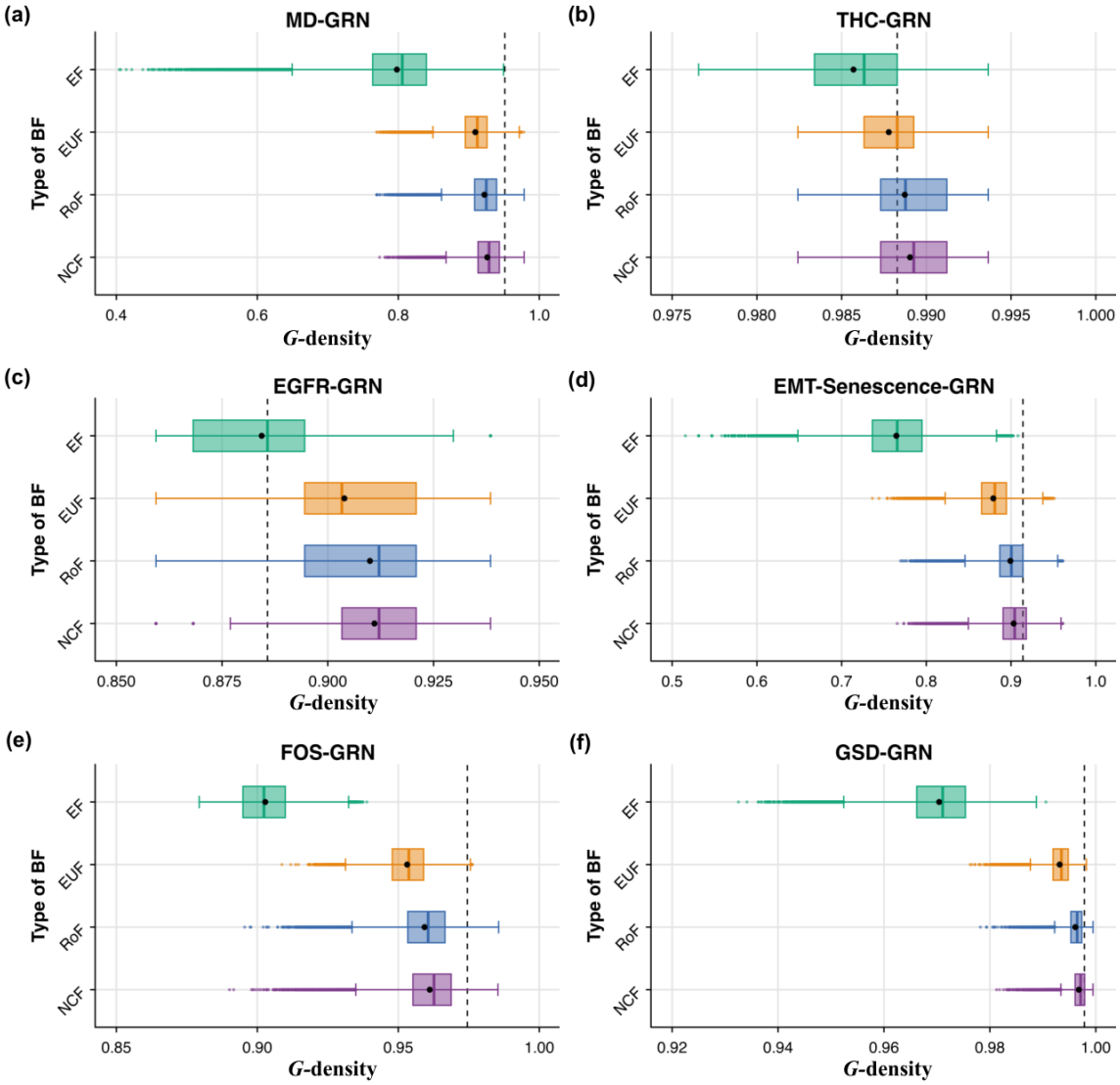


Figure B.1: Distribution of *G*-density values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to G-density and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of G-density in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have higher G-density compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the G-density of the Boolean model provided by the modelers in the published article.

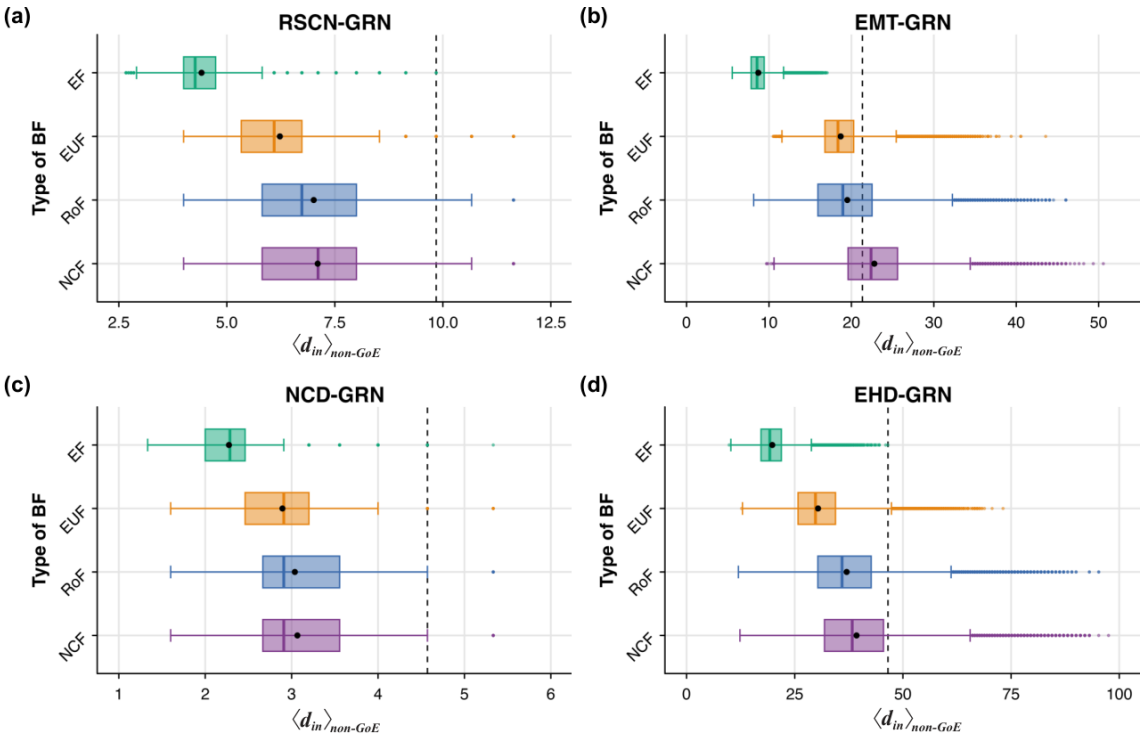


Figure B.2: Distribution of average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) for different ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\langle d_{in} \rangle_{non-GoE}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\langle d_{in} \rangle_{non-GoE}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have higher $\langle d_{in} \rangle_{non-GoE}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\langle d_{in} \rangle_{non-GoE}$ of the Boolean model provided by the modelers in the published article.

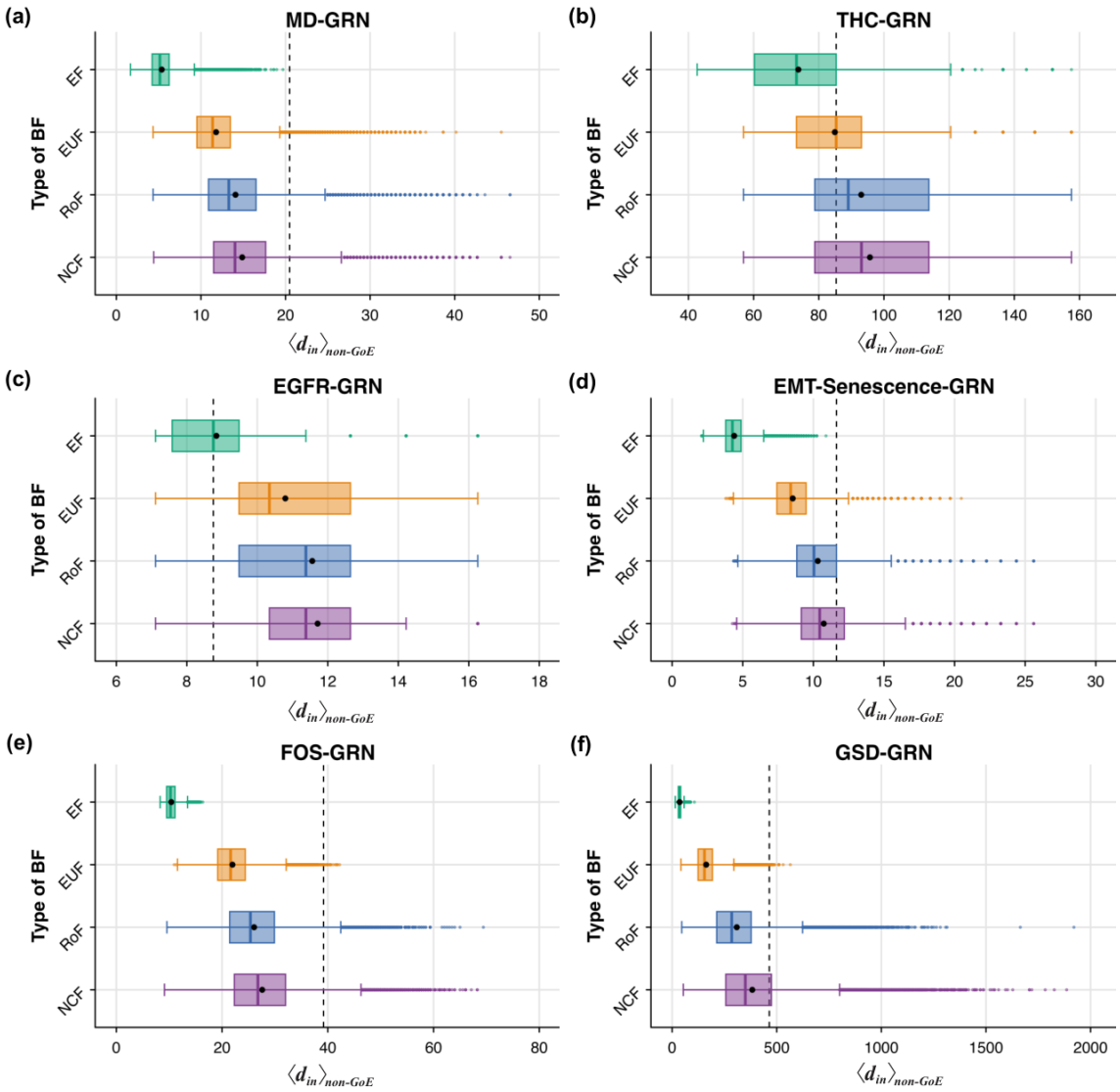


Figure B.3: Distribution of average in-degree of *non-GoE* states ($\langle d_{in} \rangle_{non-GoE}$) for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\langle d_{in} \rangle_{non-GoE}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\langle d_{in} \rangle_{non-GoE}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have higher $\langle d_{in} \rangle_{non-GoE}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\langle d_{in} \rangle_{non-GoE}$ of the Boolean model considered by the modelers in the published article.

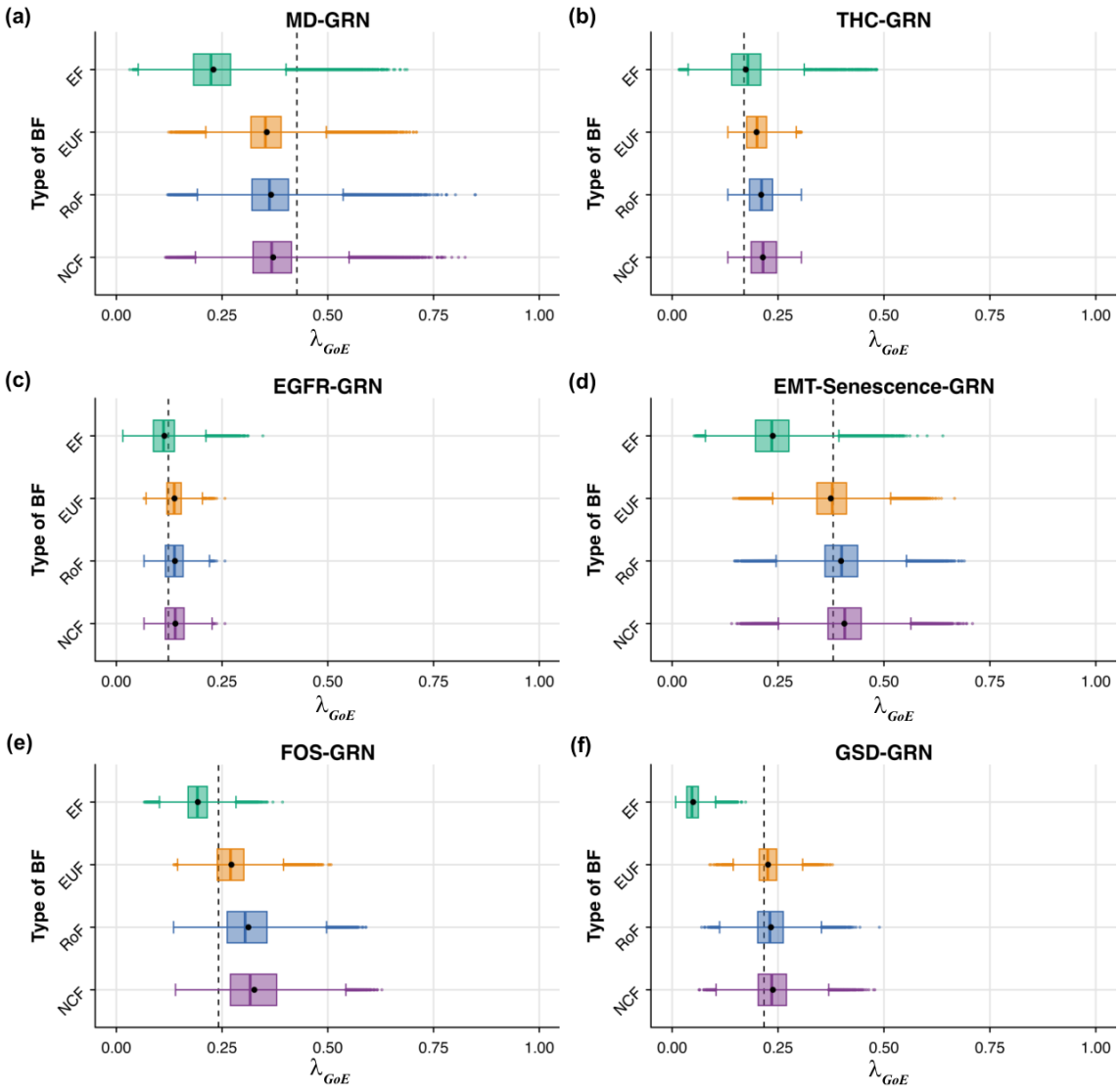


Figure B.4: Distribution of average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to λ_{GoE} and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of λ_{GoE} in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have higher convergence rates compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the λ_{GoE} of the Boolean model provided by the modelers in the published article.

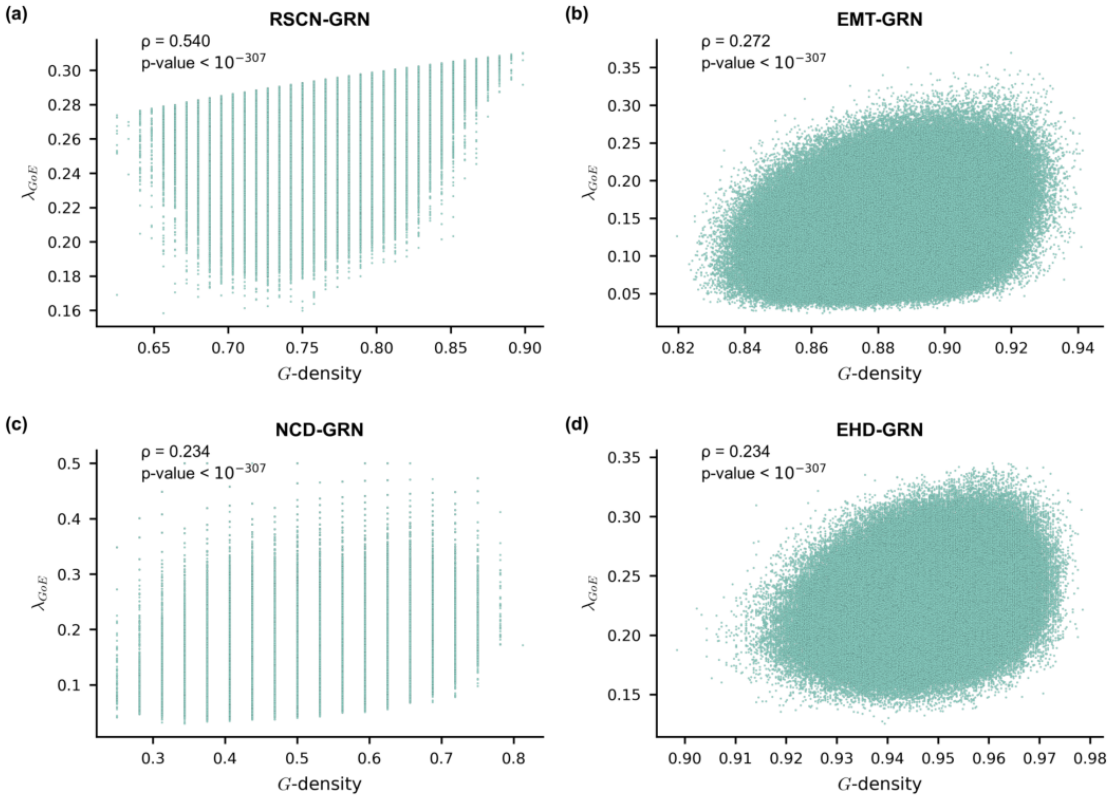


Figure B.5: Scatter plot between the *G*-density and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) for the EF ensemble generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The x and y axis of each sub-figure correspond to the G-density and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. G-density and λ_{GoE} show a moderate positive correlation for the RSCN-GRN network and a weak positive correlation otherwise.

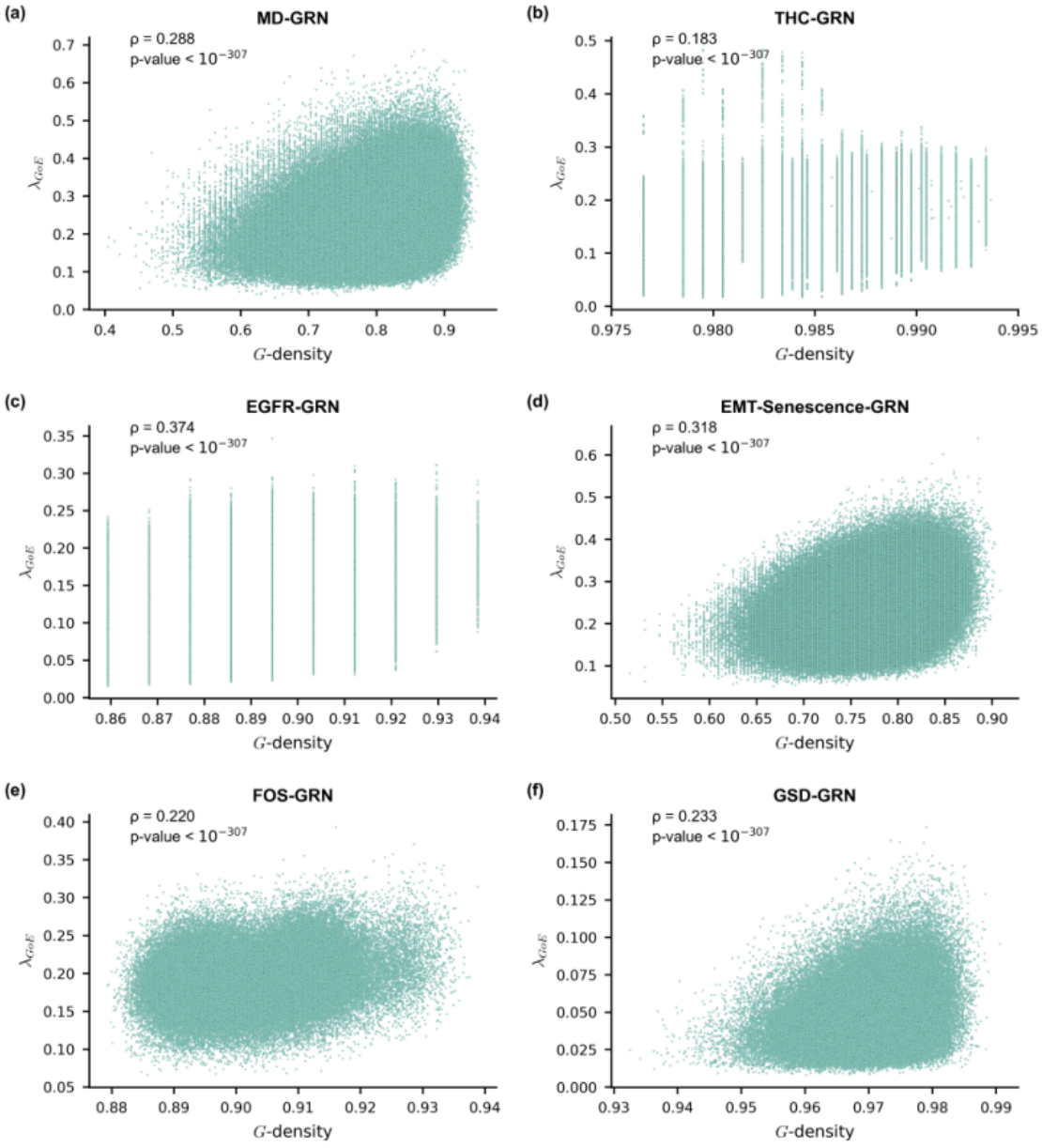


Figure B.6: Scatter plot between the *G*-density and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) for the EF ensemble generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axis of each sub-figure correspond to the G-density and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. G-density and λ_{GoE} show a weak positive correlation for all of the networks except for the THC-GRN network which shows a very weak positive correlation.

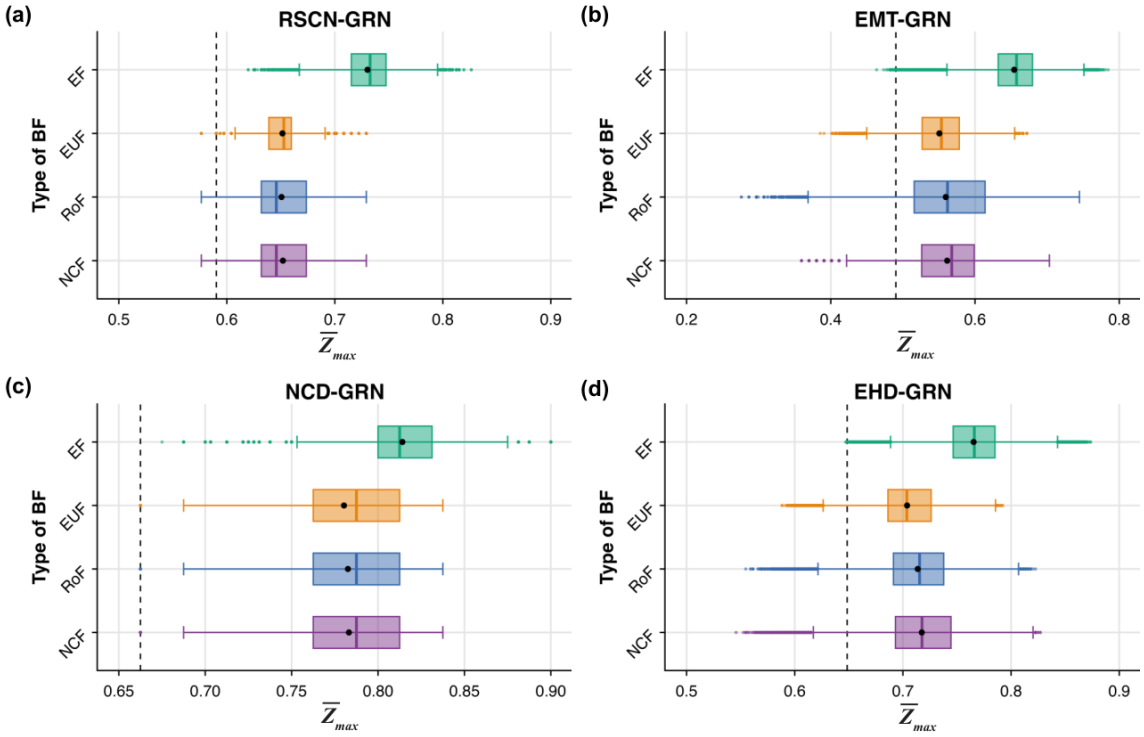


Figure B.7: Distribution of $\overline{Z}_{max}$ values for different ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{max}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{max}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{max}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{max}$ of the Boolean model provided by the modelers in the published article.

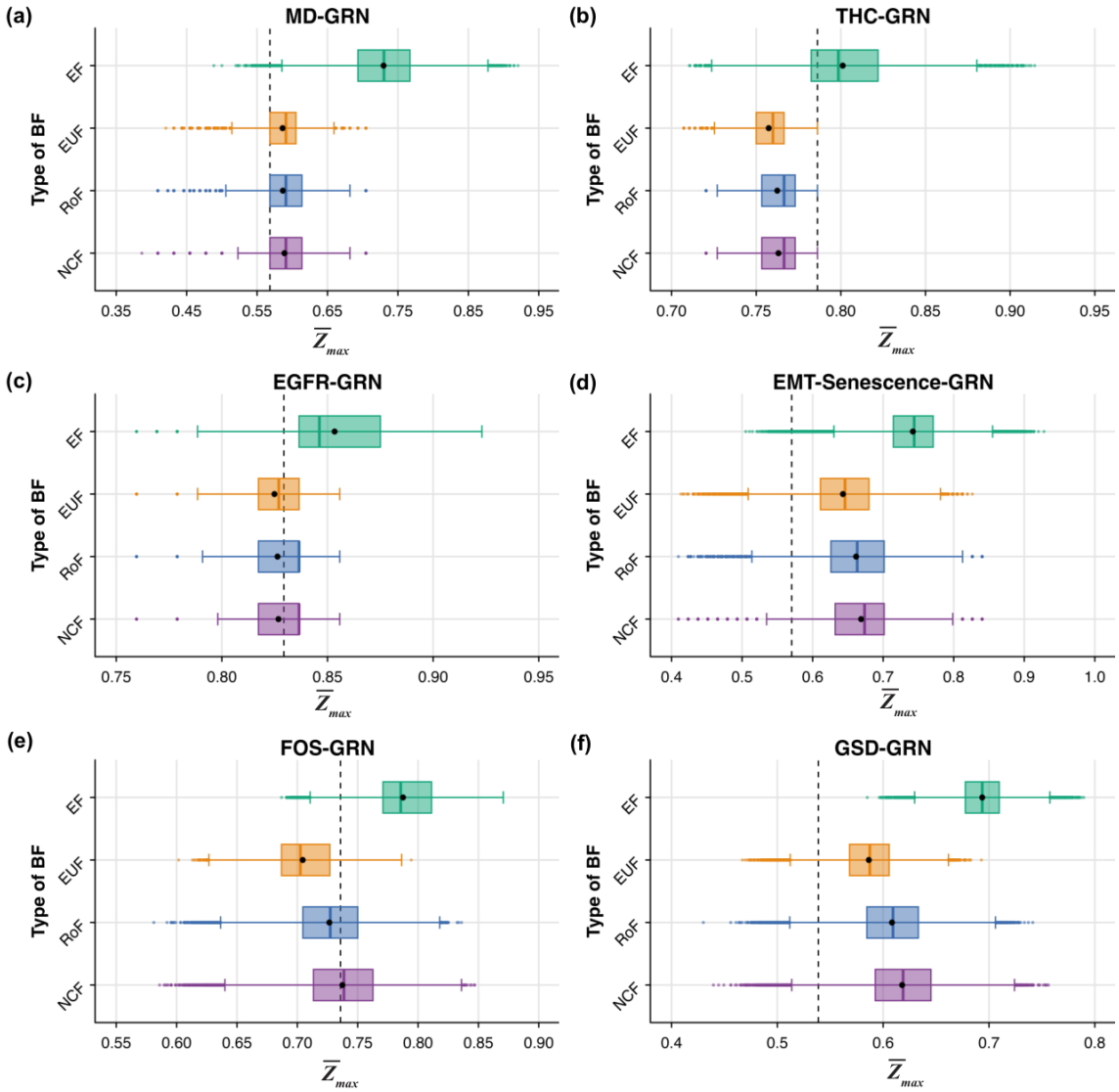


Figure B.8: Distribution of $\overline{Z}_{max}$ values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{max}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{max}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{max}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{max}$ of the Boolean model provided by the modelers in the published article.

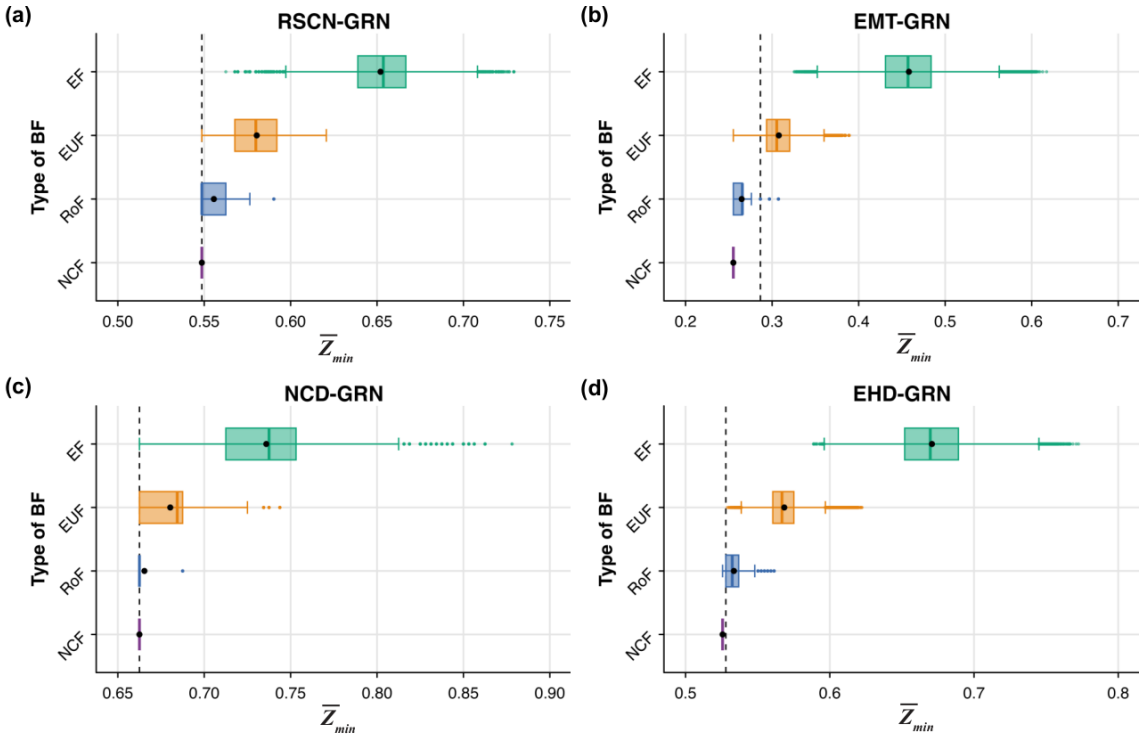


Figure B.9: Distribution of $\overline{Z}_{min}$ values for different ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{min}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{min}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{min}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{min}$ of the Boolean model provided by the modelers in the published article.

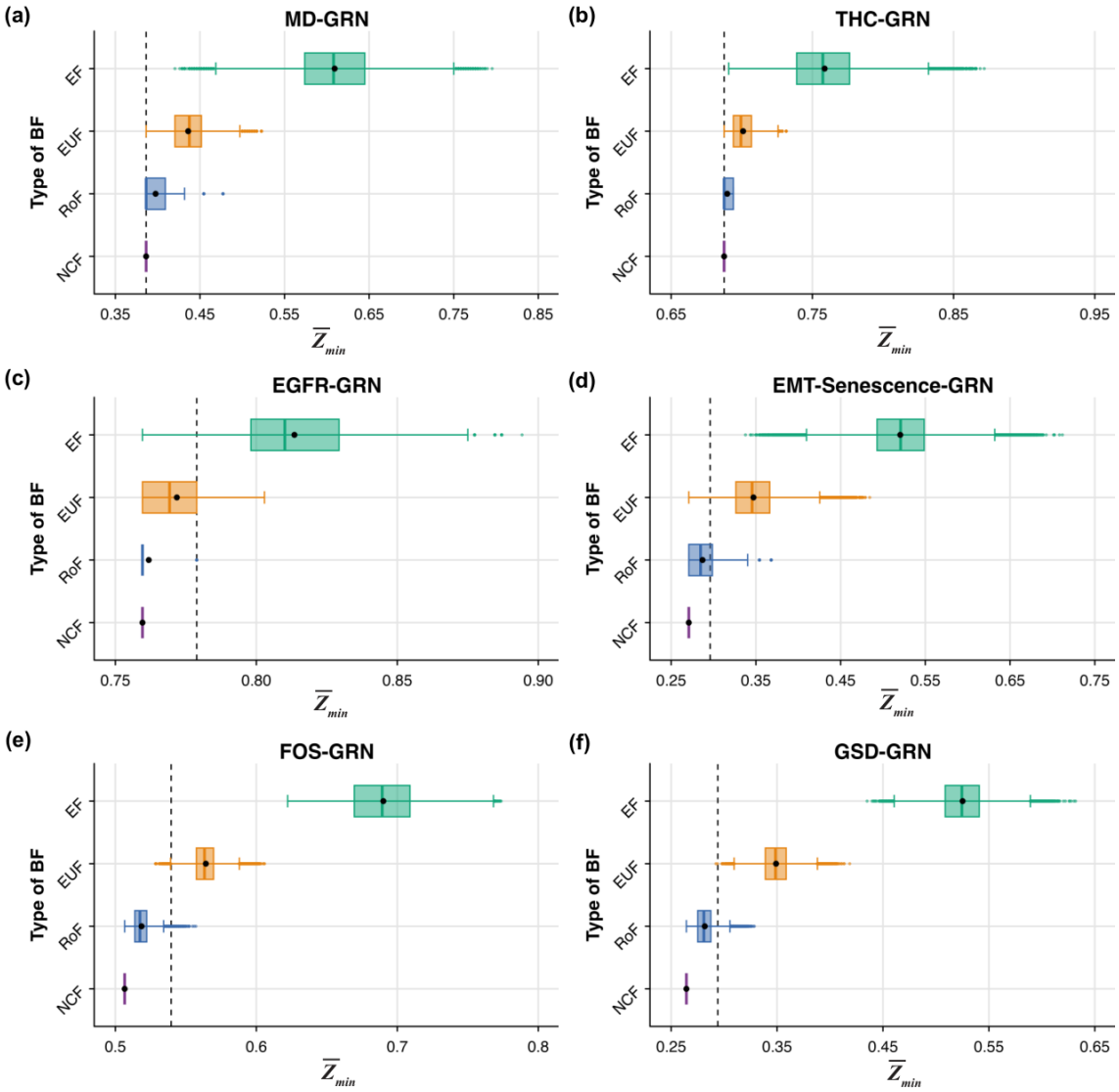


Figure B.10: Distribution of $\overline{Z}_{min}$ values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{min}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{min}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{min}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{min}$ of the Boolean model provided by the modelers in the published article.

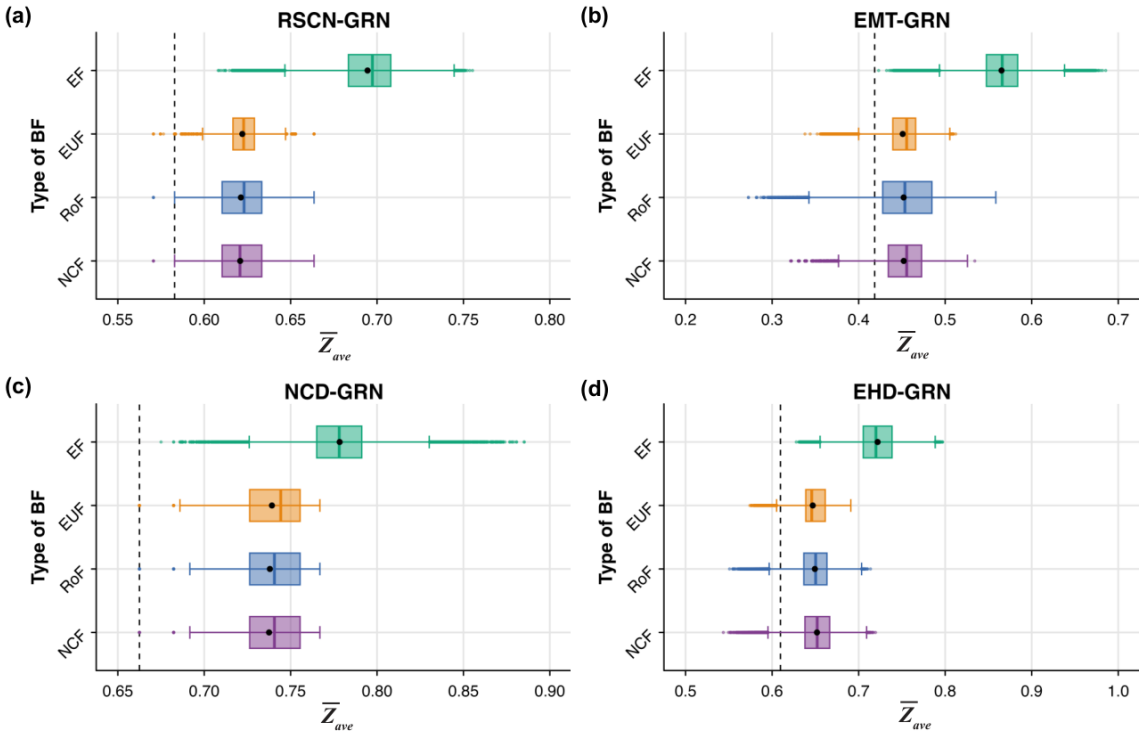


Figure B.11: Distribution of $\overline{Z}_{ave}$ values for different ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{ave}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{ave}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{ave}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{ave}$ of the Boolean model provided by the modelers in the published article.

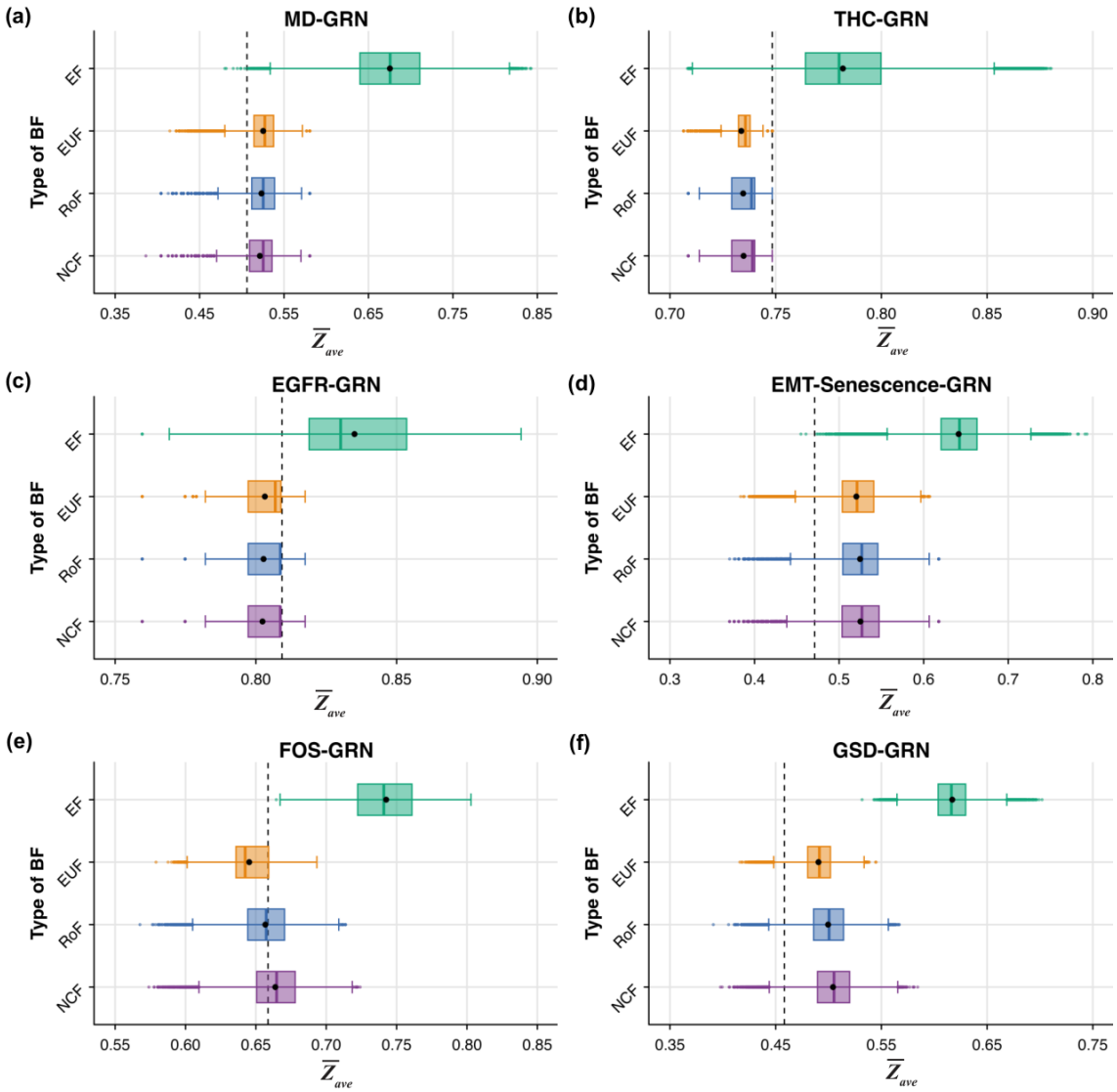


Figure B.12: Distribution of $\overline{Z}_{ave}$ values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{ave}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{ave}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{ave}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{ave}$ of the Boolean model provided by the modelers in the published article.

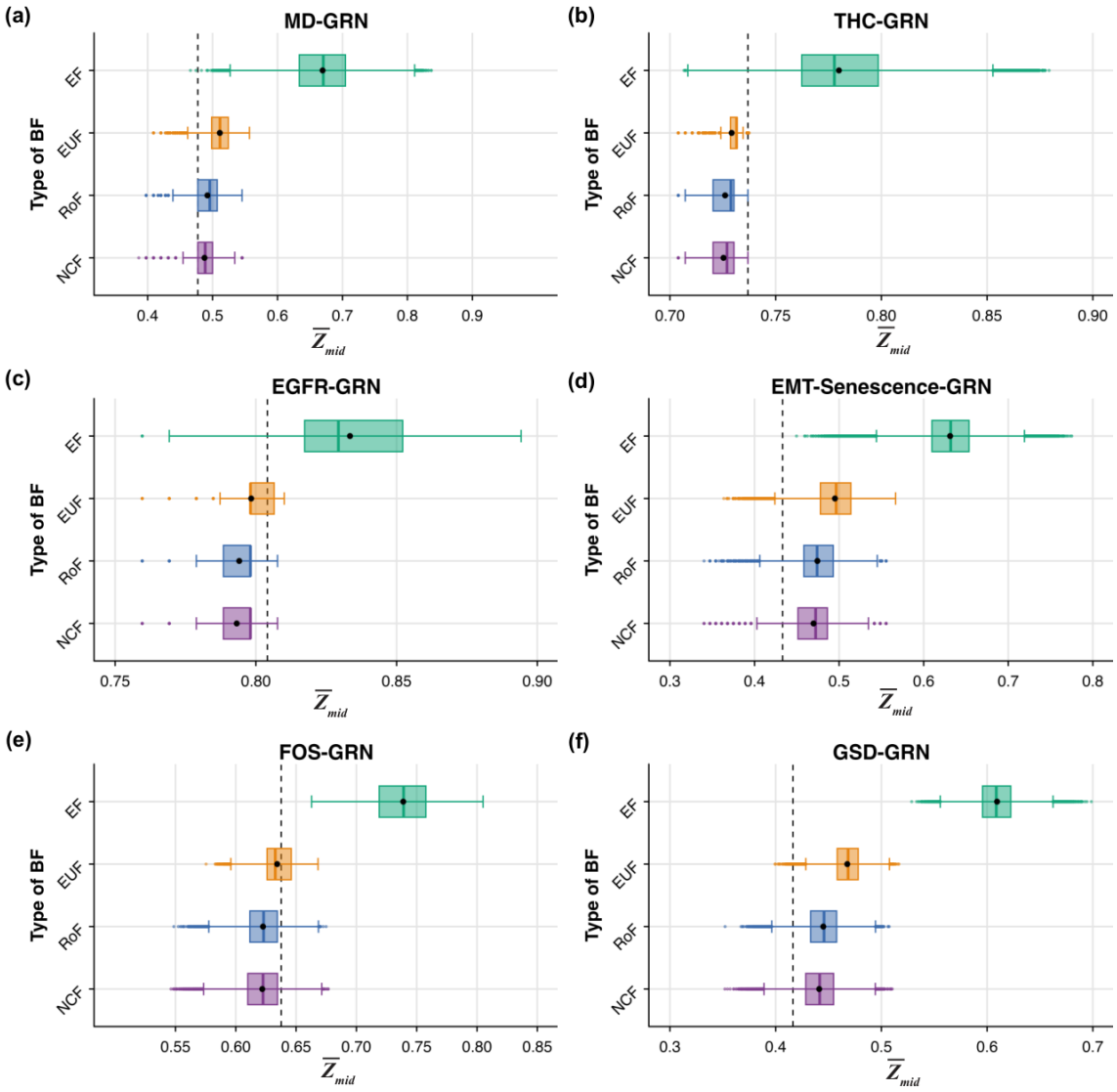


Figure B.13: Distribution of $\overline{Z}_{mid}$ values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{Z}_{mid}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{Z}_{mid}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{Z}_{mid}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{Z}_{mid}$ of the Boolean model provided by the modelers in the published article.

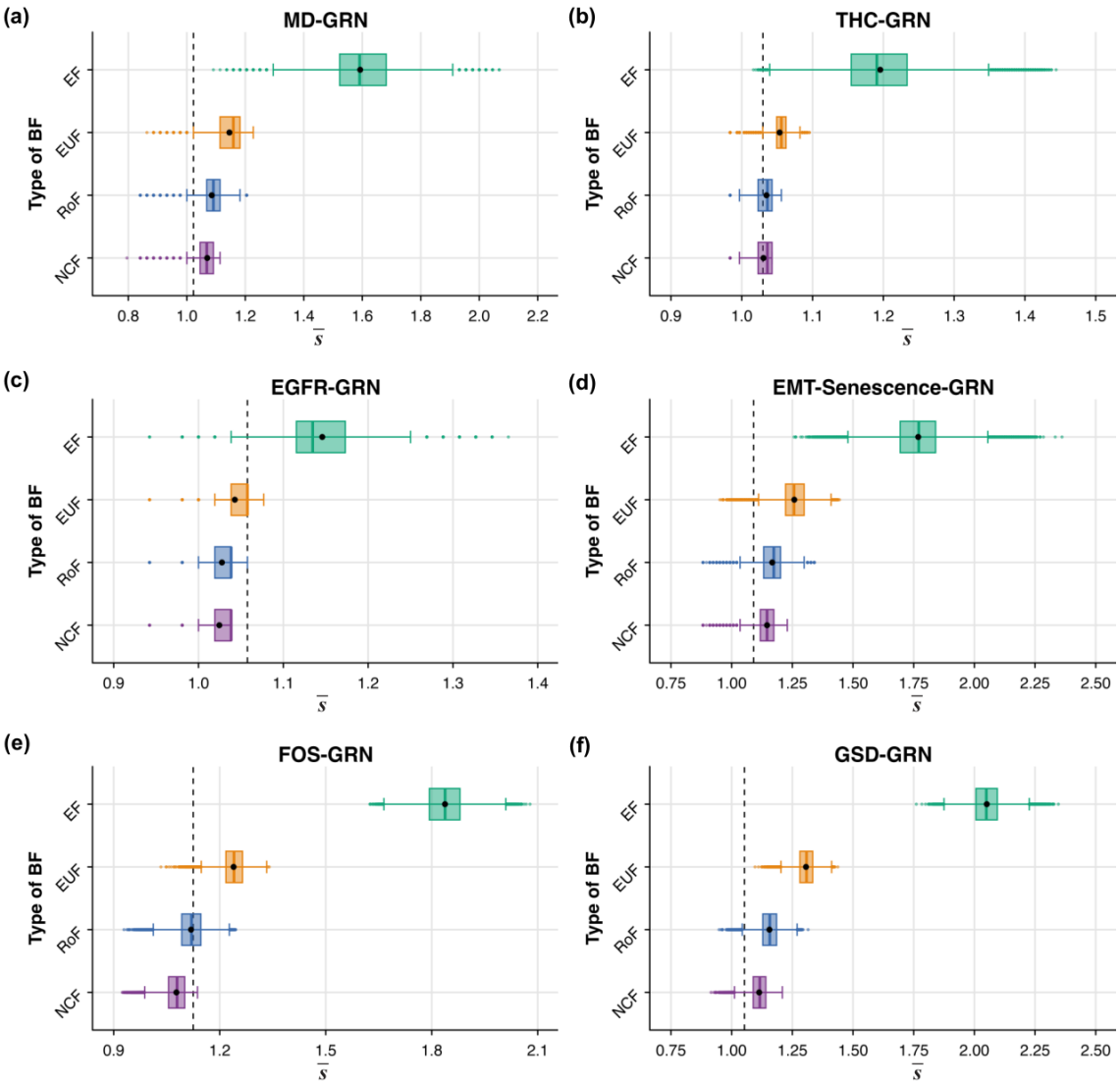


Figure B.14: Distribution of $\overline{s}$ values for different ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The x and y axes of each sub-figure correspond respectively to $\overline{s}$ and to the type of BF imposed on the nodes of the networks. The box plots display the distribution of $\overline{s}$ in the four ensembles that each use a given type of regulatory logic. The ensembles generated using biologically meaningful BFs (EUFs, RoFs, NCFs) have lower $\overline{s}$ compared to the ensemble generated using EFs. Note that we impose that ensembles recover the biological fixed points of the associated published model. The mean and median of the distribution are indicated by the black dot and the vertical line within the box respectively. The vertical dashed lines in each sub-figure correspond to the $\overline{s}$ of the Boolean model provided by the modelers in the published article.

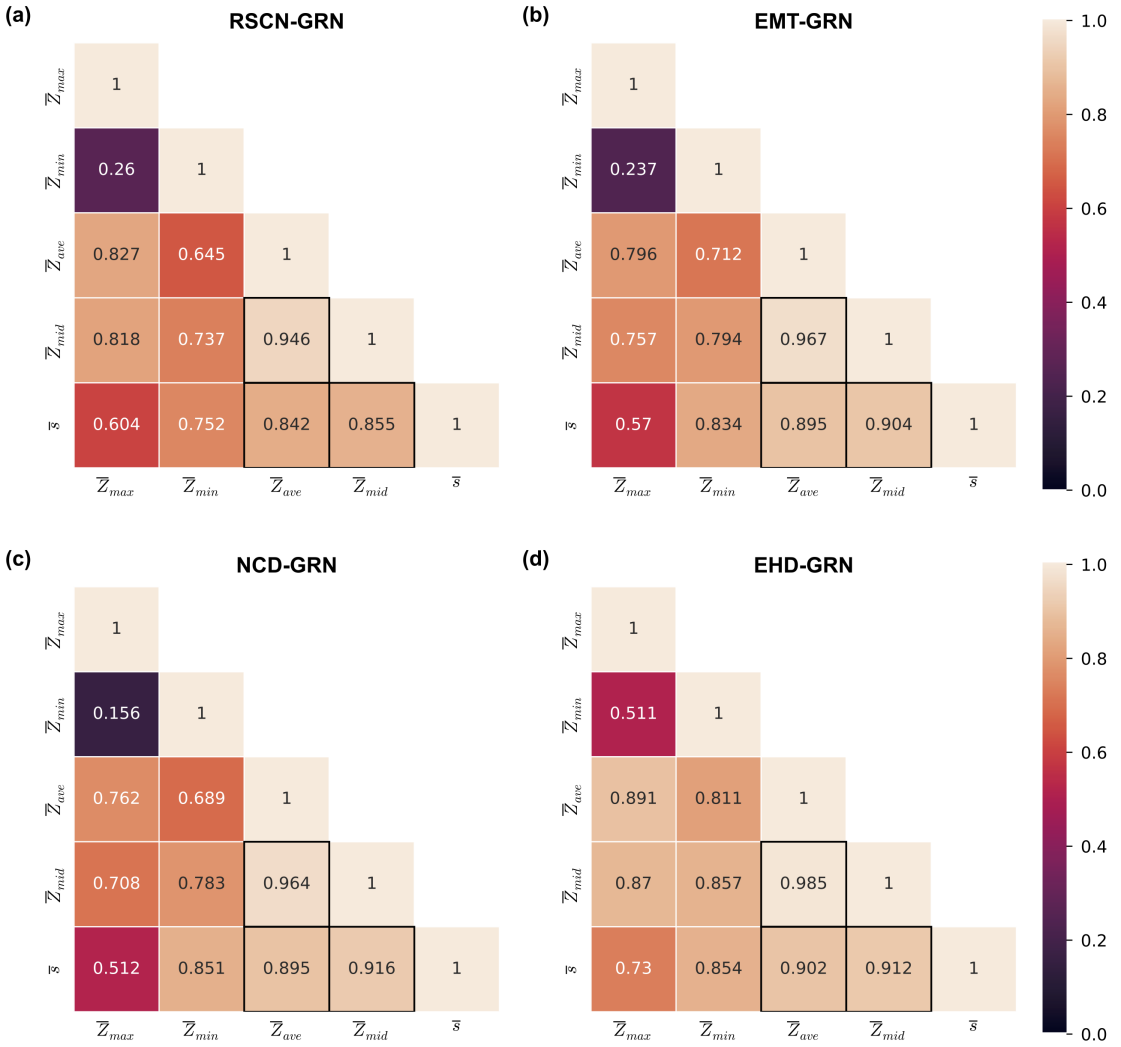


Figure B.15: Correlation heat map between the various descriptors of local dynamics on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the heat maps. The rows and columns of each sub-figure correspond to descriptors of local dynamics, namely, the network Z-parameters ($\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$) and network sensitivity ($\overline{s}$). The heat maps show the pair-wise Spearman correlations between these descriptors. All pairs of quantities show moderate to a very strong positive correlation across all models, except for one pair that shows weak positive correlation. The darkened grid highlights the descriptors that best capture the global measures of bushiness, namely, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$ and $\overline{s}$.

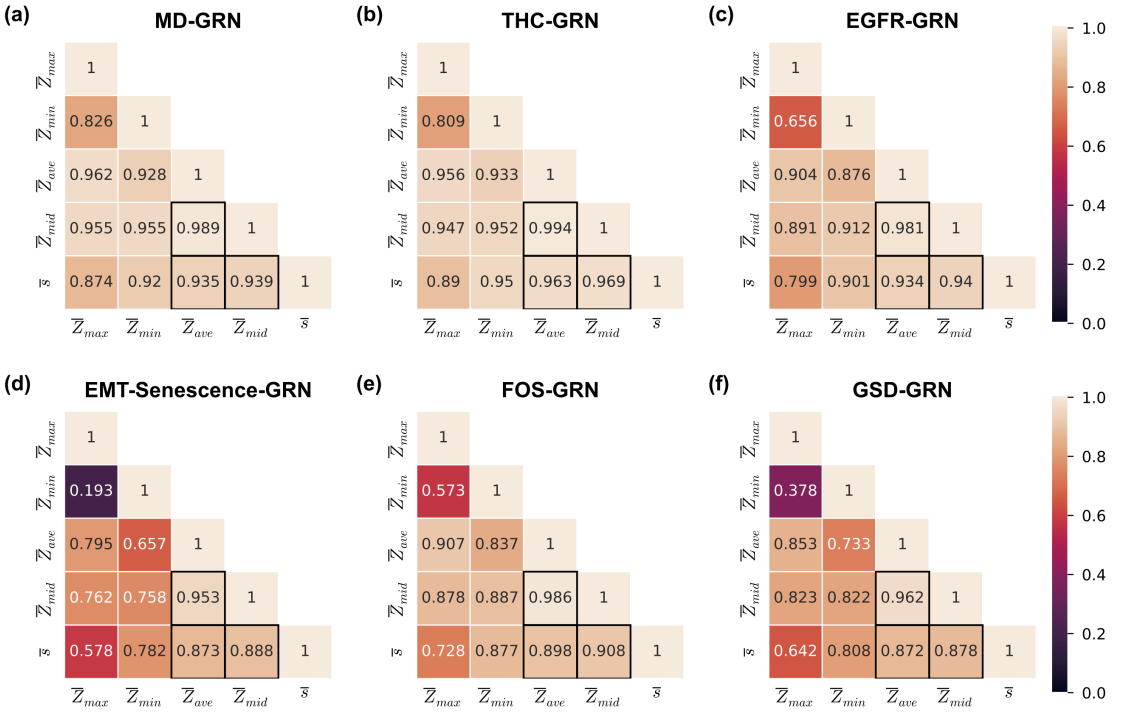


Figure B.16: Correlation heat map between the various descriptors of local dynamics on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the heat maps. The rows and columns of each sub-figure correspond to descriptors of local dynamics, namely, the network Z-parameters ($\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$) and network sensitivity ($\overline{s}$). The heat maps show the pair-wise Spearman correlations between these descriptors. All pairs of quantities typically show moderate to a very strong positive correlation across various network structures, except for the correlations of $\overline{Z}_{max}$ with $\overline{Z}_{min}$ in 2 networks in which it is weak or very weak.

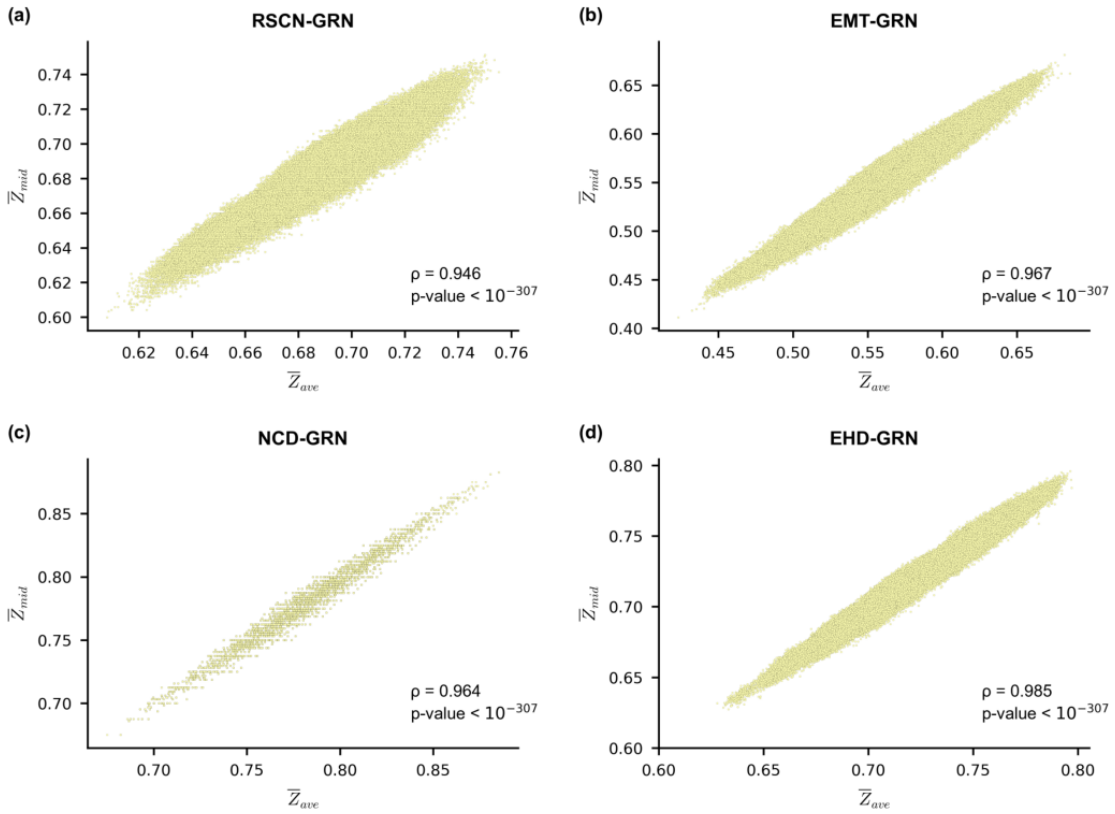


Figure B.17: Scatter plot between the $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ show a strong positive correlation for all networks.

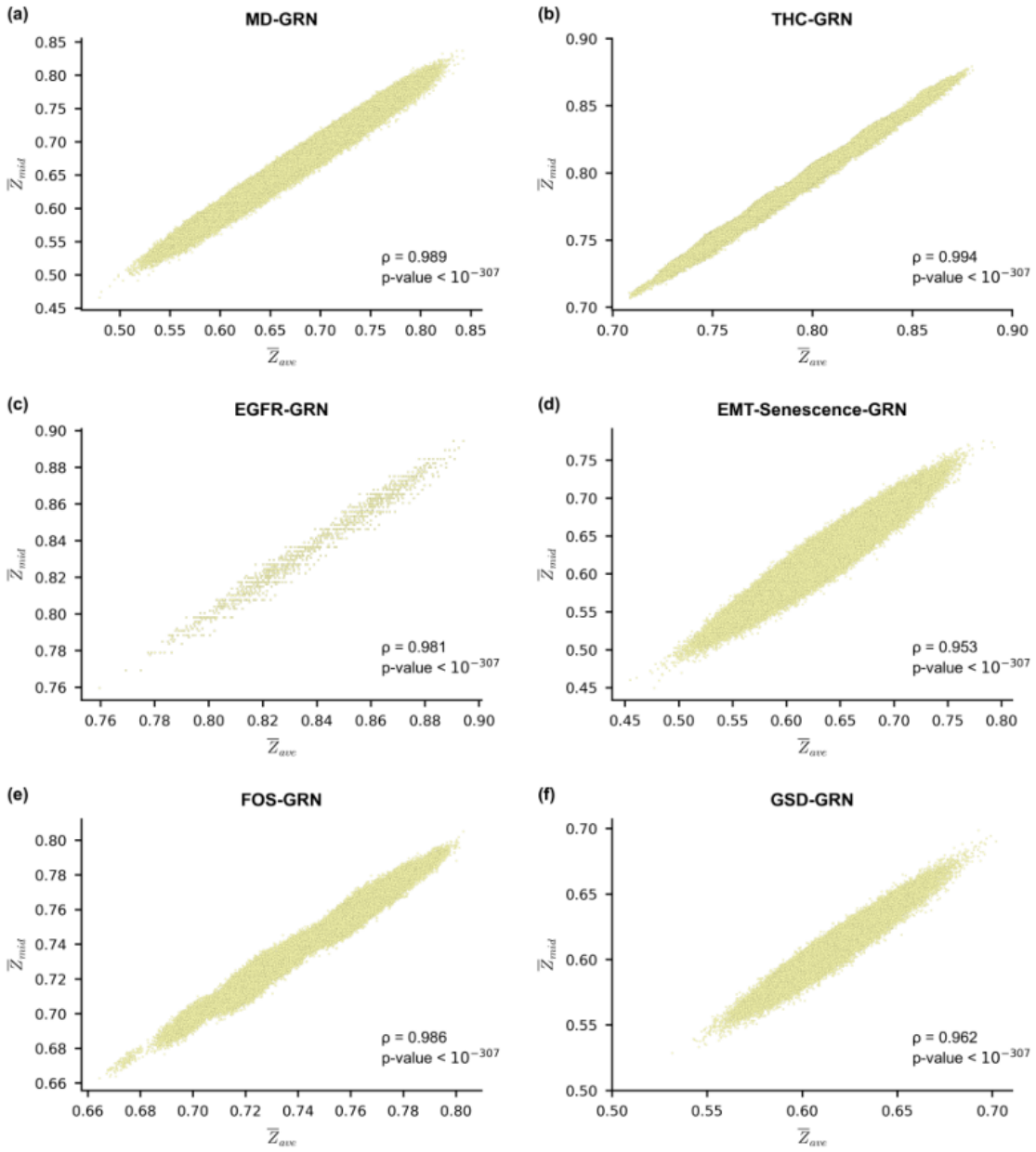


Figure B.18: Scatter plot between the $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{mid}$ and $\overline{Z}_{ave}$ show a strong positive correlation for all networks.

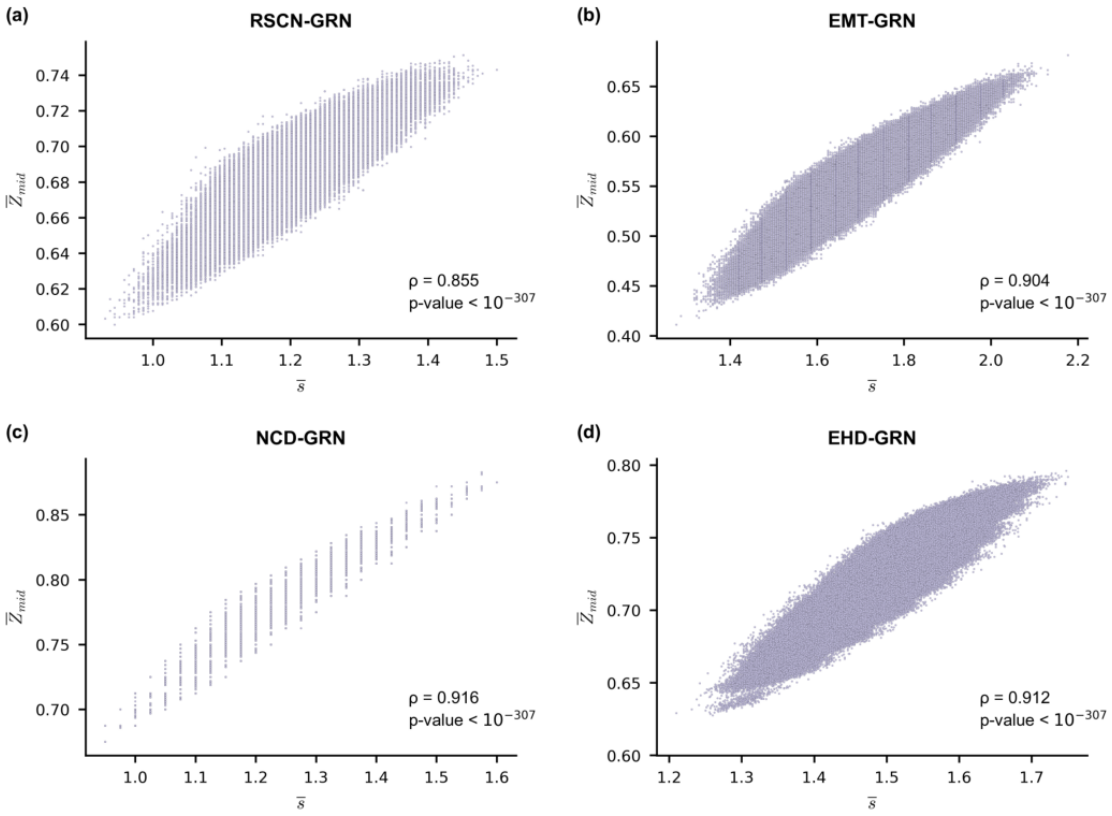


Figure B.19: Scatter plot between the $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and $\overline{Z}_{mid}$ show a strong positive correlation for all networks.

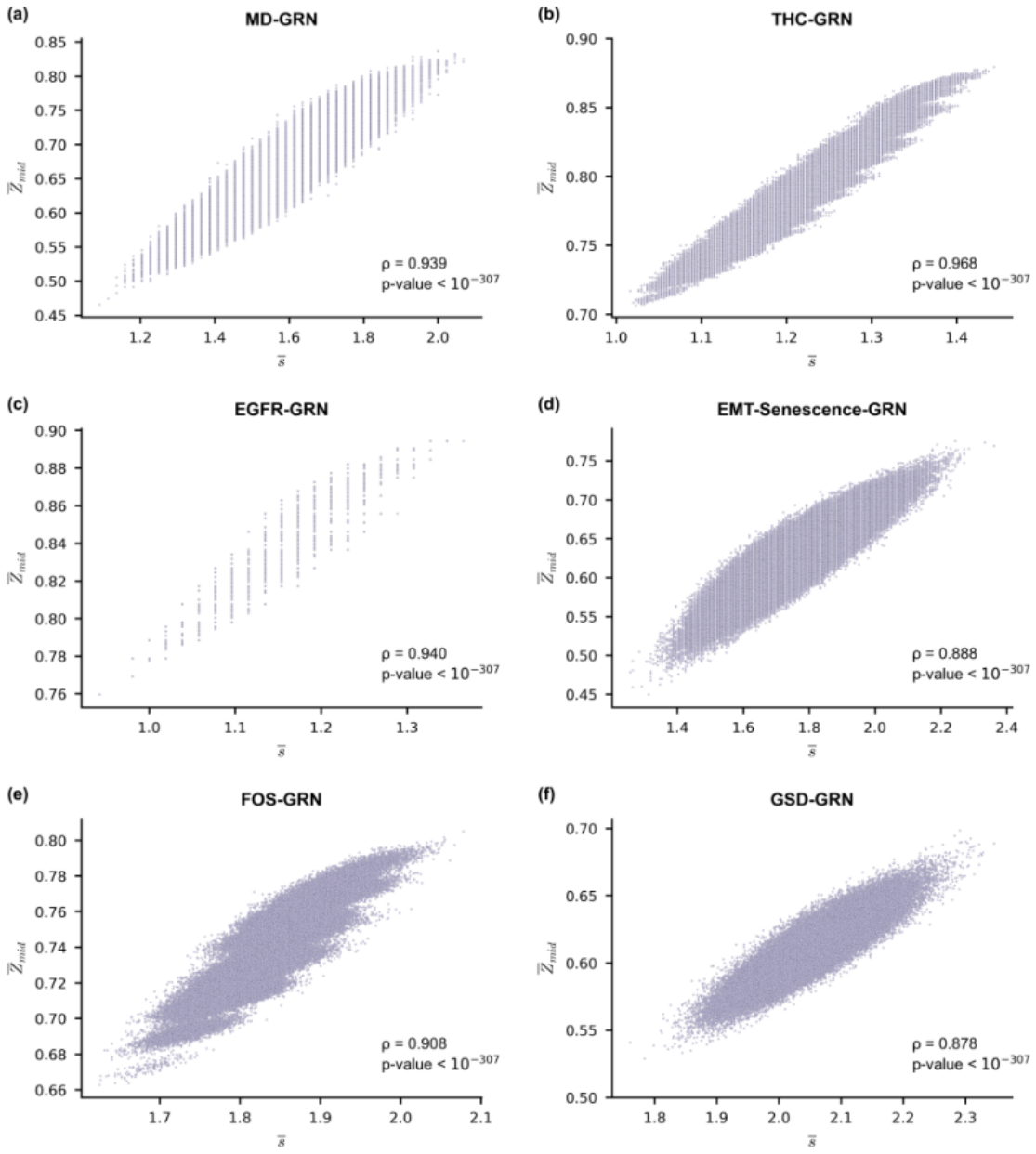


Figure B.20: Scatter plot between the $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and network sensitivity ($\overline{s}$) respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and $\overline{Z}_{mid}$ show a strong positive correlation for all networks.

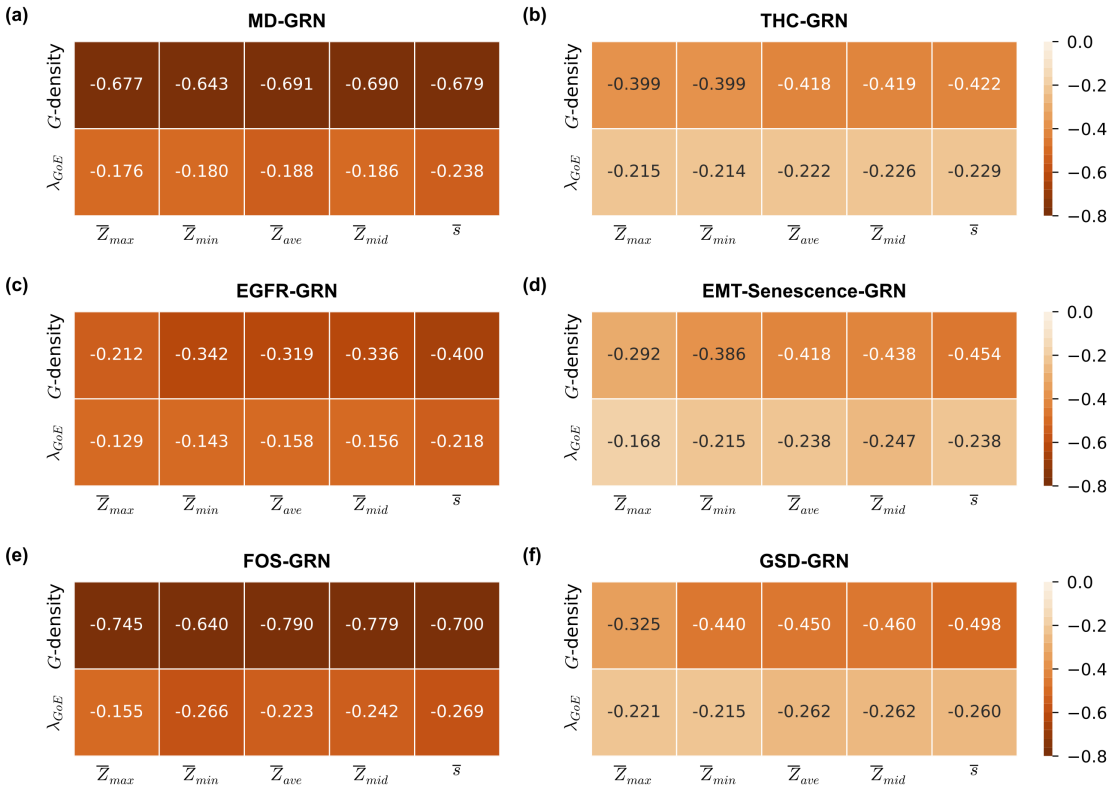


Figure B.21: Correlation heat map between descriptors of local dynamics and global measures of bushiness in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the heat maps. The rows of each sub-figure correspond to quantities computed on the STG, namely G-density and average convergence rate of trajectories originating at GoE states (λ_{GoE}). The columns of each sub-figure correspond to descriptors of local dynamics, namely, the network Z-parameters ($\overline{Z}_{max}$, $\overline{Z}_{min}$, $\overline{Z}_{ave}$, $\overline{Z}_{mid}$) and network sensitivity ($\overline{s}$). The heat maps show the pair-wise Spearman correlations between descriptors of local dynamics and global measures of bushiness defined on the STG. The descriptors of local dynamics show a moderate to strong negative correlation with the G-density and very weak to moderate negative correlation with λ_{GoE}.

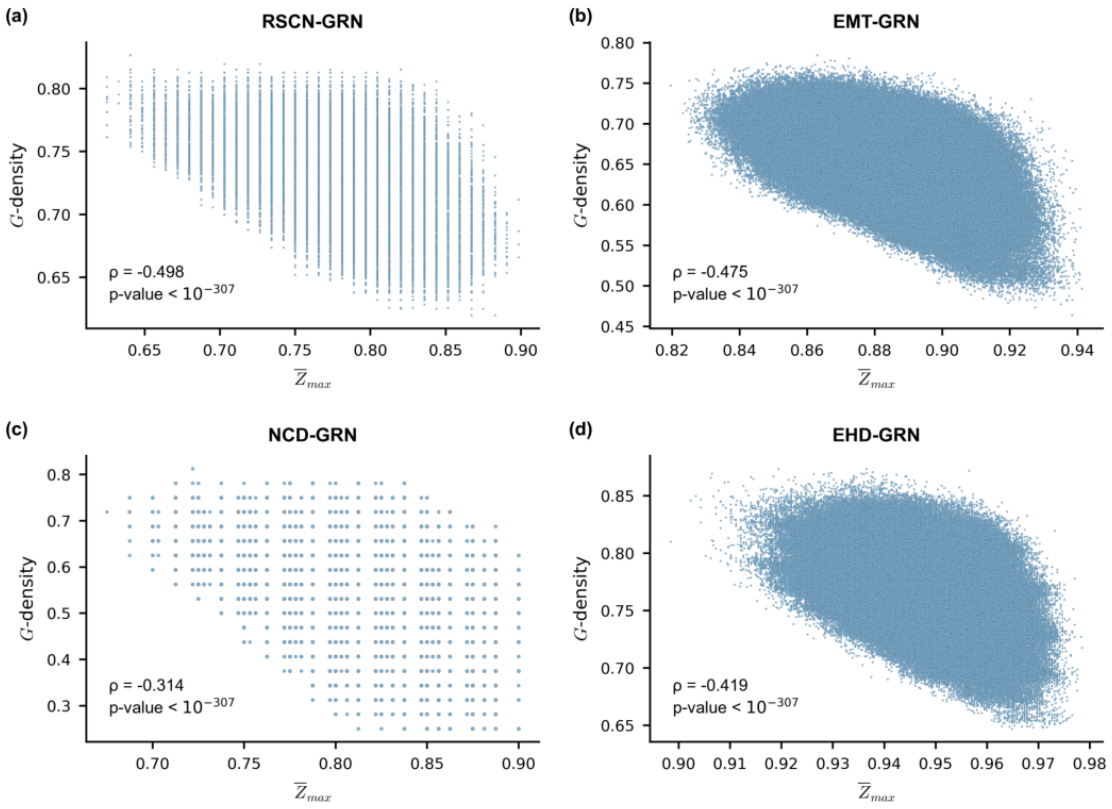


Figure B.22: Scatter plot between the $\overline{Z}_{max}$ and G-density on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{max}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{max}$ and G-density show a moderate negative correlation for all network except for NCD-GRN for which there is a weak negative correlation.

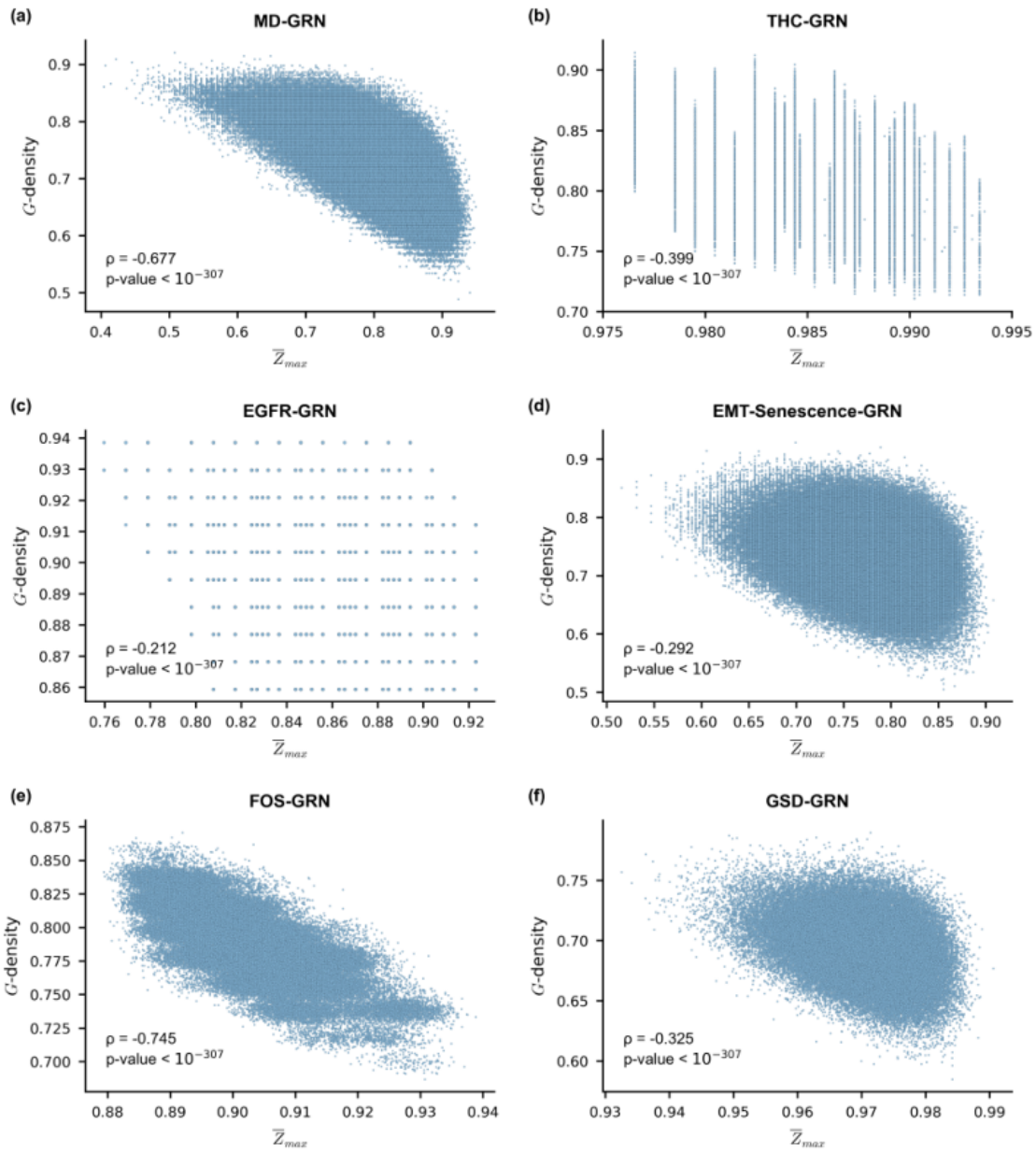


Figure B.23: Scatter plot between the $\overline{Z}_{max}$ and G-density on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{max}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{max}$ and G-density show a strong negative correlation for FOS-GRN and MD-GRN and a weak negative correlation otherwise.

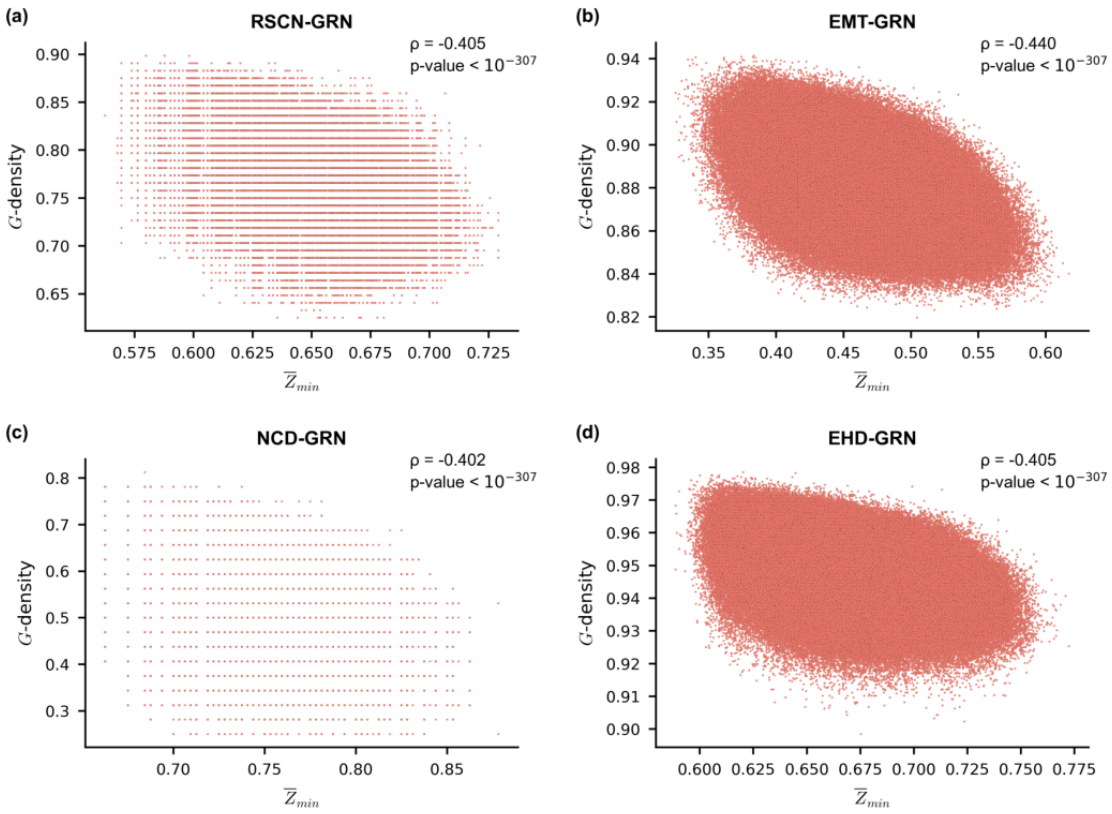


Figure B.24: Scatter plot between the $\overline{Z}_{min}$ and G-density on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{min}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{min}$ and G-density show a moderate negative correlation for all networks.

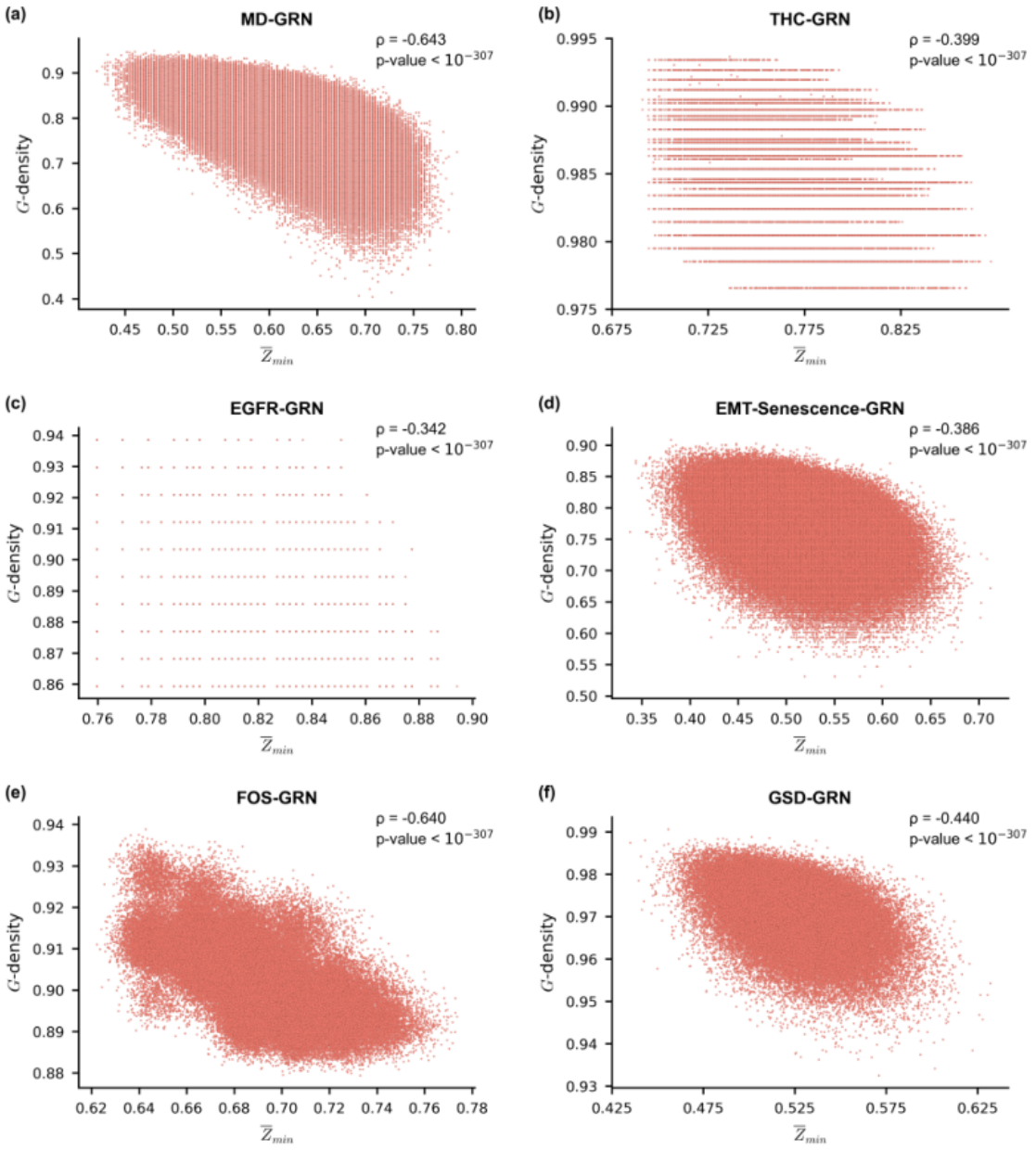


Figure B.25: Scatter plot between the $\overline{Z}_{min}$ and G-density on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{min}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{min}$ and G-density show a strong negative correlation for MD-GRN and FOS-GRN, a moderate negative correlation for GSD-GRN and a weak negative correlation otherwise.

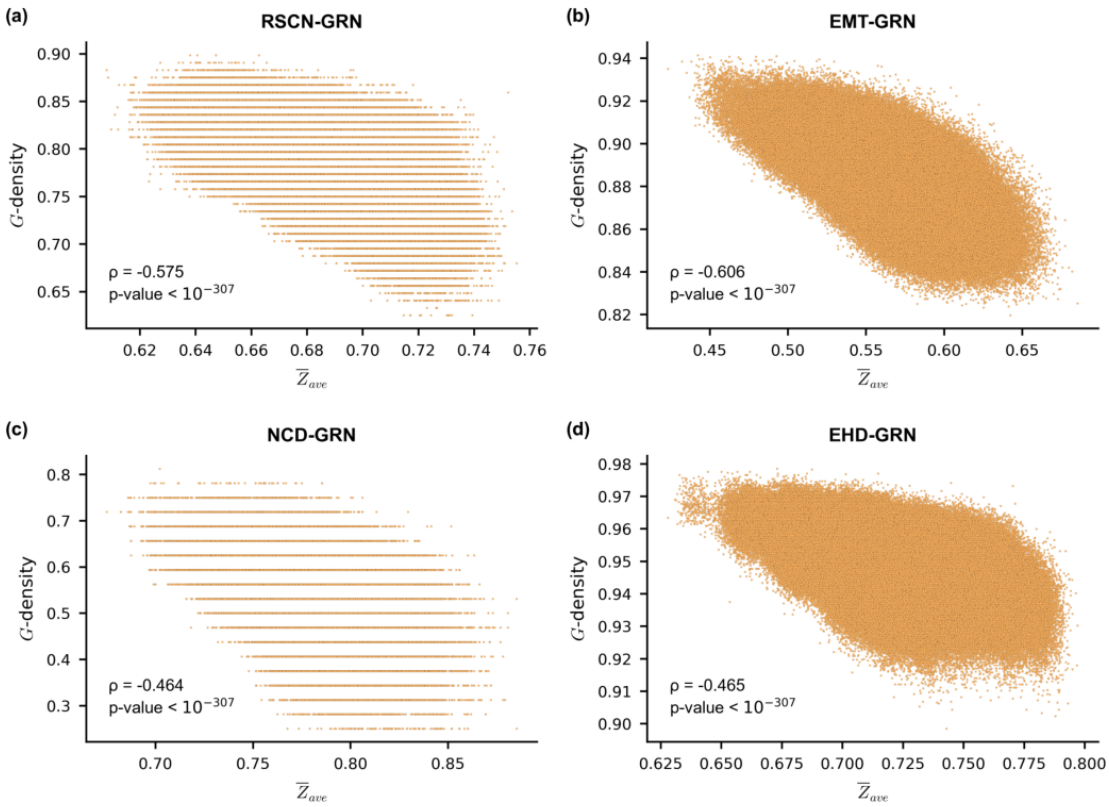


Figure B.26: Scatter plot between the $\overline{Z}_{ave}$ and G-density on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{ave}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{ave}$ and G-density show a strong negative correlation for EMT-GRN and a moderate negative correlation otherwise.

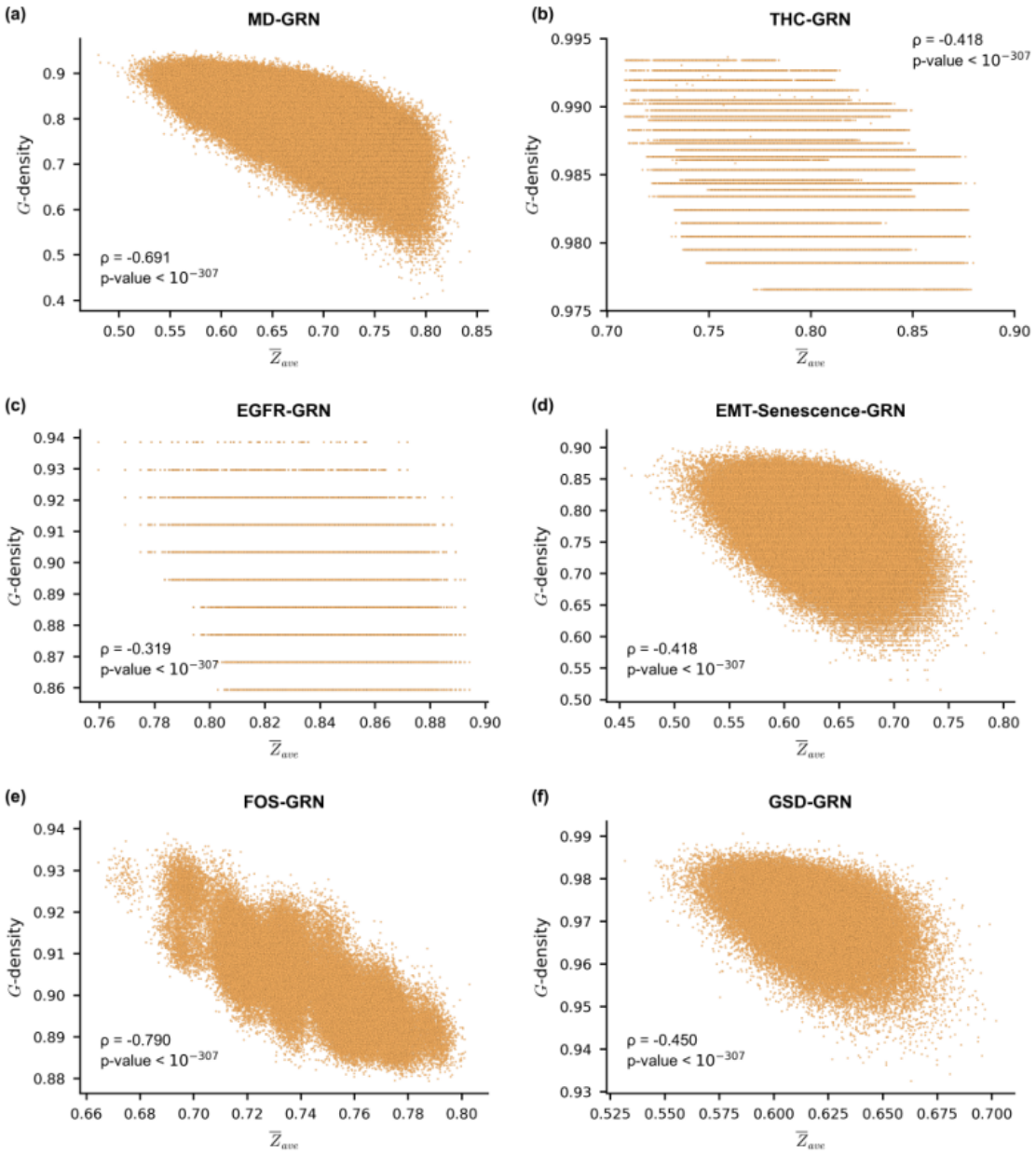


Figure B.27: Scatter plot between the $\overline{Z}_{ave}$ and G-density on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{ave}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{ave}$ and G-density show a strong negative correlation for FOS-GRN and MD-GRN, a moderate negative correlation for GSD-GRN, EMT-Senescence-GRN and THC-GRN, and a weak negative correlation for EGFR-GRN.

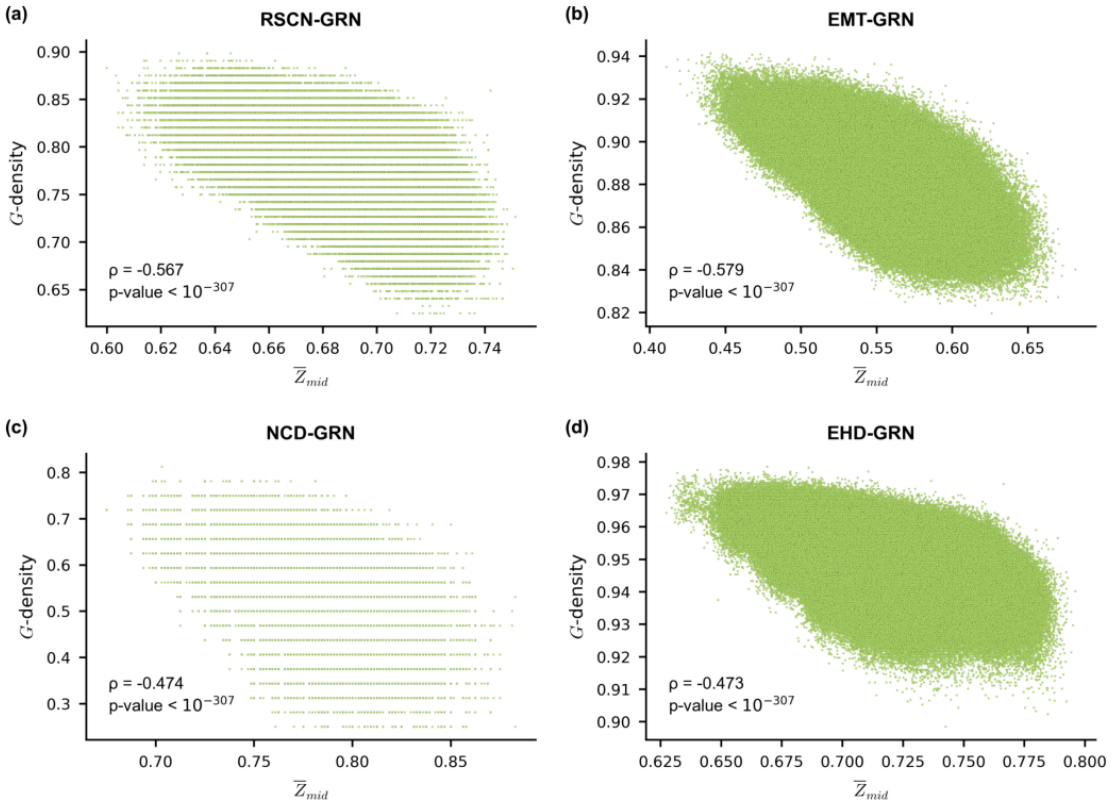


Figure B.28: Scatter plot between the $\overline{Z}_{mid}$ and G-density on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{mid}$ and G-density show a moderate negative correlation for all the networks.

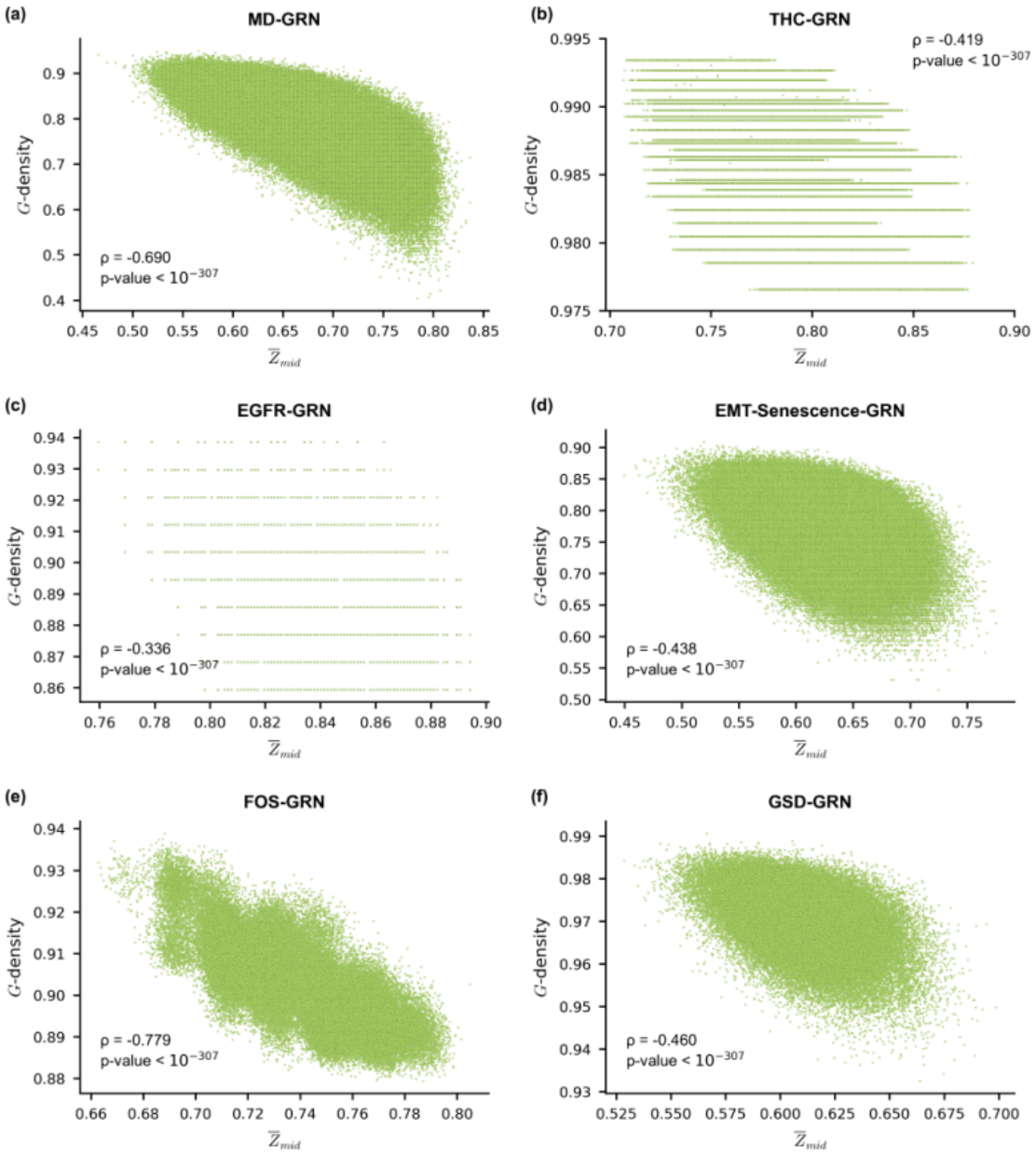


Figure B.29: Scatter plot between the $\overline{Z}_{mid}$ and G-density on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{mid}$ and G-density show a strong negative correlation for MD-GRN and FOS-GRN, a moderate negative correlation for other networks except EGFR-GRN for which there is a weak negative correlation.

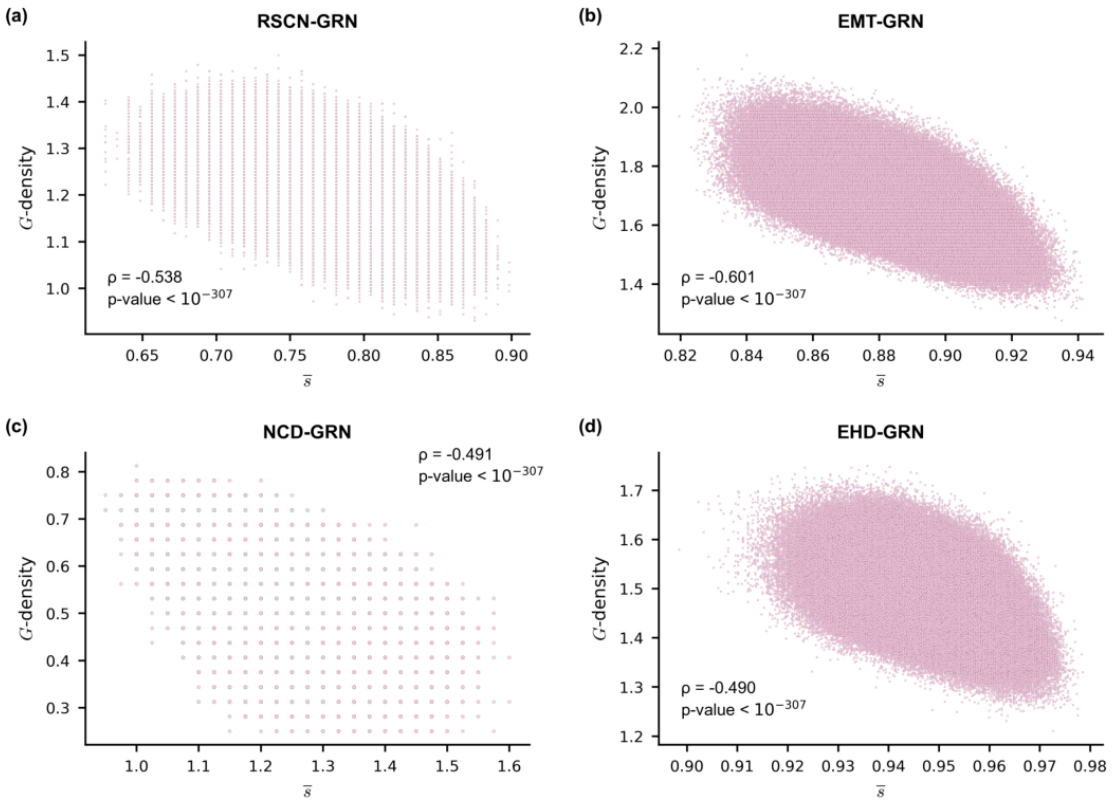


Figure B.30: Scatter plot between the network sensitivity ($\overline{s}$) and G-density on ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to the network sensitivity ($\overline{s}$) and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and G-density show a strong negative correlation for EMT-GRN and a moderate negative correlation otherwise.

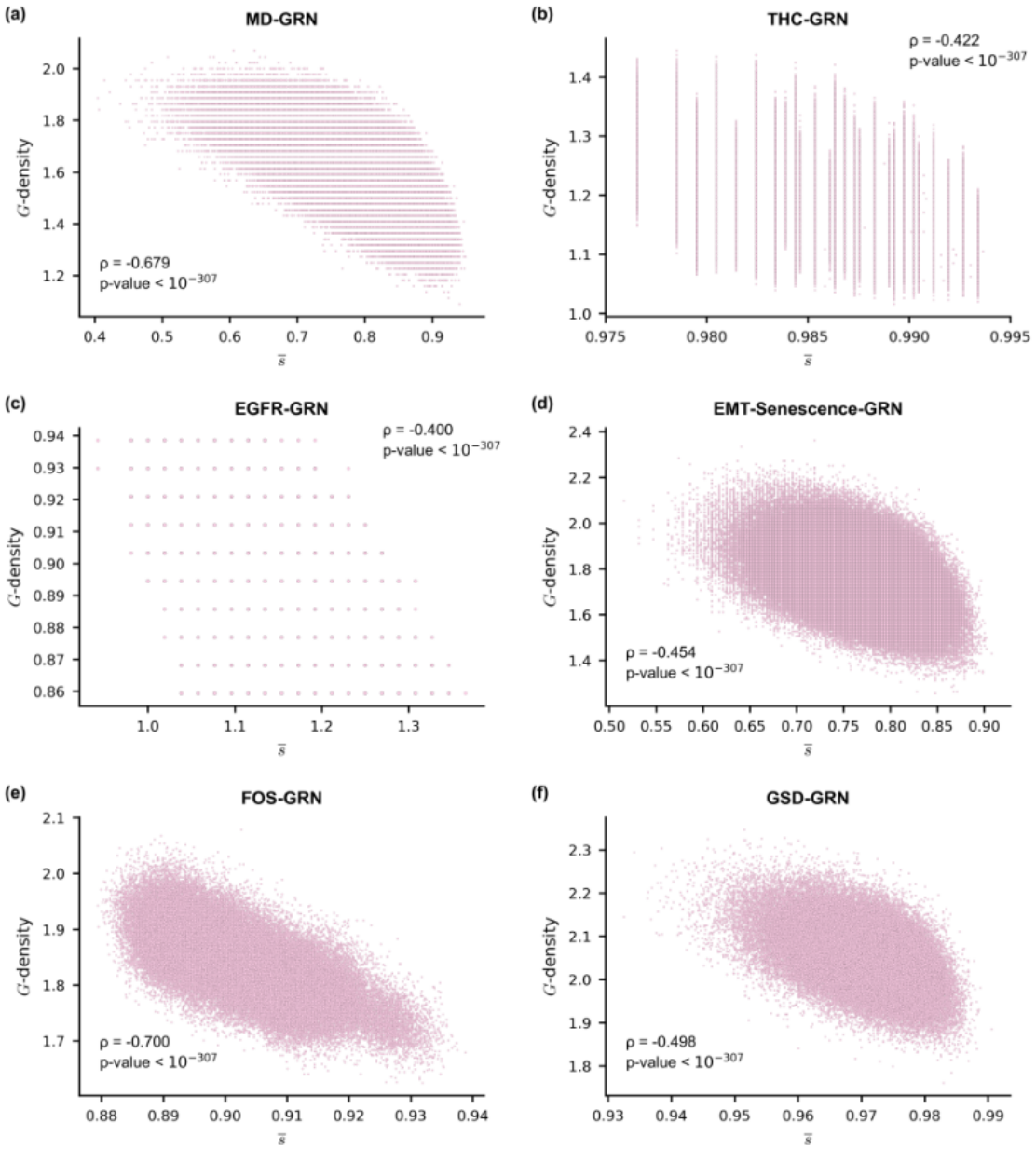


Figure B.31: Scatter plot between the network sensitivity ($\overline{s}$) and *G*-density on ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to the network sensitivity ($\overline{s}$) and G-density respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and G-density show a strong negative correlation for MD-GRN and FOS-GRN, and a moderate negative correlation otherwise.

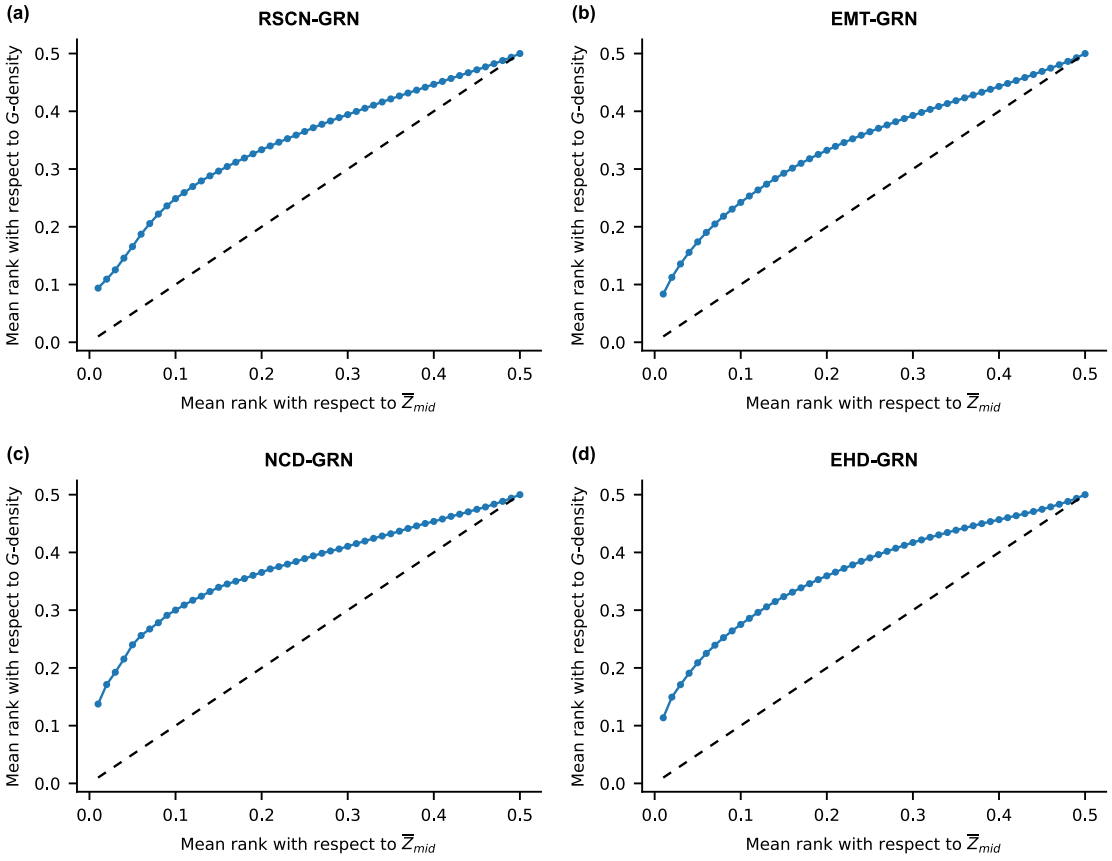


Figure B.32: Predictive power of $\overline{Z}_{mid}$ for G-density in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the plots. The x-axis corresponds to the average relative rank of the 'top' N models (ranked in the ascending order) with the 'lowest' $\overline{Z}_{mid}$ values. The values on the x-axis is given by $N/(2 * N_{total})$, where N is the number of top models and N_{total} is the total number of models. The y-axis corresponds to the average relative rank of those N models when they are ranked (in the descending) according to the G-density. The black dashed line $y = x$ indicates the ideal scenario of perfect prediction of the y value when x is known. The blue solid line indicates how predictive $\overline{Z}_{mid}$ is for the G-density for each GRN.

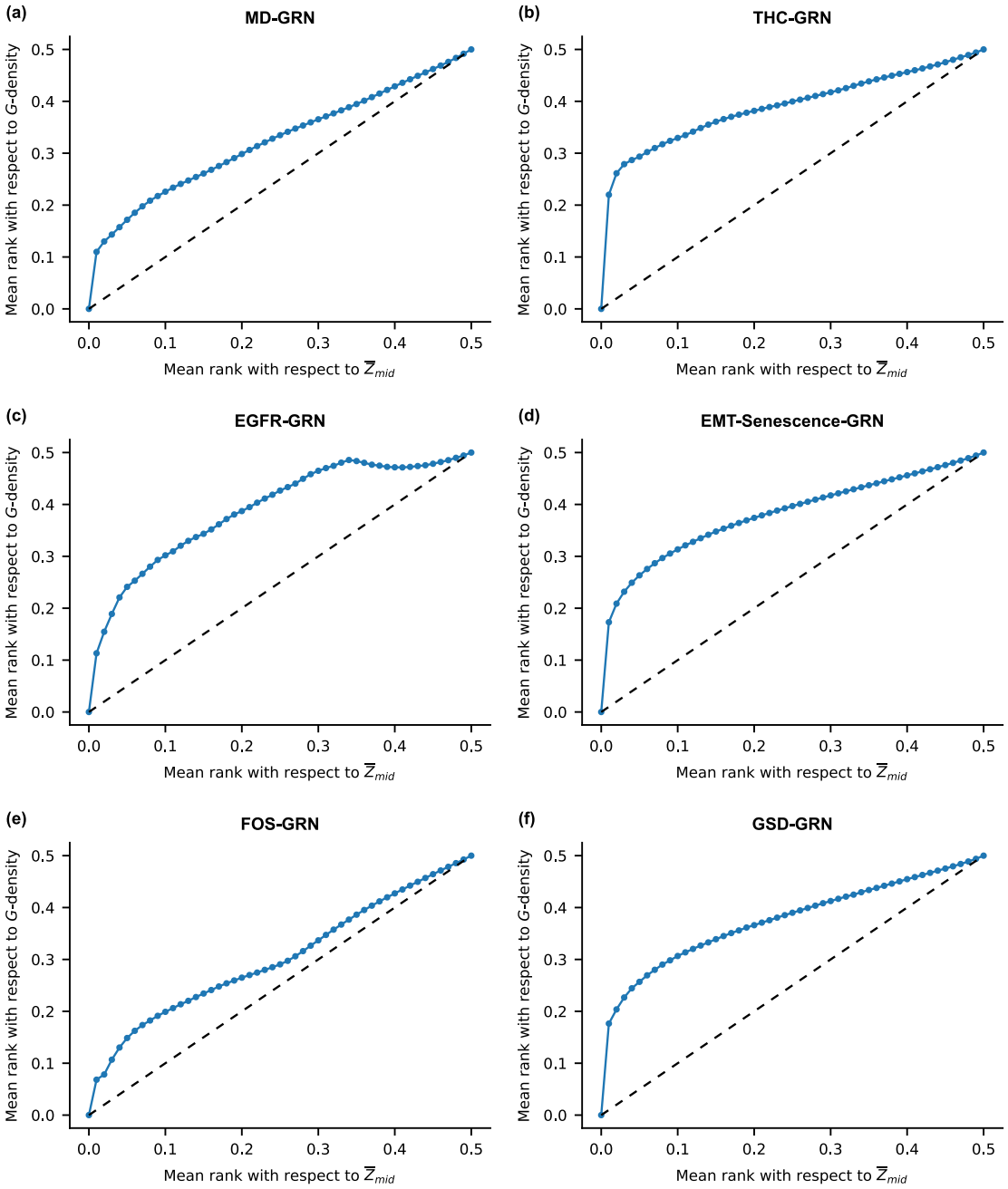


Figure B.33: Predictive power of $\overline{Z}_{mid}$ for G-density in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the plots. The x-axis corresponds to the average relative rank of the 'top' N models (ranked in the ascending order) with the 'lowest' $\overline{Z}_{mid}$ values. The values on the x-axis is given by $N/2 * N_{total}$, where N is the number of top models and N_{total} is the total number of models. The y-axis corresponds to the average relative rank of those N models when they are ranked (in the descending) according to the G-density. The black dashed line $y = x$ indicates the ideal scenario of perfect prediction of the y value when x is known. The blue solid line indicates how predictive $\overline{Z}_{mid}$ is for the G-density for each GRN.

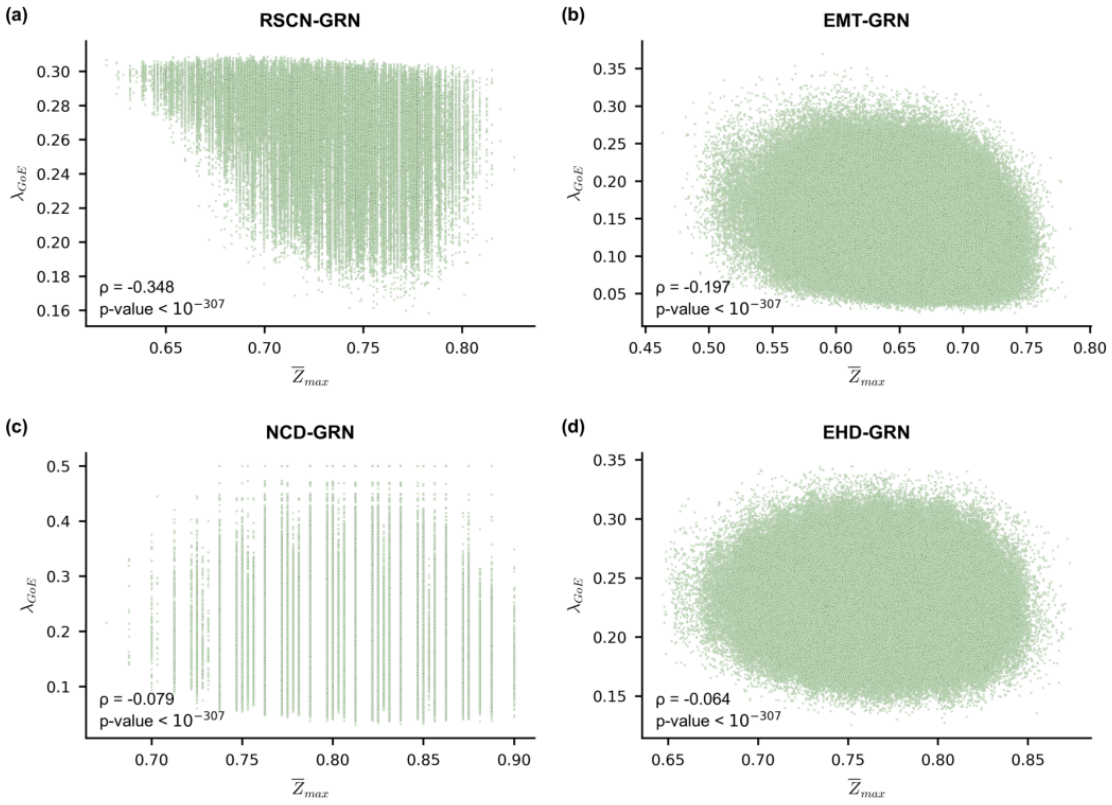


Figure B.34: Scatter plot between the $\overline{Z}_{max}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{max}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{max}$ and λ_{GoE} show a weak negative correlation for the RSCN-GRN network and a very weak negative correlation otherwise.

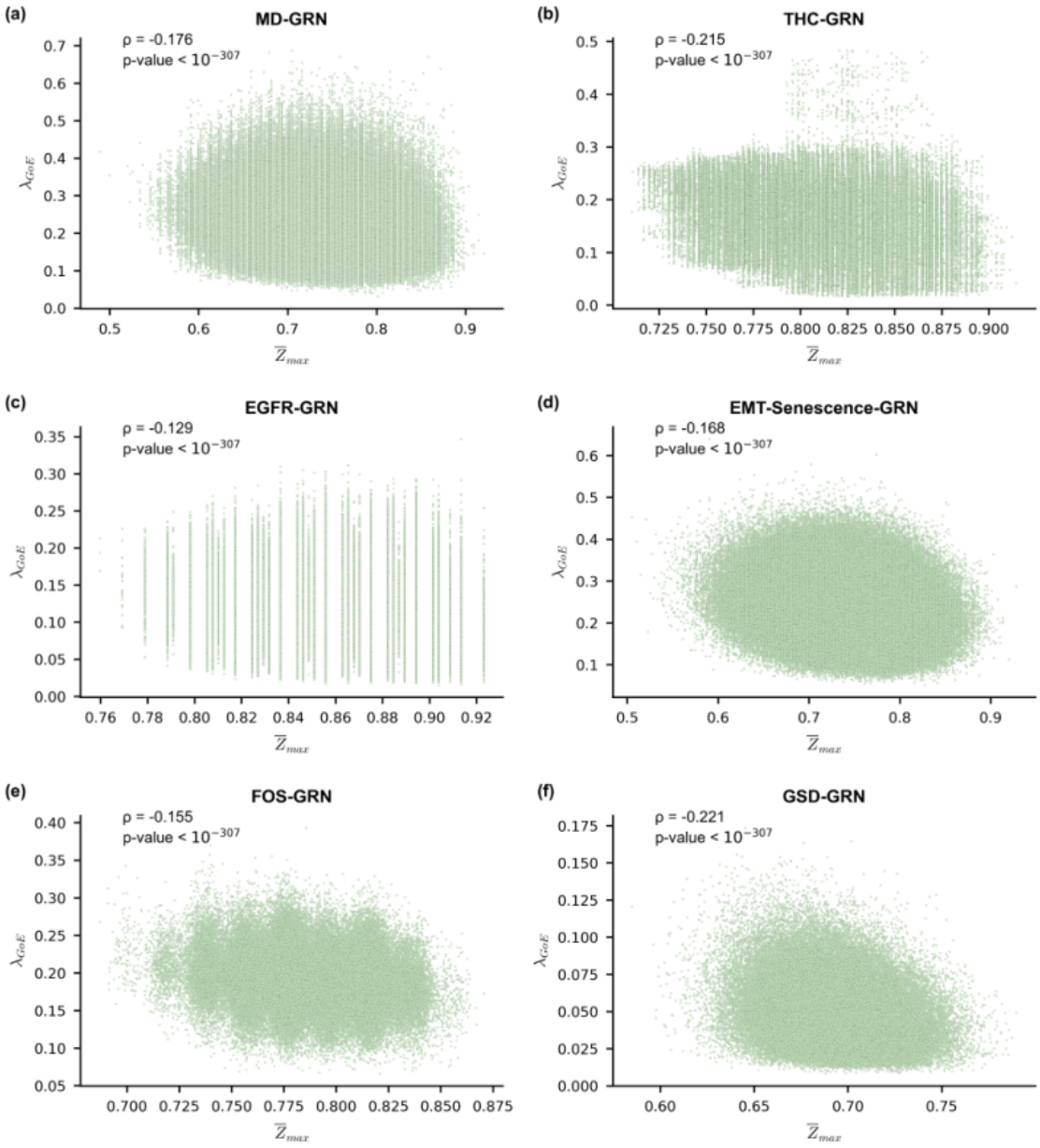


Figure B.35: Scatter plot between the $\overline{Z}_{max}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{max}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{max}$ and λ_{GoE} show a weak negative correlation for the THC-GRN and GSD-GRN networks and a very weak negative correlation otherwise.

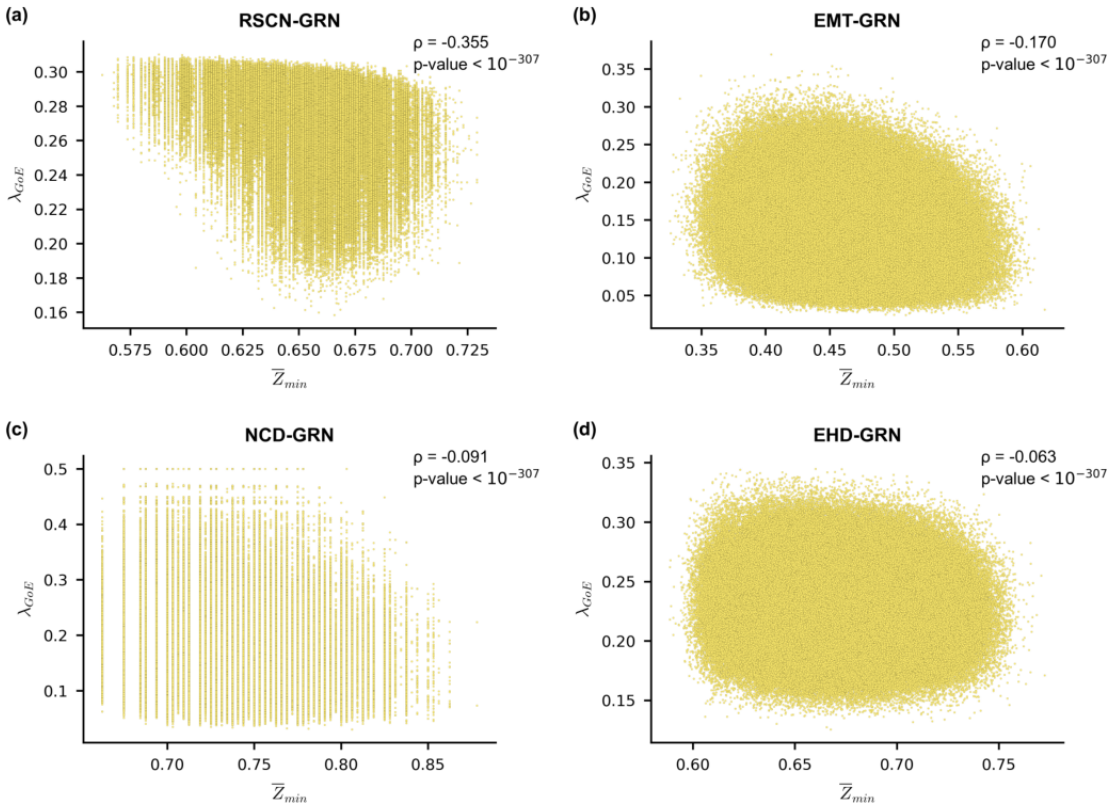


Figure B.36: Scatter plot between the $\overline{Z}_{min}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{min}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{min}$ and λ_{GoE} show a weak negative correlation for RSCN-GRN and a very weak negative correlation otherwise.

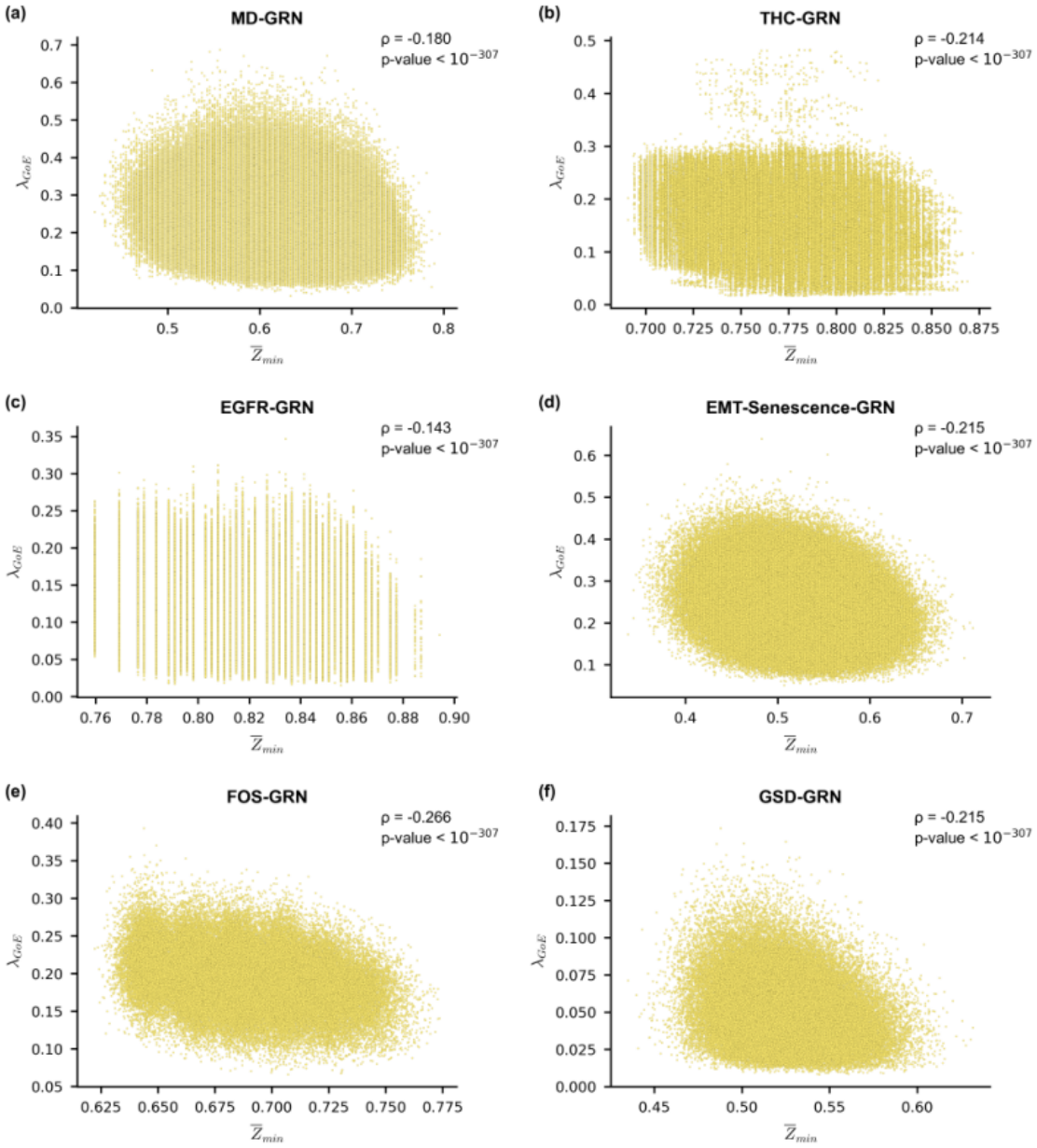


Figure B.37: Scatter plot between the $\overline{Z}_{min}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{min}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{min}$ and λ_{GoE} show a weak negative correlation for all networks except for MD-GRN and EGFR-GRN for which there is a very weak negative correlation.

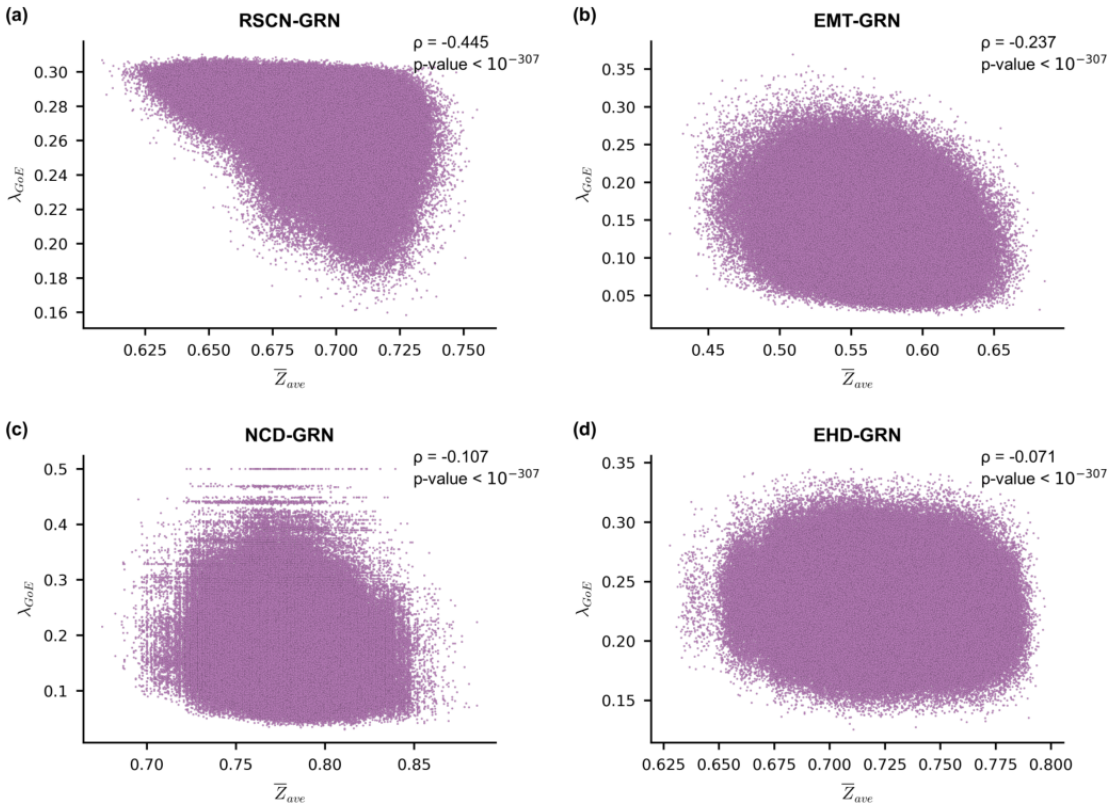


Figure B.38: Scatter plot between the $\overline{Z}_{ave}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{ave}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{ave}$ and λ_{GoE} show a moderate negative correlation for RSCN-GRN, a weak negative correlation for EMT-GRN and a very weak negative correlation otherwise.

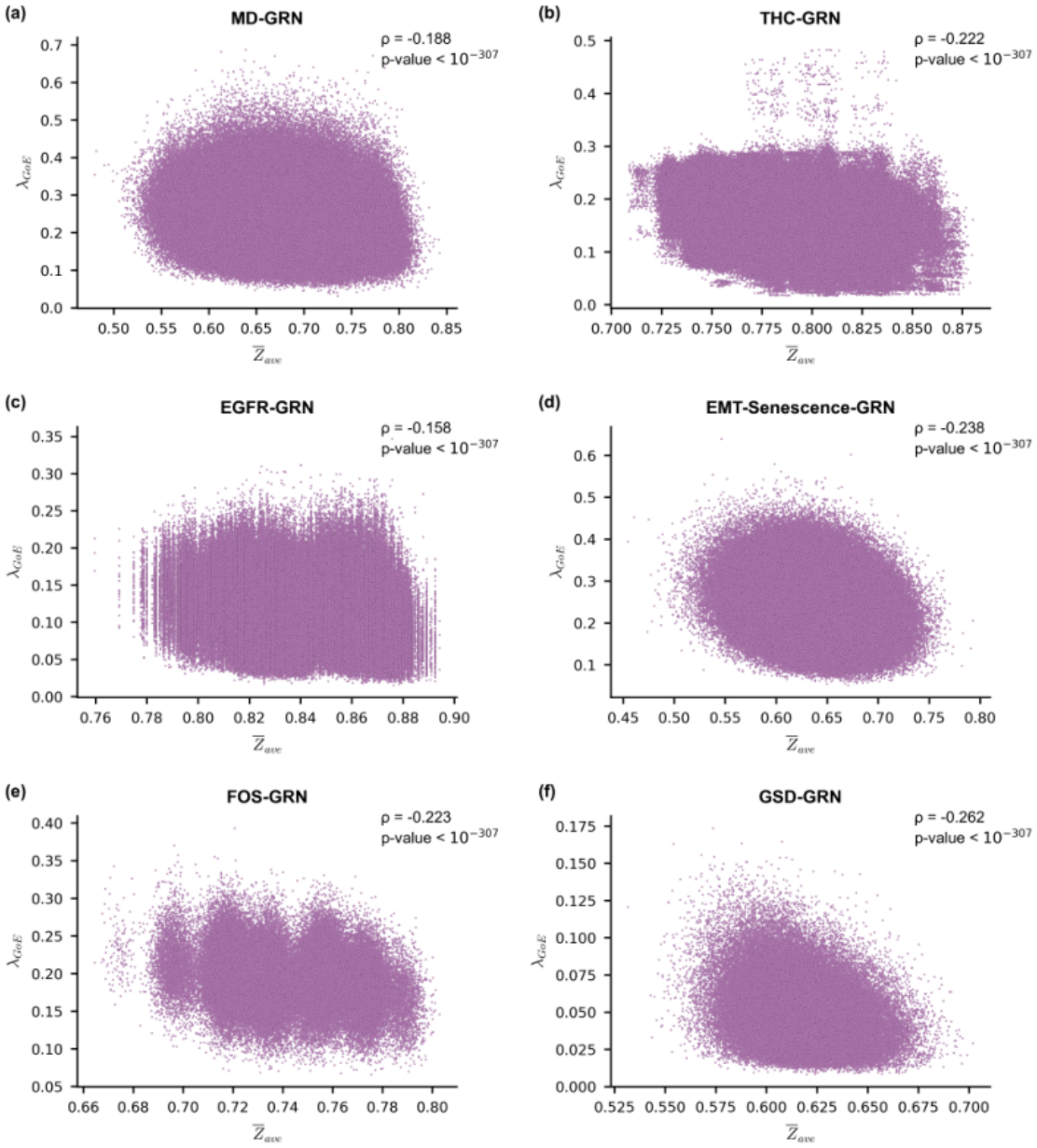


Figure B.39: Scatter plot between the $\overline{Z}_{ave}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{ave}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{ave}$ and λ_{GoE} show a weak negative correlation for THC-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN, and very weak negative correlation otherwise.

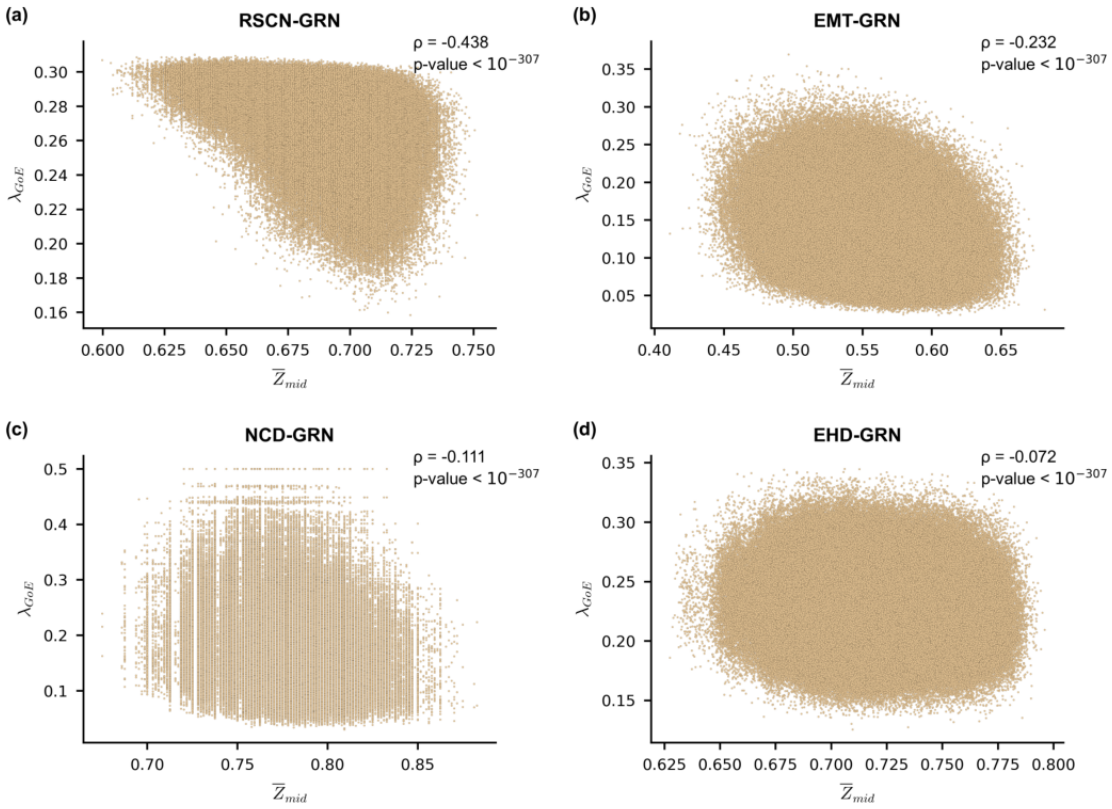


Figure B.40: Scatter plot between the $\overline{Z}_{mid}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{Z}_{mid}$ and λ_{GoE} show a moderate negative correlation for RSCN-GRN, a weak negative correlation for EMT-GRN and a weak negative correlation otherwise.

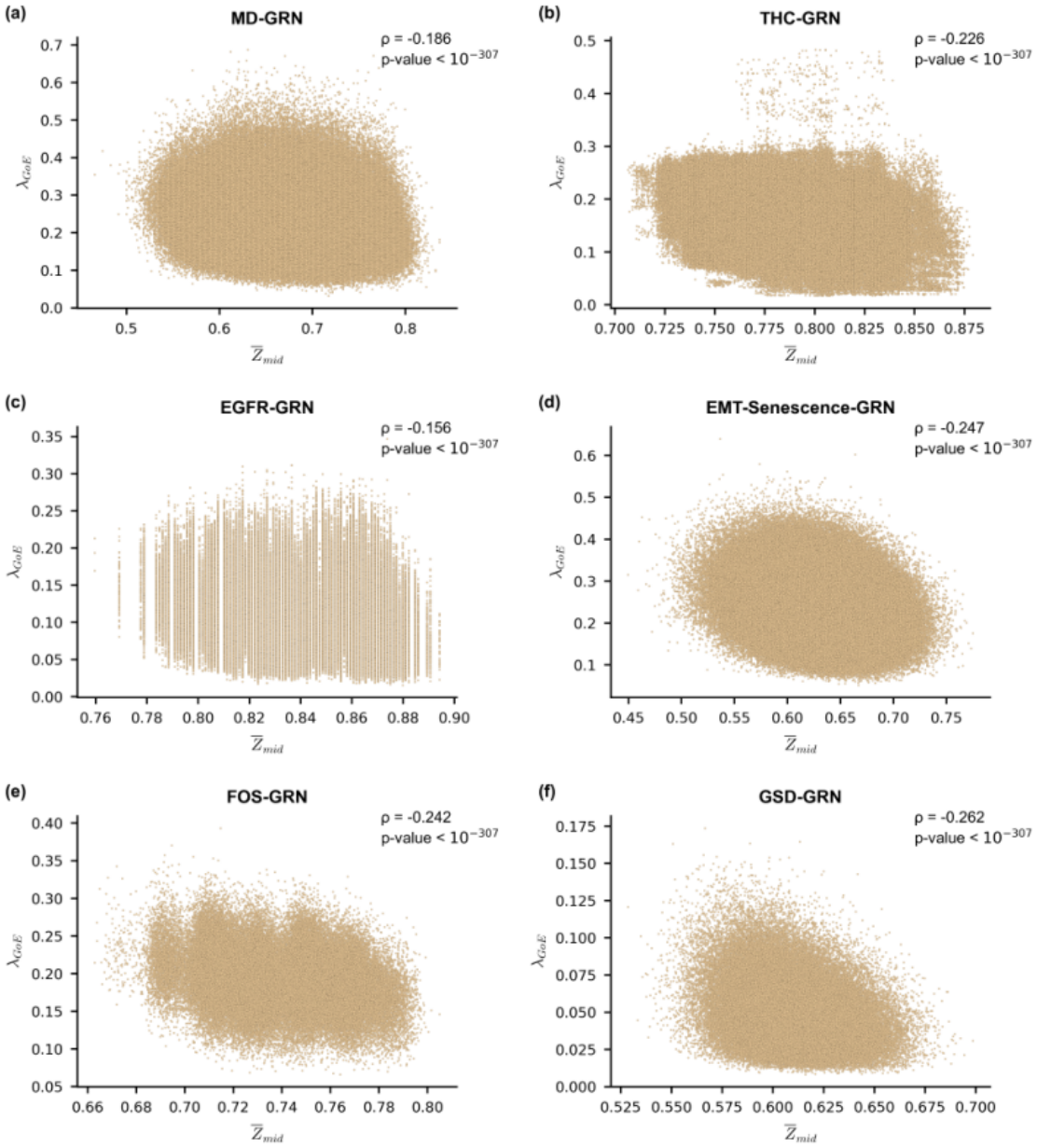


Figure B.41: Scatter plot between the $\overline{Z}_{mid}$ and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to $\overline{Z}_{mid}$ and λ_{GoE} respectively. $\overline{Z}_{mid}$ and λ_{GoE} show a weak negative correlation for THC-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN and a very weak negative correlation otherwise.

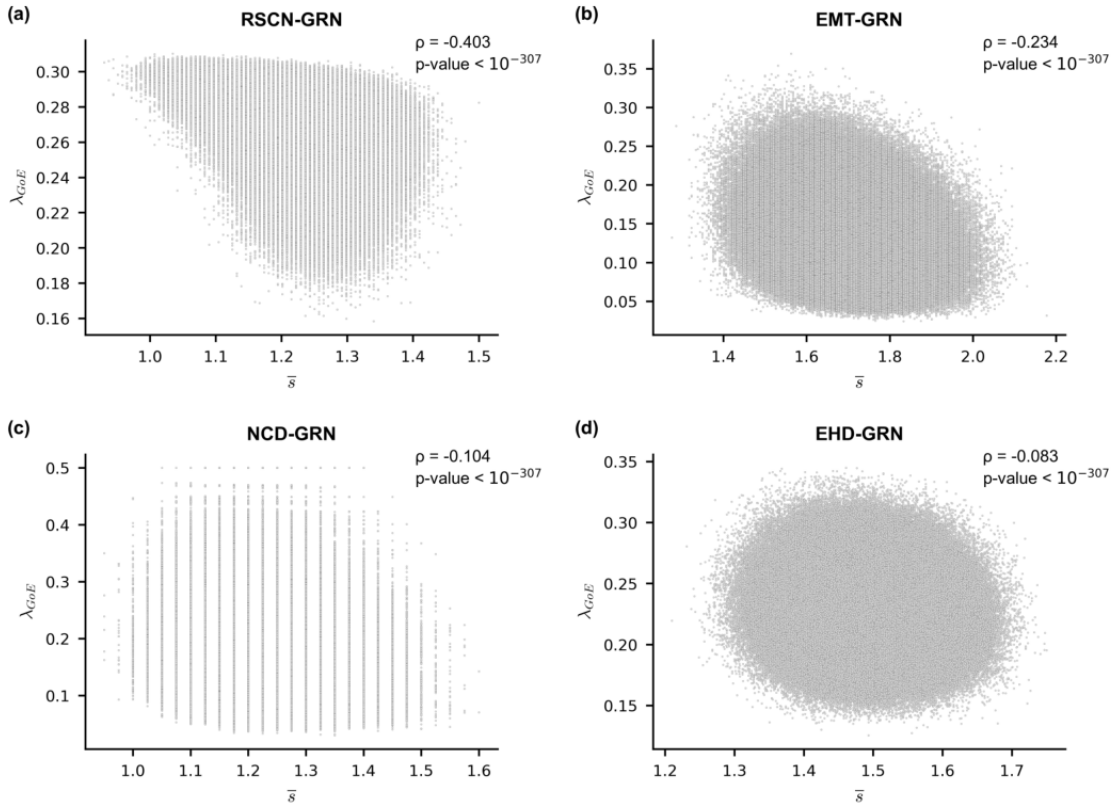


Figure B.42: Scatter plot between the network sensitivity ($\overline{s}$) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to the network sensitivity ($\overline{s}$) and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and λ_{GoE} show a moderate negative correlation for RSCN-GRN, a weak negative correlation for EMT-GRN and a very weak negative correlation otherwise.

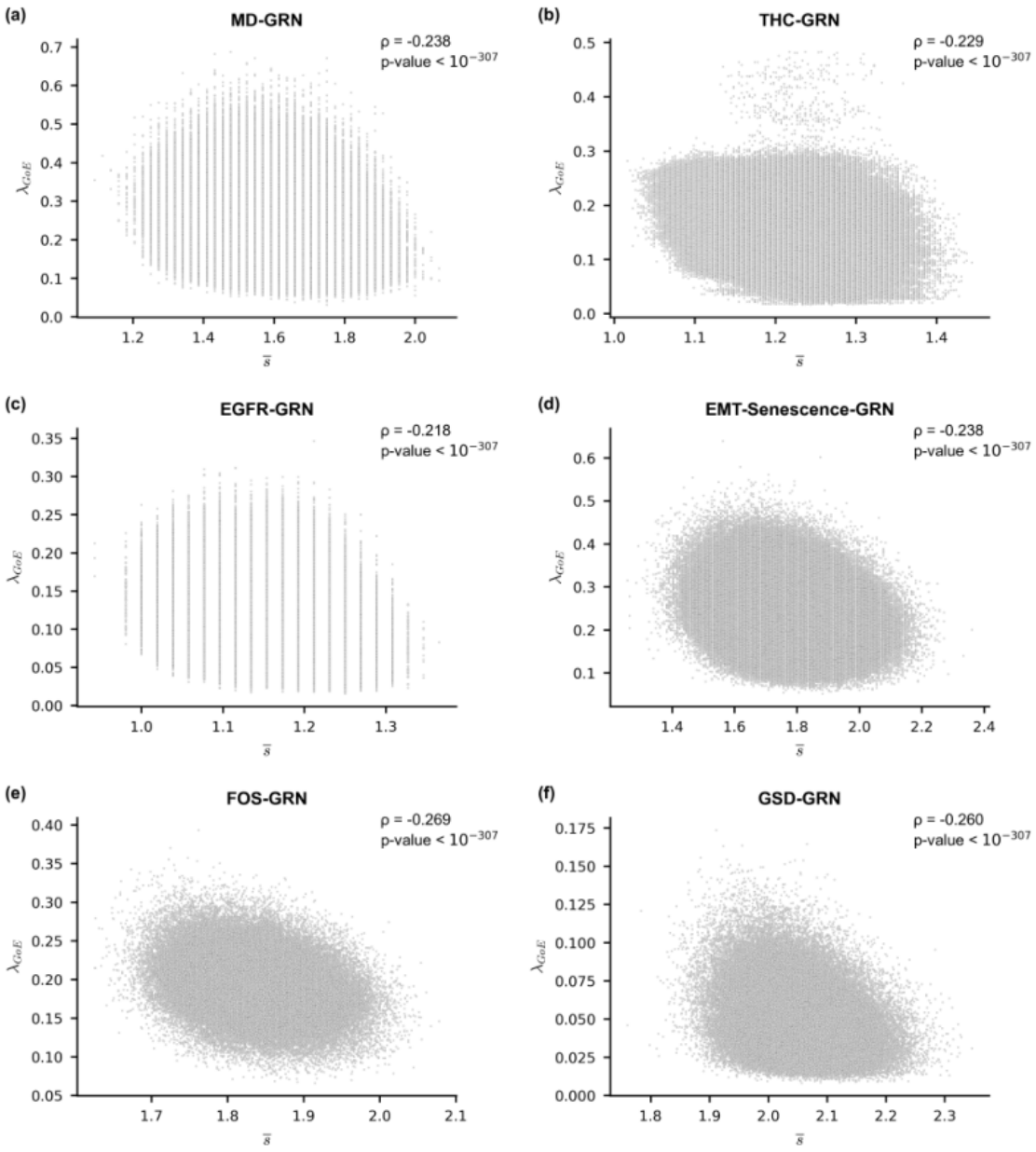


Figure B.43: Scatter plot between the network sensitivity ($\overline{s}$) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x and y axis of each sub-figure correspond to the network sensitivity ($\overline{s}$) and λ_{GoE} respectively. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. $\overline{s}$ and λ_{GoE} shows a weak negative correlation for all networks.

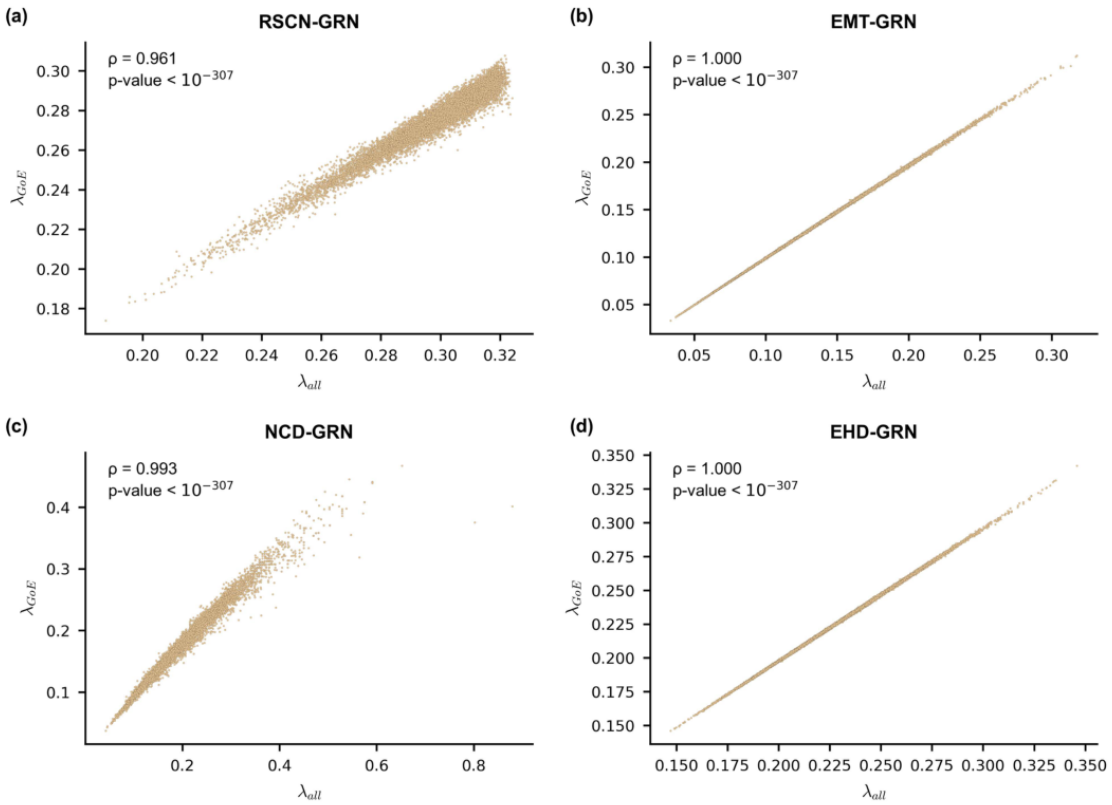


Figure B.44: Scatter plot between the average convergence rate of trajectories originating at *all* states (λ_{all}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x axis of each sub-figure is the average convergence rate of trajectories originating from all states of the STG (λ_{all}). The y axis of each sub-figure is the average convergence rate of trajectories originating only from the *GoE* states (λ_{GoE}). In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. λ_{all} and λ_{GoE} are very strongly correlated.

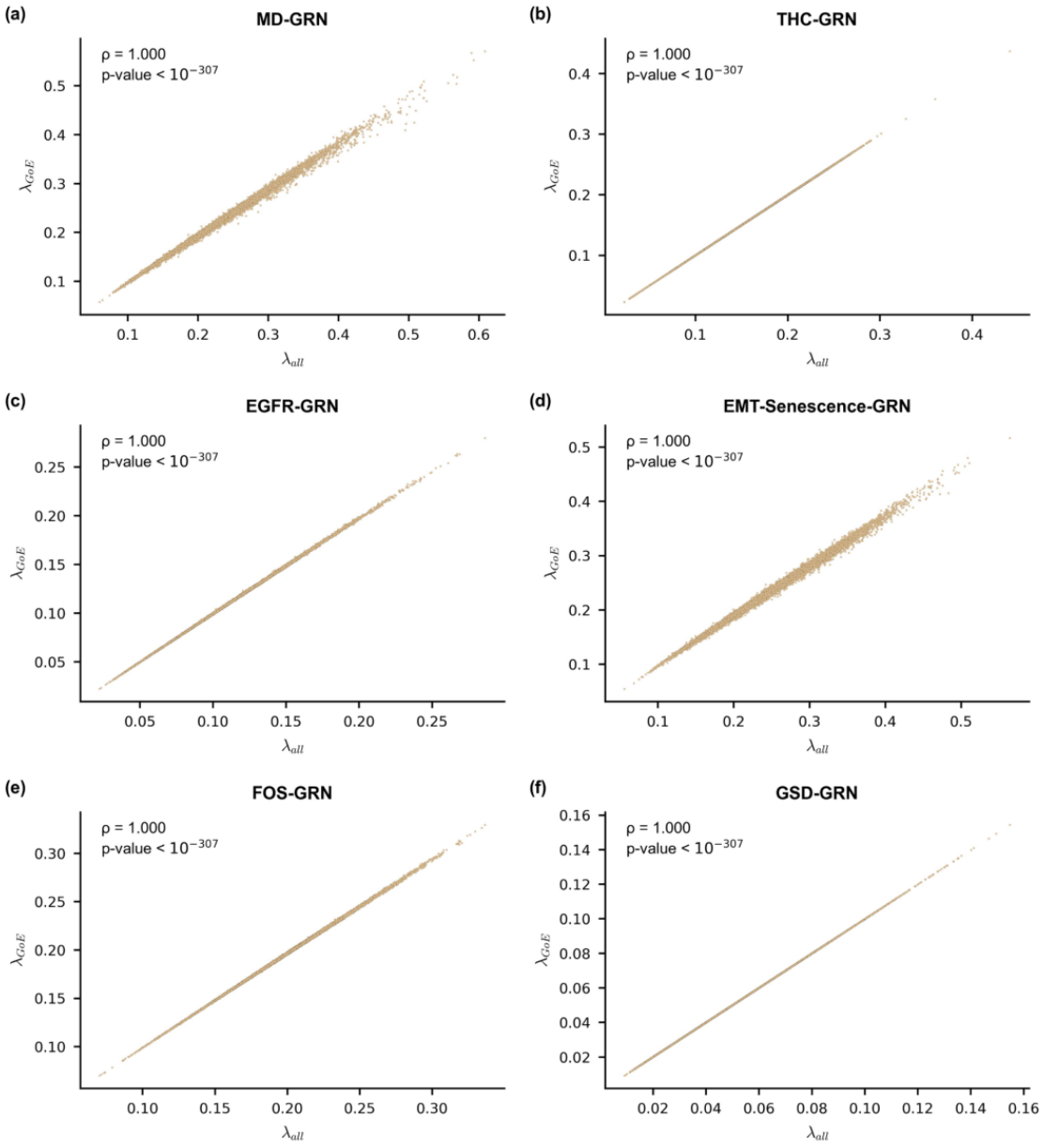


Figure B.45: Scatter plot between the average convergence rate of trajectories originating at *all* states (λ_{all}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x axis of each sub-figure is the average convergence rate of trajectories originating from all states of the STG (λ_{all}). The y axis of each sub-figure is the average convergence rate of trajectories originating only from the *GoE* states (λ_{GoE}). In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. λ_{all} and λ_{GoE} are very strongly correlated.

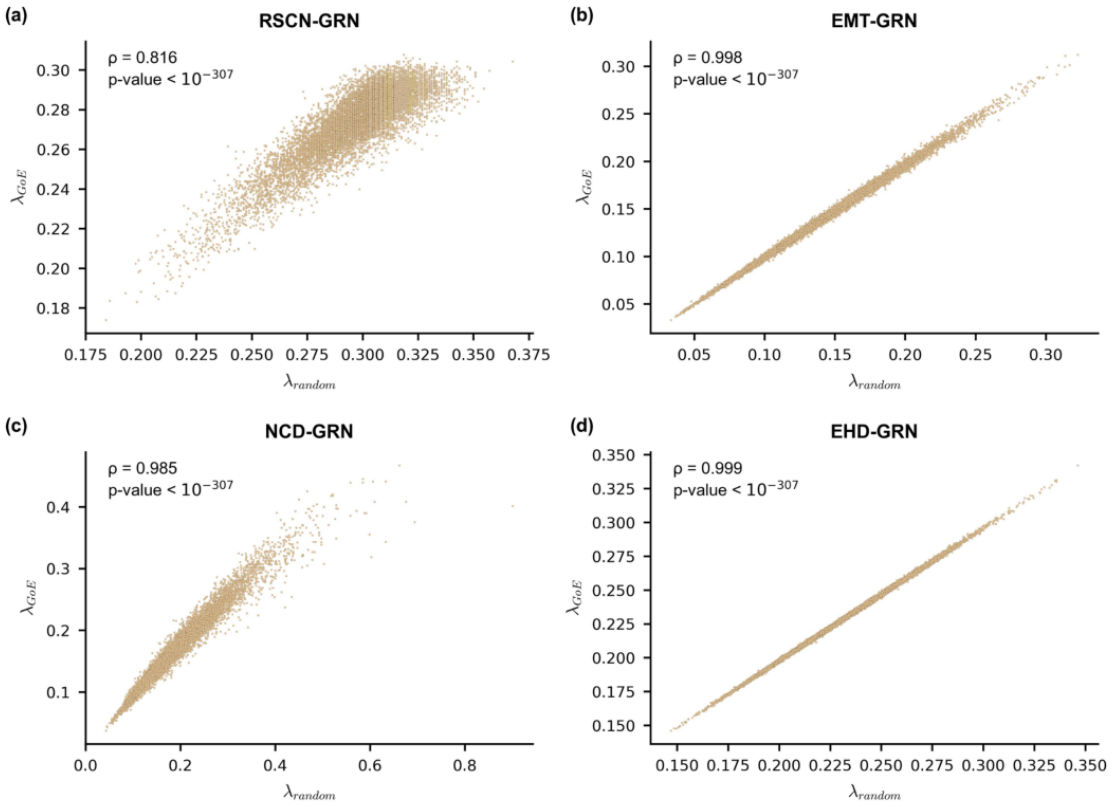


Figure B.46: Scatter plot between the average convergence rate of trajectories originating at *random* states (λ_{random}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from four published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)** and **(d)** correspond to the RSCN-GRN, EMT-GRN, NCD-GRN and EHD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x axis of each sub-figure is the average convergence rate of trajectories originating from random states of the STG (λ_{random}). The y axis of each sub-figure is the average convergence rate of trajectories originating only from the *GoE* states (λ_{GoE}). The number of random states for which trajectories were simulated were dependent on the size of the STG. For each model, the number of states chosen randomly is: 100 for RSCN-GRN, 1000 for EMT-GRN, 100 for NCD-GRN and 10000 for EHD-GRN. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. λ_{random} and λ_{GoE} are very strongly correlated.

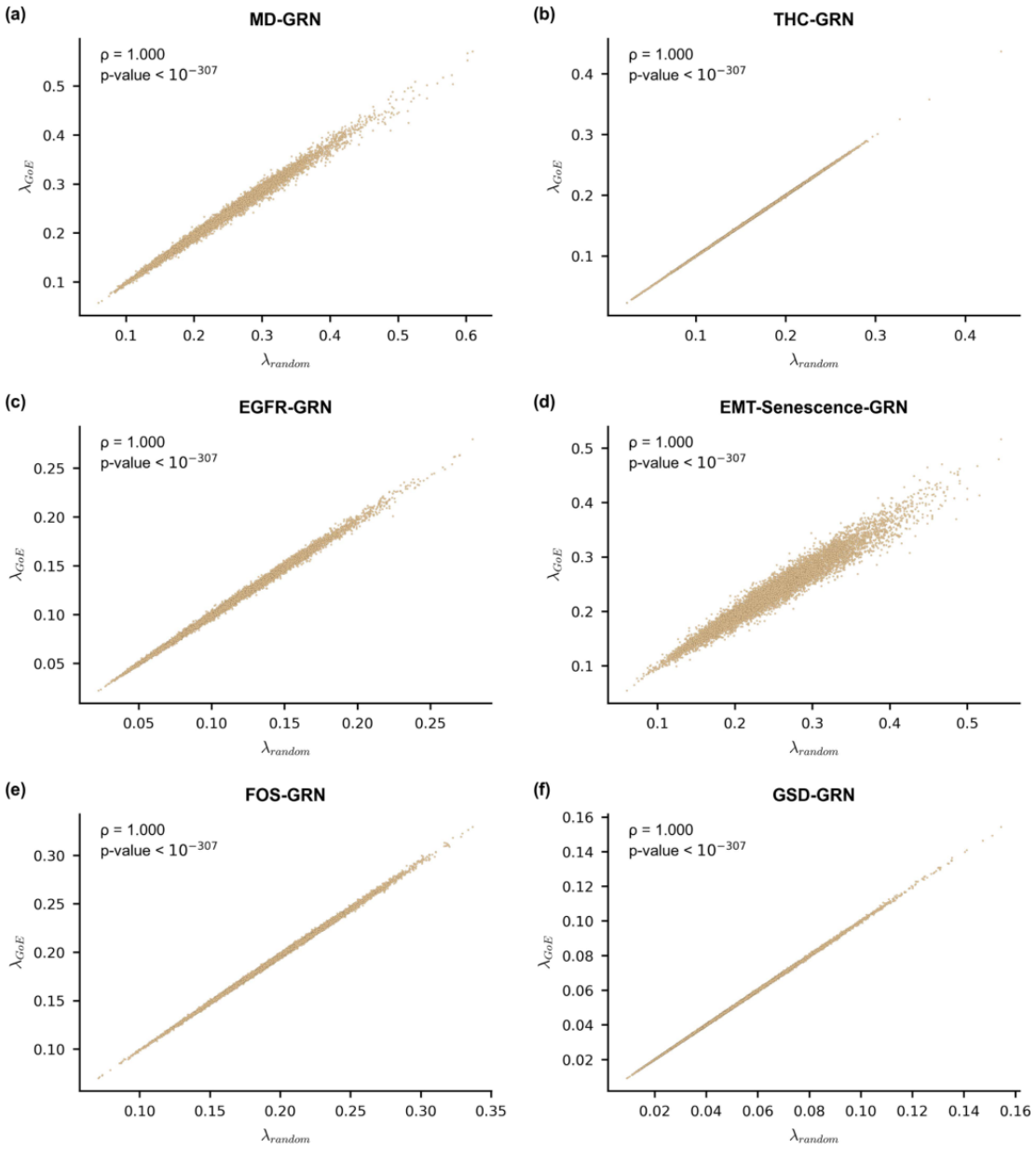


Figure B.47: Scatter plot between the average convergence rate of trajectories originating at *random* states (λ_{random}) and average convergence rate of trajectories originating at *GoE* states (λ_{GoE}) in ensembles generated using network structures from six published Boolean GRNs. The sub-figures **(a)**, **(b)**, **(c)**, **(d)**, **(e)** and **(f)** correspond to the MD-GRN, THC-GRN, EGFR-GRN, EMT-Senescence-GRN, FOS-GRN and GSD-GRN network structures respectively. The EF-ensemble for each network structure was used to generate the scatter plots. The x axis of each sub-figure is the average convergence rate of trajectories originating from random states of the STG (λ_{random}). The y axis of each sub-figure is the average convergence rate of trajectories originating only from the *GoE* states (λ_{GoE}). The number of random states for which trajectories were simulated were dependent on the size of the STG. For each model, the number of states chosen randomly is: 1000 for MD-GRN, 10000 for THC-GRN, 1000 for EGFR-GRN, 100 for EMT-Senescence-GRN, 10000 for FOS-GRN and 10000 for GSD-GRN. In each sub-figure we provide the Spearman correlation coefficient (ρ) along with the associated p-value. λ_{random} and λ_{GoE} are very strongly correlated.

Appendix C

Supplementary Material for Chapter 6

C.1 Properties of the IMR and BMR sets

Property C.1. For an IMR with k regulators, where k is odd, the BF equal to that IMR exhibits a bias of 2^{k-1} regardless of the sign combination of the regulators.

Proof. The bias of a BF is given by the number of input combinations that lead to 1 or 'on' as the output. Since, a BF with k inputs has 2^k different input combinations, it would suffice to prove that the BF will have equal number of $+1$ ('on') and -1 ('off') in the output. Assume that we have k inputs, with k odd and each input $x_j \in \{-1, +1\}$. Further, each input is assigned with a sign $\sigma_j \in \{-1, +1\}$ depending on its activatory or inhibitory characteristics. Then, the output of the BF is given by:

$$f(x_1, x_2, \ldots, x_k) = \begin{cases} +1, & \text{if } S > 0 \\ -1, & \text{if } S < 0 \end{cases}$$

where, $S = \sum_{j=1}^{k} \sigma_j x_j$.

Let us define a mapping, $\phi : \{-1, +1\}^k \to \{-1, +1\}^k$, $\phi(x_1, x_2, \ldots, x_k) = (-x_1, -x_2, \ldots, -x_k)$. Clearly, the map is a bijection since each element of $\{-1, +1\}^k$ is uniquely mapped to an element in $\{-1, +1\}^k$ and $\phi^2(x) = x \; \forall x \in \{-1, +1\}^k$.

For an element $x = (x_1, x_2, \ldots, x_k) \in \{-1, +1\}^k$, $S(x) = \sum_{j=1}^{k} \sigma_j x_j$. Now, $S(\phi(x)) = \sum_{j=1}^{k} \sigma_j(-x_j) = -\sum_{j=1}^{k} \sigma_j x_j = -S(x)$. So, whenever $S(x) > 0, S(\phi(x)) < 0$ and vice versa. Since, k is odd and each x_i can be either +1 or -1, S can never be zero. Thus, $\{-1, +1\}^k$ can be

split into two non-overlapping sets:

$$S_+ = \{x \in \{-1,+1\}^k \mid S(x) > 0\} \quad \text{and} \quad S_- = \{x \in \{-1,+1\}^k \mid S(x) < 0\}.$$

Now, since ϕ is a bijection, it maps each element in S_+ to a unique element in S_- and vice versa. So, the sets S_+ and S_- have same cardinality, and hence, the number of times $S > 0$ is $2^k/2 = 2^{k-1}$. Thus, the bias of the BF is 2^{k-1}. □

Property C.2. If an IMR rule is assigned to a node in a BN with k regulators, where k is even, then the bias of the EF equal to that rule depends on the existence of a self-regulation at that node and its sign: (a) If the node is not self-regulating (i.e., it is not among its own inputs), then for any given sign combination, the IMR corresponds to a $(k+1)$-input EF with a bias of 2^k. (b) If the node has a self-regulatory activatory edge, then the IMR corresponds to a k-input EF with a bias of 2^{k-1} for $k \geq 4$. (c) If the node has a self-regulatory inhibitory edge, then the IMR is equal to a $(k-1)$-input EF with bias of 2^{k-2}.

Proof. (a) If k is even, then there exists $x = (x_1, x_2, \ldots x_k)$ such that S (as defined in Property C.1) equals to 0 and in that case it will depend on its own state itself (say, x_{k+1}). Let us define the following sets:

$$\begin{aligned} A &= \{x \in \{-1,1\}^k \mid S(x) > 0\} \\ B &= \{x \in \{-1,1\}^k \mid S(x) < 0\} \\ C &= \{x \in \{-1,1\}^k \mid S(x) = 0\} \end{aligned}$$

If $|A| = a, |B| = b, |C| = c$ then $a + b + c = 2^k$. Each element of A, B and C will contribute two, zero and one 1's in the output, respectively. So, the bias of the corresponding BF is $2a + c$. Now, for any x,

$$S(-x) = \sum_{j=1}^{k} \sigma_j(-x_j) = -S(x)$$

This means if $S(x) > 0$, then $S(-x) < 0$ and vice versa. So, due to symmetry, $a = b$ (can be shown as in Property C.1) and hence, $c = 2^k - 2a$. So, the bias of the equivalent BF is

$2a + c = 2a + (2^k - 2a) = 2^k$.

(b) Now, suppose there exists a self-activatory edge. Without loss of generality, let us assume that x_k represents the node with the self-regulation with an activatory effect. In this case, when the signed sum $S = 0$, the output will depend on the value x_k. Since the self-regulation is activatory, the signed sum can be written as $S(x) = \sum_{j=1}^{k-1} \sigma_j x_j + \sigma_k x_k = T + x_k$. Since T is sum of an odd number of elements, namely $(k-1)$, where each element is either +1 or -1, the possible values that T can attain are: $-(k-1), -(k-3), \ldots, -1, 1, \ldots, (k-3), (k-1)$. Now, consider two input configurations: $a = (a_1, a_2, \ldots a_{k-1}, 1)$ and $a' = (a_1, a_2, \ldots a_{k-1}, -1)$ such that the sum $T = \sum_{j=1}^{k-1} \sigma_j a_j = 1$. For the configuration a, the total signed sum becomes

$$S(a) = T + x_k = 1 + 1 = 2 > 0$$

Hence, the output is 1. Again, for the configuration a', we have,

$$S(a') = T + x_k = 1 - 1 = 0$$

Since, the signed sum is 0, the output is determined by x_k, leading to an output -1. This demonstrates that the self-regulation is effective. Therefore, it is possible to map a k-input (with even k) IMR with activatory self-regulation to a k-input EF. This result holds for all even $k \geq 4$.

However, for $k = 2$, the presence of an activatory self-regulation leads to a k-input ineffective BF. Specifically, in this case the other input does not influence the output, i.e., if $x = (x_1, x_2)$ and x_2 represents the node with activatory self-regulation, then for both the sign combinations '++' and '-+', the output depends solely on x_2. The BF reduces to the EF $f(x_1, x_2) = x_2$, in both cases resulting in a bias equal to 1. Now, for any even $k \geq 4$, we aim to show that the bias of a BF equal to an IMR with an activatory self-regulation is exactly 2^{k-1}.

Suppose $k \geq 4$ and we consider $S(x) = T + x_k$. Now, T can take the following values: $-(k-1), -(k-3), \ldots, -1, 1, \ldots, (k-3), (k-1)$. Suppose N_t denotes the count of x for which T

can be equal to t. Then, by symmetry $N_t = N_{-t}$ and hence,

$$\sum_{t=-(k-1)}^{-1} N_t = \sum_{t=1}^{(k-1)} N_t = 2^{k-1} \text{ since } \sum_{t=-(k-1)}^{(k-1)} N_t = 2^k$$

For $T \leq -3$, $S < 0$ for both values of x_k. So, for all x satisfying $T \leq -3$, the output is -1. Again, for $T \geq 3$, $S > 0$ for both values of x_k. So, for all x satisfying $T \geq 3$, the output is 1. Now, when $T = 1$, if $x_k = 1$, then the output becomes 1, but if $x_k = -1$, then the output is -1. Which means when $T = 1$, half of the times we get 1 as output. Again, when $T = -1$, if $x_k = 1$, the output becomes 1 as $S = 0$ and $x_k = 1$. If $x_k = -1$ then the output is -1. So, again we see that when $T = -1$, half of the times we get 1 as output. Thus, the bias of the equivalent BF would be equal to

$$\sum_{t=3}^{(k-1)} N_t + \frac{1}{2}(N_1 + N_{-1}) = \sum_{t=1}^{(k-1)} N_t = 2^{k-1} \text{ since } N_1 = N_{-1}.$$

(c) Now consider the case when the self-regulation is inhibitory. Let the self-regulation be on the node x_k; then the total signed sum becomes $S(x) = T - x_k$ where $T = \sum_{j=1}^{(k-1)} \sigma_j x_j$. As discussed in (b), for even k, T can take the odd integer values in the range $[-(k-1), (k-1)]$. Now, we analyze the output behaviour for different values of T.

- $T \leq -3$: For both $x_k = \pm 1$, we have $S < 0$, so the output is -1.
- $T \geq 3$: For both $x_k = \pm 1$, we have $S > 0$, so the output is 1.
- $T = 1$:
  - If $x_k = 1$, then $S = 0$, so the output is 1 (since $x_k = 1$).
  - If $x_k = -1$, then $S = 2$, so the output is also 1.

  Thus, the output is 1 regardless of the value of x_k.
- $T = -1$:
  - If $x_k = 1$, then $S = -2$, so the output is -1.
  - If $x_k = -1$, then $S = 0$, so the output is -1 (since $x_k = -1$).

  Thus, the output is -1 regardless of the value of x_k.

This analysis shows that for $T = \pm 1$, the output is independent of x_k, implying that the inhibitory self-regulation becomes ineffective. Thus, it can be shown that the bias will be 2^{k-1} using the

same approach as in (b). However, since it can be reduced to a $(k-1)$-input EF by removing the ineffective self-regulation, the bias will be $\frac{2^{k-1}}{2} = 2^{k-2}$. □

Property C.3. A BF equal to an IMR with k regulators can never be canalyzing for any $k > 2$, regardless of the presence or absence of a self-regulation.

Proof. The IMR rule can be written as:

$$f(x_1, \ldots, x_k, x_t) = \begin{cases} +1, & \sum_{i=1}^{k} \sigma_i x_i > 0, \\ x_t, & \sum_{i=1}^{k} \sigma_i x_i = 0, \\ -1, & \sum_{i=1}^{k} \sigma_i x_i < 0, \end{cases}$$

where $x_i \in \{-1, +1\}$, and x_t denotes an additional variable. When k is even and there is no self-regulation, x_t acts as an external variable (representing an implicit self-regulation). If k is even and one of the k regulators is already a self-regulator, x_t is considered as an internal variable which is the self-regulator. For odd k, x_t can be treated either as an internal variable or as an external ineffective variable, since the signed sum is never zero when k is odd.

We will show that for every $1 \leq r \leq t$ and every fixed value $a \in \{+1, -1\}$, there always exists two assignments of the other variables under which f takes both values $+1$ and -1. We denote $\alpha_i = sgn(\sigma_i)$ for all i then $\sigma_i \cdot \alpha_i = 1; \sigma_i \cdot (-\alpha_i) = -1$ for all i.

Case 1: Suppose x_t is an internal variable. Suppose x_r is a canalyzing variable with $x_r = a$.

Input 1: $x_r = a, x_i = \alpha_i \; \forall \; i \neq r$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i \alpha_i = \sigma_r a + (k-1) \geq -1 + (k-1) > 0 \text{ since } k > 2$$

So, the output is $+1$ in this case.

Input 2: $x_r = a, x_i = -\alpha_i \; \forall \; i \neq r$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i (-\alpha_i) = \sigma_r a - (k-1) \leq 1 - (k-1) < 0 \text{ since } k > 2$$

So, the output is -1 in this case. So, none of the variables is a canalyzing variable.

Case 2: Suppose x_t is an external variable and x_r is a canalyzing variable with $x_r = a$. If $r \neq t$ then we can choose the same pair of inputs as in Case 1 to show that x_r cannot be canalyzing. If $r = t$ then we choose the following two inputs:

Input 1: $x_{r=t} = a, x_i = \alpha_i \; \forall \; i \neq t$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sum_{i=1}^{k} \sigma_i \alpha_i = \sum_{i=1}^{k} (+1) = k > 0$$

So, the output is $+1$ in this case.

Input 2: $x_{r=t} = a, x_i = -\alpha_i \; \forall \; i \neq t$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sum_{i=1}^{k} \sigma_i (-\alpha_i) = \sum_{i=1}^{k} (-1) = -k < 0$$

So, the output is -1 in this case. So, none of the variables is a canalyzing variable.

This completes the proof that a BF which is equal to an IMR with k regulators can never be a canalyzing function if $k > 2$ irrespective of whether there exists a self-regulation or not. □

Property C.4. Consider a BF f that corresponds to a BMR with k-regulators, one of which is a self-regulation. Then: (a) for $k > 2$, if the self-regulation is activatory, f is effective; (b) for $k > 1$, if the self-regulation is inhibitory but not all other regulators are inhibitory, f is ineffective, with the self-input being the ineffective input; (c) for $k > 0$, if all k regulators are inhibitory, f reduces to the constant 0 function; and (d) for $k = 1$ or 2, the function f is effective only when all regulators are activatory.

Proof. (a) Consider a BMR with k regulators $x_1, x_2, \ldots x_k$. Without loss of generality let us assume x_k is the target node, and hence, is the self-regulatory input. If we denote the signed sum by $S = \sum_{i=1}^{k} \sigma_i x_i$, then the function f is defined by

$$f(x_1, x_2, \ldots, x_k) = \begin{cases} 1, & \text{if } S > 0 \\ 0, & \text{if } S < 0 \\ x_k, & \text{if } S = 0 \end{cases}$$

First consider $x_i = 0$ for all $i < k$. Then toggling x_k from 0 to 1 changes the output from 0 to 1. So, the input x_k is effective. Now we want to show that all x_j (for $1 \leq j < k$) are effective.

Since $k > 2$, $\exists\ m < k$ such that $m \notin \{j, k\}$. If all $x_i = 0\ \forall\ i \neq m, j, k$ then the signed sum becomes $S = \sigma_j x_j + \sigma_m x_m + x_k$. So, the pair (σ_j, σ_m) can be equal to any tuple from the set $\{(1,1),(1,-1),(-1,1),(-1,-1)\}$. If (σ_j, σ_m) is equal to $(1,1)$ or $(1,-1)$ then fixing $x_m = x_k = 0$ and toggling x_j from 0 to 1 changes the output from 0 to 1. If (σ_j, σ_m) is equal to $(-1,1)$ then fixing $x_m = 1, x_k = 0$ and toggling x_j from 0 to 1 changes the output from 1 to 0. If (σ_j, σ_m) is equal to $(-1,-1)$ then fixing $x_m = 0, x_k = 1$ and toggling x_j from 0 to 1 changes the output from 1 to 0. This proves that f is effective when the self-regulation is activatory.

(b) Suppose the self-regulator is inhibitory then $\sigma_k = -1$. Let us denote $T = \sum_{i=1}^{k-1} \sigma_i x_i$. Then, $S = T - x_k$.

If $x_k = 0$,

$$f(x_1, x_2, \ldots, 0) = \begin{cases} 1, & \text{if } T > 0 \\ 0, & \text{if } T \leq 0 \end{cases}$$

If $x_k = 1$,

$$f(x_1, x_2, \ldots, 1) = \begin{cases} 1, & \text{if } T - 1 > 0 \\ 1, & \text{if } T - 1 = 0 \\ 0, & \text{if } T - 1 < 0 \end{cases}$$

$$= \begin{cases} 1, & \text{if } T > 0 \\ 0, & \text{if } T \leq 0 \end{cases}$$

So, the output depends only on the signed sum of first $(k-1)$ variables i.e., T. Thus, the self-regulation becomes ineffective.

Note that using the same approach as in (a), it can be shown that for all $1 \leq j < k$ $(k > 2)$ whenever $(\sigma_j, \sigma_k) \in \{(1,1),(1,-1),(-1,1)\}$, with a suitable chosen value of x_m, and $x_i = 0\ \forall i \neq k, m, j$, toggling the value of x_j from 0 to 1 will change the output. So, in all such cases x_j will be an effective input for all $1 \leq j < k$. Now, if for all $j, m < k$, $(\sigma_j, \sigma_k) = (-1,-1)$, then we can prove that none of the inputs is effective and f is the constant zero function. See (c) below for the explanation.

(c) If all the regulators are inhibitory, $\sigma_i = -1 \; \forall \; 1 \leq i \leq k$. Then, the signed sum becomes,

$$S = -\sum_{i=1}^{k} x_i$$

If $x_i = 0 \; \forall \; 1 \leq i \leq k$, then $S = 0$. So, the output is 0 because x_k is 0. Now, if there exists at least one j $(1 \leq j \leq k)$ such that $x_j = 1$, then $S < 0$. So, the output is 0. Therefore, f is the constant 0 function and no input affects the output.

(d) Consider a BMR with one input (which here is obviously the self-input). If the self-regulation is inhibitory then from (c) it follows that f is ineffective. If the self-regulation is activatory then f is the identity function, i.e., $f(x_1) = x_1$ and hence effective. Now, consider a BMR with two inputs x_1, x_2 and for which x_2 is the self-regulator. Now, if $\sigma_2 = -1$ then using (b), f will be ineffective. If $(\sigma_1, \sigma_2) = (-1, 1)$, $f(x_1, x_2) = x_2$, so f is ineffective. If $(\sigma_1, \sigma_2) = (1, 1)$, $f(x_1, x_2) = x_1 \vee x_2$. This is the only case for which f is effective. Thus, for $k = 1, 2$, f will be effective only when all the regulators are activatory. □

Property C.5. The bias of a BMR having k inputs, all external, is given by the following expression:

$$P^k_{BMR} = 2 \cdot \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k}{a}$$

where a is the number of activators among those k regulators.

Proof. A BMR with k external regulators can be mapped to a $(k+1)$-input BF. a is the number of activators. Then there are $(k-a)$ inhibitors. Suppose i and j denote the number of active activators and active inhibitors, respectively. Then, the conditions $i > j, i = j$ and $i < j$, will contribute two, one and zero 1's, respectively in the list of outputs. Thus, the bias of the BMR is given by

$$P^k_{BMR} = \sum_{i=0}^{a} \sum_{j=0}^{k-a} \binom{a}{i} \binom{k-a}{j} \cdot w(i, j), \tag{C.1}$$

where

$$w(i,j) = \begin{cases} 2 & \text{if } i > j, \\ 1 & \text{if } i = j, \\ 0 & \text{if } i < j. \end{cases}$$

So, we need to evaluate the double sum when $i > j$ and when $i = j$.

Case 1 ($i > j$): We need to determine the value of

$$P_{pos} = \sum_{\substack{i>j \\ 0\le i\le a \\ 0\le j\le k-a}} \binom{a}{i}\binom{k-a}{j}$$

Consider

$$G(x) = \sum_{i=0}^{a}\sum_{j=0}^{k-a}\binom{a}{i}\binom{k-a}{j}x^{i-j} \tag{C.2}$$

Then $G(x)$ can be written as

$$\begin{aligned} G(x) = \left(\sum_{i=0}^{a}\binom{a}{i}x^i\right)\cdot\left(\sum_{j=0}^{k-a}\binom{k-a}{j}x^{-j}\right) &= (1+x)^a\cdot(1+x^{-1})^{k-a} \\ &= x^{-(k-a)}(1+x)^k \\ &= x^{(a-k)}\sum_{s=0}^{k}\binom{k}{s}x^s \\ &= \sum_{s=0}^{k}\binom{k}{s}x^{s+a-k} \\ &= \sum_{t=-(k-a)}^{a}\binom{k}{t+k-a}x^t \end{aligned}$$

From Eq. (C.2), the coefficient of x^t in $G(x)$ is

$$C_t = \sum_{\substack{i-j=t \\ 0\le i\le a \\ 0\le j\le k-a}} \binom{a}{i}\binom{k-a}{j}$$

Since $i > j$ if and only if $t \ge 1$, by equating the sum of coefficients of the positive powers of

x, we have

$$\sum_{\substack{i>j \\ 0\le i\le a \\ 0\le j\le k-a}} \binom{a}{i}\binom{k-a}{j} = \sum_{t=1}^{a}\binom{k}{t+k-a}$$
$$= \sum_{r=k-a+1}^{k}\binom{k}{r}$$

Finally, using the binomial symmetry $\binom{k}{r} = \binom{k}{k-r}$, we have

$$P_{pos} = \sum_{\substack{i>j \\ 0\le i\le a \\ 0\le j\le k-a}} \binom{a}{i}\binom{k-a}{j} = \sum_{m=0}^{a-1}\binom{k}{m} \tag{C.3}$$

Case 2 ($i = j$): Now, we need to determine the value of

$$P_{null} = \sum_{\substack{i=j \\ 0\le i\le a \\ 0\le j\le k-a}} \binom{a}{i}\binom{k-a}{j}$$

Suppose $i = j = d$. Then

$$P_{null} = \sum_{d=0}^{min(a,\, k-a)} \binom{a}{d}\binom{k-a}{d}$$

If $min(a, k-a) = a$,

$$P_{null} = \sum_{d=0}^{a}\binom{a}{d}\binom{k-a}{d} = \sum_{d=0}^{a}\binom{a}{a-d}\binom{k-a}{d}$$

Using Vandermonde's convolution, we get,

$$P_{null} = \binom{a+(k-a)}{a} = \binom{k}{a}$$

Similarly, if $min(a, k-a) = k-a$,

$$P_{null} = \sum_{d=0}^{k-a}\binom{a}{d}\binom{k-a}{d} = \sum_{d=0}^{a}\binom{a}{d}\binom{k-a}{(k-a)-d}$$

Again, using Vandermonde's convolution,

$$P_{null} = \binom{a + (k-a)}{k-a} = \binom{k}{k-a} = \binom{k}{a}$$

So, irrespective of $min(a, k-a) = k-a$ or a, the value of P_{null} is given by

$$P_{null} = \binom{k}{a} \tag{C.4}$$

Using Eqs. (C.1), (C.3) and (C.4), we have

$$P^k_{BMR} = 2 \cdot P_{pos} + P_{null} = 2 \cdot \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k}{a}$$

□

Corollary C.5.1. Suppose the BF f corresponds to a BMR with k external regulators, out of which a are activatory. Then, (a) f is an NCF if and only if all the regulators are inhibitory or all are activatory; (b) f is of odd bias if and only if $a \wedge \bar{k} = 0$, where $a \wedge \bar{k}$ indicates the bitwise AND operation of a and $\bar{k}$; (c) All such f will be of odd bias if and only if k is a Mersenne number.

Proof. (a) If all the regulators are inhibitory then $a = 0$. So, using Property C.5

$$P^k_{BMR} = \binom{k}{0} = 1$$

Since, the bias is 1, it indicates an AND function, and hence, an NCF.

Now, if all the regulators are activatory, then $a = k$. Using Property C.5

$$P^k_{BMR} = 2 \cdot \sum_{m=0}^{k-1} \binom{k}{m} + \binom{k}{k}$$

Now, using the identity $\sum_{m=0}^{k} \binom{k}{m} = 2^k$ we have,

$$\begin{aligned} P^k_{BMR} &= 2 \cdot \left[2^k - \binom{k}{k}\right] + \binom{k}{k} \\ &= 2^{k+1} - 1 \end{aligned}$$

Since, the bias is $2^{k+1} - 1$, all except one of the output values are equal to 1, which indicates

it to be an OR function and hence an NCF.

Lastly, in the mixed case, f cannot even be canalyzing (see Property C.7) and consequently it is not an NCF.

(b) From Property C.5, it follows that the bias can be odd if and only if $\binom{k}{a}$ is odd. As a consequence of Lucas's theorem [199] we know that $\binom{k}{a}$ is odd if and only if every binary bit of a is less than or equal to the corresponding bit of k [200]. In other words, $\binom{k}{a}$ is odd if and only if $a \wedge \overline{k} = 0$, where $a \wedge \overline{k}$ indicates the bitwise AND operation of a and $\overline{k}$ (i.e., writing a and k as binary strings of length k, one performs the bitwise AND operation between a and $\overline{k}$, and then, one converts the output to integer and checks if it is zero) [200].

(c) For a given k, all the BFs will be of odd bias if and only if for all $0 \leq a \leq k$, $\binom{k}{a}$ is odd. Now, from (b) it follows that it is possible only when all the binary bits of k are 1, and that is possible only if k is of the form $2^r - 1$ for some positive integer r, i.e., only when k is a Mersenne number. □

Property C.6. The bias of a k-input BF f corresponding to a BMR with k regulators – one of which is a self-regulation – is given by the following expression:

$$P^k_{BMR} = \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k-1}{a-\delta}$$

where a is the number of activators among those k regulators, $\delta = 0$ if the self-regulation is activatory and $\delta = 1$ if the self-regulation is inhibitory.

Proof. Suppose the k regulators of a node x_r are $x_1, \ldots, x_{r-1}, x_r, x_{r+1}, \ldots, x_k$. The BMR at node x_r can thus be represented as a BF with k inputs. The output of this function is 1 under the following two conditions: (a) the signed sum of the inputs is positive; (b) the signed sum is 0 and $x_r = 1$. Each instance in which either of these conditions hold contributes a 1 to the output of the function. Therefore,

$$P^k_{BMR} = P_{pos} + P_{null}(x_r = 1)$$

The value of P_{pos} remains identical to that derived in Property C.5:

$$P_{pos} = \sum_{m=0}^{a-1} \binom{k}{m} \tag{C.5}$$

Now we need to evaluate the value of $P_{null}(x_r = 1)$, i.e., the number of input combinations for which the signed sum is 0 and $x_r = 1$. Suppose, a, i, j denote the number of activators, number of active activators and number of active inhibitors, respectively. The signed sum is 0 when $i = j$, leading to

$$P_{null} = \sum_{j=0}^{min(a,k-a)} \binom{a}{j}\binom{k-a}{j}$$

Among these we need to count only the input combinations where $x_r = 1$. Since the self-regulation at x_r can be activatory or inhibitory in nature, we consider these two cases separately.

Case 1 (x_r is activatory): Since, $x_r = 1$ and it is self-activatory, we need to choose $(j-1)$ other active activators from the remaining $(a-1)$ activator slots. So, in this case,

$$P_{null}(x_r = 1) = \sum_{j=0}^{min(a,k-a)} \binom{a-1}{j-1}\binom{k-a}{j} \tag{C.6}$$

If $min(a, k-a) = a$,

$$\begin{aligned} P_{null}(x_r = 1) &= \sum_{j=0}^{a} \binom{a-1}{j-1}\binom{k-a}{j} \\ &= \sum_{j=0}^{a} \binom{a-1}{a-j}\binom{k-a}{j} \\ &= \binom{k-1}{a} \text{ (using Vandermonde's convolution)} \end{aligned}$$

If $min(a, k-a) = k-a$,

$$\begin{aligned} P_{null}(x_r = 1) &= \sum_{j=0}^{k-a} \binom{a-1}{j-1}\binom{k-a}{j} \\ &= \sum_{i=-1}^{k-a-1} \binom{a-1}{i}\binom{k-a}{i+1} \\ &= \sum_{i=0}^{k-a-1} \binom{a-1}{i}\binom{k-a}{(k-a-1)-i} \\ &= \binom{k-1}{k-a-1} \text{ (using Vandermonde's convolution)} \\ &= \binom{k-1}{a} \end{aligned}$$

Thus, when x_r is activatory,

$$P_{null}(x_r = 1) = \binom{k-1}{a}$$

Case 2 (x_r is inhibitory): Since, $x_r = 1$ and it is self-inhibitory, we need to choose $(j-1)$ other active inhibitors from the remaining $(k-a-1)$ inhibitor slots. So, in this case,

$$P_{null}(x_r = 1) = \sum_{j=0}^{min(a,k-a)} \binom{a}{j}\binom{k-a-1}{j-1} \tag{C.7}$$

If $min(a, k-a) = a$,

$$\begin{aligned} P_{null}(x_r = 1) &= \sum_{j=0}^{a} \binom{a}{j}\binom{k-a-1}{j-1} \\ &= \sum_{i=-1}^{a-1} \binom{a}{i+1}\binom{k-a-1}{i} \\ &= \sum_{i=0}^{a-1} \binom{a}{(a-1)-i}\binom{k-a-1}{i} \\ &= \binom{k-1}{a-1} \text{ (using Vandermonde's convolution)} \end{aligned}$$

If $min(a, k-a) = k-a$,

$$\begin{aligned} P_{null}(x_r = 1) &= \sum_{j=0}^{k-a} \binom{a}{j}\binom{k-a-1}{j-1} \\ &= \sum_{j=0}^{k-a} \binom{a}{j}\binom{k-a-1}{(k-a)-j} \\ &= \binom{k-1}{k-a} \text{ (using Vandermonde's convolution)} \\ &= \binom{k-1}{a-1} \end{aligned}$$

Thus, when x_r is inhibitory,

$$P_{null}(x_r = 1) = \binom{k-1}{a-1}$$

A single formula can then summarize the two cases:

$$P^k_{BMR} = \sum_{m=0}^{a-1} \binom{k}{m} + \binom{k-1}{a-\delta}$$

where $\delta = 0$ if x_r is activatory and $\delta = 1$ if x_r is inhibitory. If all the regulations including the self-regulation are inhibitory, the BMR corresponds to a constant-zero BF (The formula correctly yields a bias of 0, using the convention $\binom{\alpha}{\beta} = 0$ if $\beta < 0$). For $k > 2$, if the self-regulation is inhibitory and if there exists at least one activator among the other regulators, then x_r is the only ineffective input to f (see Property C.4). For $k = 1$ and 2, the BF corresponding to a BMR with self-regulation is effective if and only if all regulators are activatory. When f is ineffective, after converting the f to an EF, the bias will be $\frac{1}{2}P^k_{BMR}$. □

Property C.7. A BF which is equal to a BMR with k regulators, all being external, is canalyzing if and only if all k regulators are either activatory or all are inhibitory.

Proof. The output of a BF equal to a BMR is determined by the values of the signed sum $S = \sum_{j=1}^{k} \sigma_j x_j$ and x_{k+1} ($k+1$ being the target node) using the following rule:

$$f(x_1, x_2, \ldots, x_k, x_{k+1}) = \begin{cases} 1, & \text{if } S > 0 \\ 0, & \text{if } S < 0 \\ x_{k+1}, & \text{if } S = 0 \end{cases}$$

We first prove that if all the regulators are either activatory or inhibitory, then the corresponding BF is canalyzing. If all the regulators are activatory (i.e., $\sigma_i = +1 \; \forall \; 1 \leq i \leq k$), then the signed sum is given by $S = \sum_{j=1}^{k} x_j$. In this case, if $x_1 = 1$ then regardless of the values of remaining inputs, we have $S > 0$, and hence the output is 1. Therefore, x_1 acts as a canalyzing variable with canalyzing input value 1 and canalyzed output value 1. Similarly, when all the regulators are inhibitory, setting any $x_i = 1$, regardless of the other inputs the signed sum is negative and hence the output is 0. Thus, each x_i for $1 \leq i \leq k+1$ is a canalyzing input with canalyzing input value 1 and canalyzed output value 0.

Now we will show that if not all the regulators are of same type then the BF cannot be canalyzing. Suppose, f is canalyzing but not all regulator signs are of the same type. So, $\exists \; p, q \; (p \neq q)$ such that $\sigma_p = +1$ and $\sigma_q = -1$. Since, f is canalyzing then there must be some variable x_r and $a \in \{0, 1\}$ such that $f(x_1, \ldots, x_{r-1}, a, x_{r+1}, \ldots x_{k+1}) = c$ for some fixed value of $c \in \{0, 1\}$. Now, let us consider different cases separately by fixing $x_r = a$ and comparing the outcomes for two distinct input configurations.

Case 1: $\sigma_r = +1$

Subcase 1.1: $a = 0$

Input 1: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 0$. Then, the output is 0.

Input 2: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 1$. Then, the output is 1.

Subcase 1.2: $a = 1$

Input 1: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 0$. Then, the output is 1.

Input 2: $x_q = 1$, $x_i = 0 \; \forall \; i \neq q, r$ and $x_{k+1} = 0$. Then, the output is 0.

Case 2: $\sigma_r = -1$

Subcase 2.1: $a = 0$

Input 1: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 0$. Then, the output is 0.

Input 2: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 1$. Then, the output is 1.

Subcase 2.2: $a = 1$

Input 1: $x_i = 0 \; \forall \; i \neq r$ and $x_{k+1} = 0$. Then, the output is 0.

Input 2: $x_p = 1$, $x_i = 0 \; \forall \; i \neq p, r$ and $x_{k+1} = 1$. Then, the output is 1.

Thus, we observe that in every possible condition, a contradiction arises, implying that none of the variables x_i for $1 \leq i \leq k$ can serve as a canalyzing variable. It is also straightforward to verify that x_{k+1} cannot be canalyzing either. To illustrate, consider the following two input configurations: (a) $x_p = 1$ and $x_i = 0$ for all $i \neq p$, and (b) $x_q = 1$ and $x_i = 0$ for all $i \neq q$. For any fixed value of x_{k+1}, the first input configurations lead to the output 1 whereas the second input leads to the output 0. This demonstrates that a BF equal to a BMR with k external regulators (i.e., without a self-regulation) is canalyzing if and only if all k regulators are either activatory or all are inhibitory. □

Property C.8. A BF that is equal to a BMR with k regulators including a self-regulation is canalyzing if and only if at most one regulator is activatory or at most one regulator is inhibitory.

Proof. We first show that if there is at most one inhibitory regulator or at most one activatory regulator, then the BF f is canalyzing. Let k regulators be $x_1, x_2, \ldots, x_k$. Without loss of generality, assume that x_k is the self-regulatory input.

Case 1 (At most one activator): If there is no activator then f is constant zero function (see Property C.4(c)) and hence canalyzing. When there is exactly one activator x_r, consider the case $x_r = 0$ as it will be canalyzing. Indeed,

$$S = \sum_{i \neq r} \sigma_i x_i = -\sum_{i \neq r} x_i \leq 0$$

Clearly, if $S < 0$ the output is 0; if instead $S = 0$, then necessarily $x_i = 0$ for all i and in particular $x_k = 0$, so the output is also 0. Since all cases for which $x_r = 0$ lead to a output of 0, x_r is a canalyzing variable.

Case 2 (At most one inhibitor): If there is no inhibitor then the signed sum is $\sum_{i=1}^{k} x_i$. So if any $x_i = 1$, the signed sum is positive and hence the output is fixed to 1. If there is exactly one inhibitor x_r and $r \neq k$, then if we fix $x_k = 1$, the signed sum becomes

$$S = -x_r + \sum_{i \neq r,k} x_i + 1 \geq 0$$

in which case the output is always 1 since $x_k = 1$. Now, if instead $k = r$, then x_k is ineffective (see Property C.4(b)) and the signed sum becomes $S = \sum_{i=1}^{k-1} x_i$. Hence for any x_i $(1 \leq i < k)$ fixed to 1, the signed sum is positive and hence the output is fixed to 1.

We have shown that if there is at most one activator or at most one inhibitor among the regulators, the BF f equal to the BMR is canalyzing. We now prove the converse. Specifically, we show that if there exist at least two activatory and at least two inhibitory regulators, then f cannot be canalyzing. Suppose x_t is a canalyzing variable for f with the canalyzing input value $a \in \{0, 1\}$ where $t < k$. Since, there exists at least two activators and at least two inhibitors $\exists\ p, q \neq t$ such that $\sigma_p = +1, \sigma_q = -1$.

Input case 1: $x_t = a, x_p = 1, x_{i \notin \{t,p\}} = 0, x_k = 1$. The signed sum $S = \sigma_t a + 1 \geq 0$. So the output is always 1 as $x_k = 1$.

Input case 2: $x_t = a, x_q = 1, x_{i \notin \{t,q\}} = 0, x_k = 0$. The signed sum $S = \sigma_t a - 1 \leq 0$. So the output is always 0 as $x_k = 0$.

As a result, none of the x_t can be a canalyzing variables for $t < k$. Now we will show that x_k cannot be canalyzing either. Let us fix $x_k = a \in \{0, 1\}$. Choosing $p, q < k$ such that $\sigma_p = +1, \sigma_q = -1$ allows us to select two input combinations such that one leads to the output 1 whereas the other leads to the output 0 while keeping $x_k = a$ fixed.

Input case 1: $x_p = 1$ and $x_i = 0\ \forall\ i \neq p, k$. Then $S = 1 + \sigma_k a \geq 0$. Now if $1 + \sigma_k a > 0$, the output is 1. S can be equal to 0 only when $\sigma_k = -1, a = 1$. In that case too, the output is 1 since $x_k = a = 1$.

Input case 2: Here we further break it down into two subcases:

Subcase 1 ($\sigma_k = 1$): Choose $m \neq q$ such that $\sigma_m = -1$. Use the input: $x_k = a, x_q = x_m = 1$, all other $x_i = 0$. Then, $S = -2 + a < 0$. So, the output is always 0.

Subcase 2 ($\sigma_k = -1$): $x_k = a, x_q = 1$, all other $x_i = 0$. Then, $S = -1 - a < 0$. The output is again 0.

So, this proves that a BF which is equal to a BMR with k regulators and contains a self-regulation is canalyzing if and only if there is at most one activator or at most one inhibitor. □

Property C.9. Let f_1 and f_2 denote the BFs corresponding to a k-input IMR and a k-input BMR, respectively, satisfying a given sign assignment of the regulators. Then, the following statements hold: (a) If k is even, then $f_1 = f_2$ if and only if the number of activatory and inhibitory regulators

are equal. (b) If k is odd and $k > 1$, then f_1 and f_2 are never equal.

Proof. (a) First, consider the case when k is even and one of the k regulators is the self-regulator. Without loss of generality, assume that it is the k^{th} regulator that is the target node itself. Let S_I and S_B denote the signed sums corresponding to IMR and BMR, respectively.

$$S_I(x') = \sum_{i=1}^{k} \sigma_i x'_i, \;\; S_B(x) = \sum_{i=1}^{k} \sigma_i x_i$$

where $x = (x_1, x_2, \ldots, x_k)$, $x' = (x'_1, x'_2, \ldots, x'_k)$ and $x'_i = 2x_i - 1$. For IMR (respectively, BMR), the output is active if $S_I > 0$ (similarly, $S_B > 0$) or $S_I = 0$ and $x'_k = +1$ (respectively, $S_B = 0$ and $x_k = 1$). Now, $S_I(x') = \sum_{i=1}^{k} \sigma_i x'_i = \sum_{i=1}^{k} \sigma_i(2x_i - 1) = 2S_B(x) - \Sigma$, where $\Sigma = \sum_{i=1}^{k} \sigma_i$. Thus, we have the following three conditions of equivalency: (i) $S_I > 0 \iff S_B > \frac{\Sigma}{2}$; (ii) $S_I = 0 \iff S_B = \frac{\Sigma}{2}$; and (iii) $x'_k = +1 \iff x_k = 1$ (since $x'_k = 2x_k - 1$).

Now if the number of activators and inhibitors are equal, then $\Sigma = 0$. Then, using the three equivalency conditions, f_1 and f_2 are either both active or both inactive for every input x' to IMR and for the corresponding x input to BMR, i.e., $f_1 = f_2$ up to the mapping between x and x'.

We can now prove the converse. We will show that if the number of activators and inhibitors are unequal, i.e., if $\Sigma \neq 0$, then there exists x' such that $f_1(x') \neq f_2(x)$. To proceed, we note that since k is even, Σ is an even integer.

Case 1 ($\Sigma > 0$): Since Σ is positive, the number of activators is strictly greater than 1 as $k \geq 2$. So, there exists $j \neq k$ such that $\sigma_j = +1$. For f_2, consider the input x where $x_j = 1$ while for all other $i \neq j$, $x_i = 0$. Then, $S_B(x) = 1 > 0$ and hence the output is 'on' for f_2. Now, from the equivalency condition, the output for f_1 will be 'on' if $S_B(x) > \frac{\Sigma}{2}$ or $S_B(x) = \frac{\Sigma}{2}$ and $x_k = 1$. But none of these conditions are satisfied since $\frac{\Sigma}{2} \geq 1$ and $x_k = 0$ and as a result f_1 and f_2 cannot be equal.

Case 2 ($\Sigma < 0$): Since Σ is a negative even integer, $\frac{\Sigma}{2} \leq -1$. Now, for f_2 consider the input x such that $x_i = 0$ for all i. Then, $S_B(x) = 0$ and $x_k = 0$ and hence the output would be 0 ('off'). Furthermore, since $S_B = 0$ and $0 > -1 \geq \frac{\Sigma}{2}$, for f_1 the output will be +1. So again we see that f_1 and f_2 are not equal.

Exactly the same arguments can be applied if the target is not an explicit regulator. This proves the first part that if k is even then $f_1 = f_2$ if and only if the number of activators and in-

hibitors are equal for a given sign assignment.

(b) Now we consider the case when k is odd. If the target is not an explicit regulator, then f_1 corresponds to a k-input BF whereas f_2 corresponds to a $(k+1)$-input BF. Hence, in this case $f_1 \neq f_2$. However, we now examine the possibility of $f_1 = f_2$ when the target is an explicit regulator. We can demonstrate using a counter-example that f_1 can never be equal to f_2 in this case either. Since k is odd, the signed sum can never be zero, leading us to consider the two cases: (i) $\Sigma > 0$ and (ii) $\Sigma < 0$.

Case 1 ($\Sigma > 0$):

Subcase 1.1 ($\Sigma = 1$): Then there are $\frac{(k+1)}{2}$ activators and $\frac{(k-1)}{2}$ inhibitors. We turn all the inhibitors on and $\frac{(k-1)}{2}$ activators on, ensuring $x_k = 1$. We keep all other regulators off. Then $S_B = 0$ (since $k \geq 3$) and since $x_k = 1$ the output is active. However, $S_I = 2S_B - \Sigma = -1 < 0$. So, the output is inactive. Hence, $f_1 \neq f_2$.

Subcase 1.2 ($\Sigma \geq 3$): Since Σ is positive and k is odd, there exists at least $\frac{(k+1)}{2}$ activators, i.e., at least 2 activators because $k \geq 3$. Thus, there exists $j \neq k$ such that $\sigma_j = +1$. For f_2, consider an input x such that $x_j = 1$ and $x_i = 0$ for all $i \neq j$. Then $S_B(x) = 1$ and hence the output is 'active'. But, $S_I(x') = 2.1 - \Sigma \leq 2 - 3 = -1 < 0$. So, the output is inactive and hence we can conclude $f_1 \neq f_2$.

Case 2 ($\Sigma < 0$): We consider an input x such that $x_i = 0$ for all i. Then $S_B(x) = 0$ and since $x_k = 0$, the output is inactive. Now, $S_I(x') = 2.0 - \Sigma = -\Sigma > 0$ since $\Sigma < 0$. Hence, the output will be active. Thus, $f_1 \neq f_2$.

This proves that if k is odd and $k > 1$, then f_1 and f_2 can never be equal, regardless of the sign assignments. However, when $k = 1$ and the self-regulation is activatory, then f_1 and f_2 are equal. This is the only situation at odd k for which $f_1 = f_2$. □

Property C.10. A k-input BF which is equal to an IMR has the maximum average sensitivity among all k-input unate functions (UFs).

Proof. A result in [201] suggests that for any non-constant monotone BF f with k inputs, the average sensitivity $\overline{s}(f)$ has the following bounds:

$$\frac{k}{2^k} \leq \overline{s}(f) \leq \binom{k}{\lfloor k/2 \rfloor} \cdot \frac{\lceil k/2 \rceil}{2^{k-1}}$$

The upper bound is tight for the function 'MAJORITY'. In Boolean logic a majority function outputs 'False' when half or more inputs are 'False', and 'True' otherwise. The Majority rule exactly corresponds to the IMR when all regulatory inputs are activatory (i.e., have positive signs) with the exception that in case of a tie (i.e., when the number of 'True' and 'False' inputs are equal) the majority function outputs 'False'. However, in the IMR rule, a tie (i.e., when the signed sum equals zero) results in the output being determined by the state of the target node itself. IMR generalizes the Majority functions by allowing all possible assignments of regulatory signs. Monotone functions form a specific subset of the unate functions (UFs). Any UF can be transformed into a monotone function by appropriately negating some of its inputs [202]. Since the average sensitivity remains invariant under input negation, the maximum value of $\overline{s}(f)$ over the space of all UFs will be the same as that over all monotone functions and is achieved by the IMRs. It is important to note that the distinction in the tie-breaking scenarios for the two cases does not affect the value of the average sensitivity [201]. □

Corollary C.10.1. For any odd k, a k-input BF equal to an IMR has the same average sensitivity as a $(k+1)$-input BF equal to an IMR.

Proof. Since k is odd, introduce m such that $k = 2m+1$. The average sensitivity of the BF with k inputs that is equal to an IMR is given by

$$\overline{s}(f_k) = \binom{k}{\lfloor k/2 \rfloor} \cdot \frac{\lceil k/2 \rceil}{2^{k-1}}$$

Noting that $\lfloor k/2 \rfloor = m$ and $\lceil k/2 \rceil = m+1$, we have

$$\overline{s}(f_k) = \frac{\binom{2m+1}{m} \cdot (m+1)}{2^{2m}}$$

Thus the average senstivity of the BF with $(k+1)$ inputs that is equal to an IMR is given by

$$\overline{s}(f_{k+1}) = \frac{\binom{2m+2}{m+1} \cdot (m+1)}{2^{2m+1}} = \frac{2m+2}{m+1} \cdot \frac{\binom{2m+1}{m} \cdot (m+1)}{2^{2m+1}} = \frac{\binom{2m+1}{m} \cdot (m+1)}{2^{2m}} = \overline{s}(f_k)$$

which proves the statement. □

Property C.11. A k-input BF that corresponds to a k-input Ising* rule exhibits odd bias if and

only if k is a power of two.

Proof. The Ising* rule, a modification of IMR proposed in [190], is defined identically to IMR except in its handling of a zero signed sum. Under IMR, a vanishing sum causes the target node to retain its current state. In contrast, Ising* forces the target node into the 'off' state (-1) whenever the signed sum is zero. Formally, the Ising* rule can be written as:

$$f(x_1, \ldots, x_k) = \begin{cases} +1, & \sum_{i=1}^{k} \sigma_i x_i > 0, \\ -1, & \sum_{i=1}^{k} \sigma_i x_i \leq 0, \end{cases} \tag{C.8}$$

where $x_i \in \{-1, +1\}$. So a k-input Ising* rule can always be mapped to a k-input BF, regardless of self-regulation or parity of k. When k is odd there are no zero-sum cases, so Ising* coincides exactly with the original IMR variant. In that case (see Property C.1), the only odd bias possibility is $k = 1 (= 2^0)$.

We now turn to the even k case. Suppose there are k inputs, with k even and each input $x_i \in \{-1, +1\}$. Further, each input is assigned with a sign $\sigma_i \in \{-1, +1\}$ depending on its activatory or inhibitory nature. Define $y_i = \sigma_i x_i$, so the signed sum is given by

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sum_{i=1}^{k} y_i,$$

where each $y_i \in \{-1, +1\}$. Clearly, $S = 0$ precisely when half of the y_i equal to $+1$ and half equal -1, i.e.,

$$|\{i : y_i = +1\}| = |\{i : y_i = -1\}| = \frac{k}{2}.$$

Since, there are $\binom{k}{k/2}$ ways to choose which $k/2$ inputs are $+1$, exactly $\binom{k}{k/2}$ of the 2^k input combinations satisfy $S = 0$, and each of these leads to an output of -1, under Ising*. By symmetry, the remaining inputs split evenly between $S > 0$ (output $+1$) and $S < 0$ (output -1). Hence, the bias (number of $+1$ outputs) is given by:

$$\begin{aligned} P^k_{\text{Ising*}} &= \frac{2^k - \binom{k}{k/2}}{2} \\ &= 2^{k-1} - \frac{1}{2}\binom{k}{k/2} \end{aligned}$$

Thus, for any even $k \geq 2$, the bias of a k-input Ising* rule is odd if and only if $\frac{1}{2}\binom{k}{k/2}$ is odd. As a consequence of Kummer's theorem [203], for even k, $\frac{1}{2}\binom{k}{k/2}$ is odd if and only if k is a power of two [204]. Note that, for $k = 2$, the bias of the Ising* rule equals 1, which is possible only if it is the AND function and therefore an NCF. □

Property C.12. A k-input BF that corresponds to an Ising* rule is never canalyzing for $k > 2$.

Proof. We consider the definition of Ising* as provided in Property C.11 (see Eq. (C.8)). We will show that for each $1 \leq r \leq k$ and for every fixed value $a \in \{+1, -1\}$, there always exist two assignments of the remaining variables under which f takes both the values $+1$ and -1. We denote $\alpha_i = sign(\sigma_i)$ for all i then $\sigma_i \cdot \alpha_i = 1; \sigma_i \cdot (-\alpha_i) = -1$ for all i.

Suppose x_r is a canalyzing variable with $x_r = a$.

Input 1: $x_r = a, x_i = \alpha_i \; \forall \; i \neq r$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i \alpha_i = \sigma_r a + (k-1) \geq -1 + (k-1) > 0 \text{ since } k > 2$$

So, the output is $+1$ in this case.

Input 2: $x_r = a, x_i = -\alpha_i \; \forall \; i \neq r$. Then, the signed sum

$$S = \sum_{i=1}^{k} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i x_i = \sigma_r a + \sum_{i \neq r} \sigma_i (-\alpha_i) = \sigma_r a - (k-1) \leq 1 - (k-1) < 0 \text{ since } k > 2$$

So, the output is -1 in this case. So, none of the variables is a canalyzing variable. This completes the proof. □

Property C.13. Consider a BF f that corresponds to a 01* rule with k regulators, one of which is a self-regulation. Then for $k > 1$, if the self-regulation is inhibitory, but not all other regulators are inhibitory, f is ineffective, with the self-input being the ineffective input.

Proof. The preprint [190] introduces a modified version of BMR rule, denoted 01*, which coincides with the original BMR rule except in two cases. First, if all regulators are inhibitory and every input is off, 01* outputs 'on' (1). Second, if all regulators are activatory and every input is off, it outputs 'off' (0). So, for any other sign combination of the regulators, 01* coincides with

BMRs and hence the Property follows from the previous Property C.4 for BMRs. However, unlike the BMRs, where the case of all inhibitions including a self-regulation produce the constant 0 function, 01* yield an AND rule and is therefore an NCF. When all regulators are activatory and one of them is a self-regulation, 01* again coincides with BMR and yield an OR rule and thus an NCF. □

C.2 Generation of models with nested canalyzing functions preserving biological attractors

Here, we outline the systematic procedure we have undertaken for constructing Boolean models with scNCFs for a GRN with fixed topology and known biological attractors. In Section 2.7 in Chapter 2, it was discussed how, given a network topology along with the signs of the interactions and the biological attractors, one can sample BFs restricted to different classes such that they satisfy the attractor constraints. Here, we employ that methodology to construct suitable scNCF models. We first impose the requirement that the BN reproduces the desired biological attractors. This constraint is implemented as follows. For a given gene in the network, we extract from each biological fixed point the expression state of that gene together with the expression states of its regulators. Each such input configuration fixes the corresponding output entry of the gene's truth table to match the observed expression state in the fixed point. This procedure is applied independently to every gene in the network. If conflicting constraints arise, namely, if two fixed points prescribe different output values for the same entry in the truth table, then the given network topology cannot realize those fixed points simultaneously. Each fixed point constrains exactly one row of every truth table, so that M fixed points impose constraints on at most M rows (some constraints may be redundant). While this guarantees recovery of all biological fixed points, it does not preclude the existence of additional attractors. Note that for cyclic attractors, it is possible to impose analogous constraints by fixing entries in the truth table using consecutive states along the cycle. Next, we restrict the BF at each node to be consistent with the known activatory or inhibitory signs of its regulators, thereby enforcing sign-conformity with the network structure. Finally, we restrict the admissible BFs to the class of NCFs; in our work, all BFs of this class have been enumerated by recursion on k, once and for all, up to $k = 9$ [64, 148]. This yields,

independently for each node, a list of admissible scNCFs that are consistent with the biological attractors. All possible combinations of these node-specific scNCFs then define the full set of Boolean models that both preserve the network topology and recover the biological attractors.

Because the admissible scNCFs are generated independently at each node, this framework also allows selective modification of update rules. In our analysis in Section 6.5.5 in Chapter 6, for four published models (network IDs 61, 69, 95 and 212 in BBM dataset) we retained the update rules at nodes that were already governed by NCFs, while replacing non-NCF rules at other nodes with all possible scNCFs consistent with the biological attractors. This procedure yields all admissible scNCF-based models for these networks under the stated constraints. Practical implementations of this procedure for constructing scNCF models that preserve biological attractors for a given network topology are available in our GitHub repositories https://github.com/asamallab/TMRvsBioLogic and https://github.com/asamallab/LDLM.

C.3 Supplementary tables for Chapter 6

Table C.1: Numbers of various subtypes of the EThFs in the theoretical space. Here k corresponds to the number of effective inputs to a BF. The columns under the grouping 'Number of BFs' report the total count of BFs in theory belonging to the classes EThF, NCF, IMR, BMR, non-NCF BMR, NCF BMR and non-BMR NCF, across varying values of k.

k	Number of BFs						
	EThF	NCF	IMR	BMR	non-NCF BMR	NCF BMR	non-BMR NCF
1	2	2	2	1	0	1	1
2	8	8	0	3	0	3	5
3	72	64	8	16	3	13	51
4	1536	736	32	43	38	5	731
5	86080	10624	32	106	100	6	10618
6	14487040	183936	192	249	242	7	183929
7	8274797440	3715072	128	568	560	8	3715064
8	17494930604032	85755372	1024	1271	1262	9	85755363
9	144222448789966828	2226939904	512	2806	2796	10	2226939894

Table C.2: Numbers and proportions of various subtypes of the EThFs in the BBM dataset. Here k corresponds to the number of effective inputs to a BF. The columns under the grouping 'Number of BFs' report the total count of BFs in the BBM dataset ('Total' column), along with the counts of BFs belonging to the classes EThF, NCF, IMR, BMR, non-NCF BMR, NCF BMR and non-BMR NCF, across varying values of k. The columns under grouping 'Fraction of subtypes within EThFs' presents the relative proportions of the different subtypes of EThF within EThFs for each k.

k	Number of BFs								Fraction of subtypes within EThFs					
	Total	EThF	NCF	IMR	BMR	non-NCF BMR	NCF BMR	non-BMR NCF	NCF	IMR	BMR	non-NCF BMR	NCF BMR	non-BMR NCF
1	2013	2013	2013	2013	1771	0	1771	242	1.0	1.0	0.88	0.0	0.88	0.12
2	1634	1634	1634	0	902	0	902	732	1.0	0.0	0.552	0.0	0.552	0.448
3	914	908	895	13	312	12	300	595	0.986	0.014	0.344	0.013	0.33	0.655
4	552	515	506	0	85	2	83	423	0.983	0.0	0.165	0.004	0.161	0.821
5	371	318	303	0	52	5	47	256	0.953	0.0	0.164	0.016	0.148	0.805
6	222	168	160	0	10	0	10	150	0.952	0.0	0.06	0.0	0.06	0.893
7	122	84	81	1	3	1	2	79	0.964	0.012	0.036	0.012	0.024	0.94
8	74	51	50	0	3	0	3	47	0.98	0.0	0.059	0.0	0.059	0.922
9	51	31	29	0	1	0	1	28	0.935	0.0	0.032	0.0	0.032	0.903

Table C.3: Relative enrichment ratios (E_R) and statistical significance (p-values) for EThF subtypes in the BBM dataset. Here, k is the number of effective inputs per BF. Under 'E_R of subtypes within EThFs', the columns 'NCF', 'IMR' and 'BMR' display the relative enrichment ratios (E_R) for each subtype at given k (see Chapter 6, Section 6.4.3). $E_R > 1$ indicates enrichment of the corresponding subtype within EThFs in the BBM dataset. No entry appears for subtype IMR at $k = 2$ because no 2-input EF is equal to such an IMR. Under 'p-values associated with E_R of different subtypes', the same columns report the significance of each enrichment (with $p < 0.05$ indicating statistical significance). These p-values have been used to annotate Figure 6.3(c) in Chapter 6. For $k = 1$, the subtypes NCF and IMR coincide with EThF, and for $k = 2$, NCF coincides with EThF while IMR is absent; accordingly, p-values are not shown in these cases.

k	E_R of subtypes within EThFs			p-values associated with E_R of different subtypes		
	NCF	IMR	BMR	NCF	IMR	BMR
1	1	1	1.76	-	-	4.85×10^{-287}
2	1	-	1.47	-	-	9.59×10^{-48}
3	1.11	0.13	1.55	3.09×10^{-30}	1	4.47×10^{-17}
4	2.05	0	5.90	3.98×10^{-146}	1	4.27×10^{-39}
5	7.72	0	1.33×10^{2}	1.29×10^{-251}	1	7.3×10^{-92}
6	75.01	0	3.46×10^{3}	4.67×10^{-291}	1	8.43×10^{-33}
7	2.15×10^{3}	7.7×10^{5}	5.2×10^{5}	6.42×10^{-267}	1.3×10^{-6}	3.08×10^{-17}
8	2.0×10^{5}	0	8.1×10^{8}	1.68×10^{-264}	1	7.99×10^{-27}
9	6.06×10^{7}	0	1.66×10^{12}	1.38×10^{-224}	1	6.03×10^{-13}

Table C.4: Relative enrichment ratios (E_R) and statistical significance (p-values) for the EThF subtypes 'non-BMR NCF', 'NCF BMR' and 'non-NCF BMR' in the BBM dataset. Here, k is the number of effective inputs per BF. A, B and C denote the subtypes non-BMR NCF, NCF BMR and non-NCF BMR, respectively. Under 'E_R of subtypes within EThFs', the columns display the relative enrichment ratios (E_R) for each subtype at given k (see Chapter 6, Section 6.4.3). $E_R > 1$ indicates enrichment of the corresponding subtype within EThFs in the BBM dataset. No entry appears for column 'C' at $k = 1$ and 2 because there are no EFs of the type 'non-NCF BMR' for $k = 1$ and 2. Under 'p-values associated with E_R of different subtypes and their order', the same columns report the significance of each enrichment. These p-values have been used to annotate Figure 6.3(d) in Chapter 6. As for $k = 1$ and 2, no EF belongs to the non-NCF BMR category, so the corresponding p-values are omitted. The 'Order' column ranks subtypes by increasing p-value; larger p-values indicate weaker statistical significance for enrichment.

k	E_R of subtypes within EThFs			p-values associated with E_R of different subtypes and their order			
	A	B	C	A	B	C	Order
1	0.24	1.76	-	1	4.85×10^{-287}	-	B < A
2	0.72	1.47	-	1	9.59×10^{-48}	-	B < A
3	0.93	1.83	0.32	1.00	1.99×10^{-27}	1	B < A < C
4	1.73	49.51	0.16	3.2×10^{-59}	2.44×10^{-110}	1.00	B < A < C
5	6.53	2.12×10^{3}	13.53	4.87×10^{-170}	1.87×10^{-139}	4.11×10^{-5}	A < B < C
6	70.33	1.23×10^{5}	0	1.95×10^{-261}	2.61×10^{-48}	1	A < B < C
7	2.09×10^{3}	2.46×10^{7}	1.76×10^{5}	1.03×10^{-257}	3.26×10^{-15}	5.68×10^{-6}	A < B < C
8	1.88×10^{5}	1.14×10^{11}	0	6.98×10^{-245}	2.84×10^{-33}	1	A < B < C
9	5.85×10^{7}	4.65×10^{14}	0	8.62×10^{-216}	2.15×10^{-15}	1	A < B < C

Table C.5: 24 selected models from the BBM dataset. This table lists the 24 published models selected through our multi-step filtration pipeline (see Chapter 6, Figure 6.4). The column 'ID in BBM dataset' provides the unique identifier assigned to the network model in the BBM dataset. The column 'System' briefly describes the biological system modeled, and the column 'References' points to the corresponding peer-reviewed publications.

Serial Number	ID in BBM datatset	System	References
1	7	Mammalian cortical area development	[151]
2	8	Death receptor signaling	[107]
3	10	Early heart development	[152]
4	22	Terminal differentiation of B cells	[205]
5	23	Mammalian cell cycle	[23]
6	31	Control of cell-cycle transcription	[206]
7	40	T helper cell differentiation	[110]
8	61	Tumor microenvironment in acute lymphoblastic leukemia	[207]
9	63	The *lac* operon in *E. coli*	[208]
10	67	Regulation of L-arabinose operon in *E. coli*	[209]
11	69	*A. fumigatus* iron regulation and oxidative stress response	[210]
12	74	T-LGL survival signaling network (reduced)	[36]
13	85	Reprogramming of testes-derived germline stem (GS) cells to germline pluripotent stem cells (GPS)	[211]
14	86	Molecular pathways enabling tumour cell invasion and migration	[212]
15	88	miR-9 controlled neurogenesis	[213]
16	95	Cell-cycle control	[214]
17	99	*C. albicans* yeast to hyphal transition	[215]
18	100	Acute myeloid leukemia	[216]
19	109	Asymmetric cell division in *C. crescentus* (A)	[217]
20	110	Asymmetric cell division in *C. crescentus* (B)	[217]
21	133	*A. thaliana* root stem cell niche	[149]
22	200	Non-small cell lung cancer cell lines	[218]
23	208	Early hematopoiesis aging	[39]
24	212	Transport, synthesis, and metabolism of Tryptophan in *E. coli*	[219]

Table C.6: Summary of models following replacement of non-NCF BFs with NCFs. Each row corresponds to a specific network, identified by its ID as given in the BBM dataset. Column 'N' gives the number of nodes in the network, and column 'M' indicates the number of nodes whose assigned BF in the published model is non-NCF. The column 'Number of replaced models' denotes the total number of candidate models obtained by replacing all non-NCFs with valid scNCFs, such that the resulting model still recovers the biological attractors. The column 'Number of models for analysis' refers to the subset of these models that were randomly sampled (if too large) or fully enumerated for further dynamical analysis. The next three columns report how many models exhibit attractor recovery scores that are higher than or equal to the published model and higher than the IMR- and BMR-type variants, respectively. The final three columns show how many models yield JS distances (from the gold standard basin fraction distribution) that are lower than or equal to that of the published model and are lower than those for the IMR- and BMR-type variants, respectively.

ID	N	M	Number of replaced models	Number of models considered for analysis	Number of models with $\text{ARS}_{repl} \geq \text{ARS}_{model}$	Number of models with $\text{ARS}_{repl} > \text{ARS}_{\text{IMR}}$	Number of models with $\text{ARS}_{repl} > \text{ARS}_{\text{BMR}}$	Number of models with $D_{repl} \leq D_{model}$	Number of models with $D_{repl} < D_{\text{IMR}}$	Number of models with $D_{repl} < D_{\text{BMR}}$
61	26	2	686244	5000	1640	5000	5000	4064	5000	5000
69	22	2	96	96	96	96	96	96	96	96
95	10	4	2116	2116	576	2080	2116	313	2116	2116
212	16	1	48	48	9	48	48	4	48	48

C.4 Supplementary figures for Chapter 6

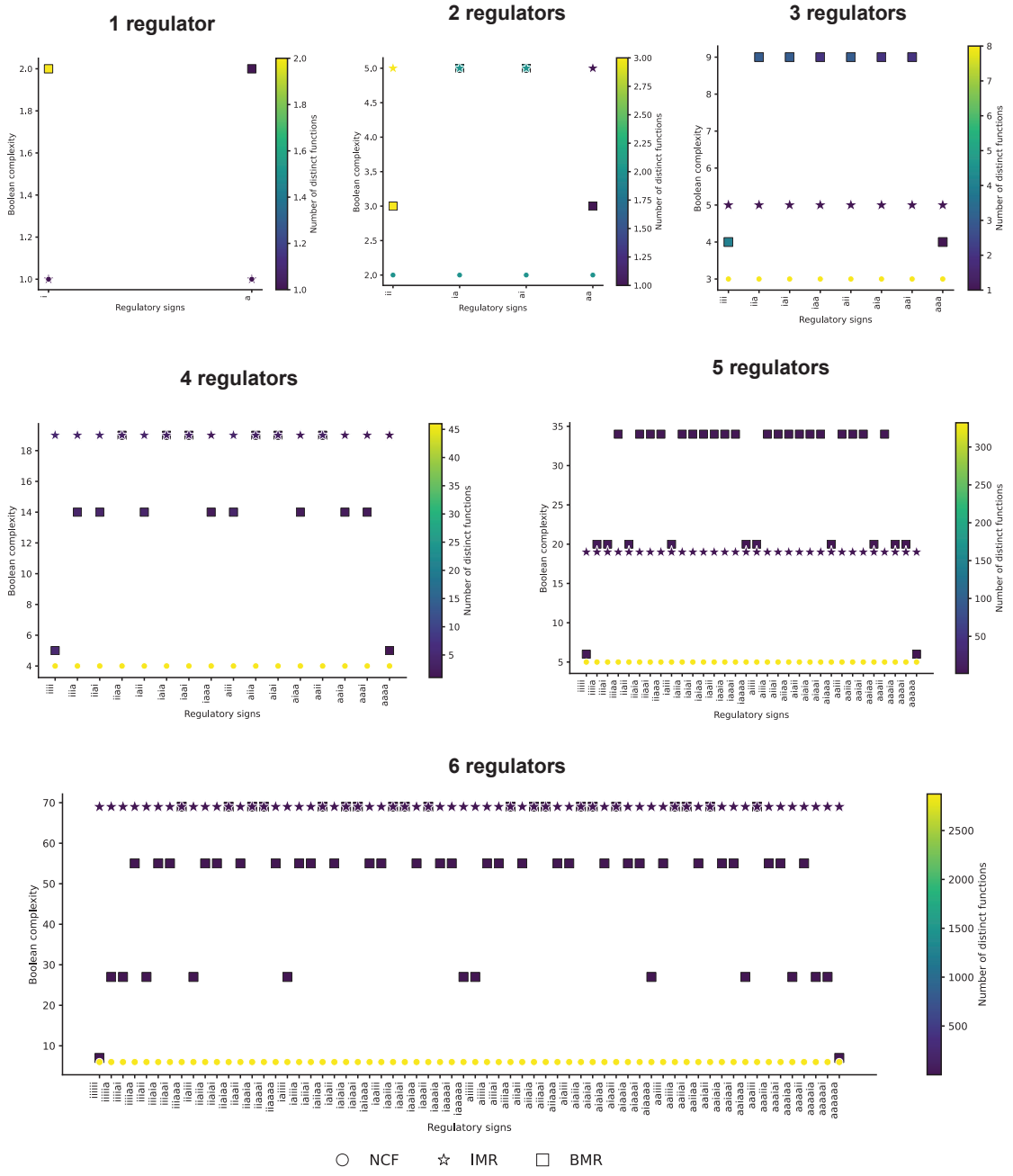


Figure C.1: TMRs generally exhibit substantially higher Boolean complexity than NCFs across different sign combinations of regulators. Each subplot corresponds to a fixed number of regulators k (≤ 6) and shows the Boolean complexity of IMRs, BMRs, and NCFs for different combinations of regulator signs; the number of distinct functions that have a specific value of the Boolean complexity is shown using a color gradient. While NCFs always have Boolean complexity exactly equal to the number of regulators k, IMRs and BMRs generally exhibit much higher Boolean complexity across sign combinations.

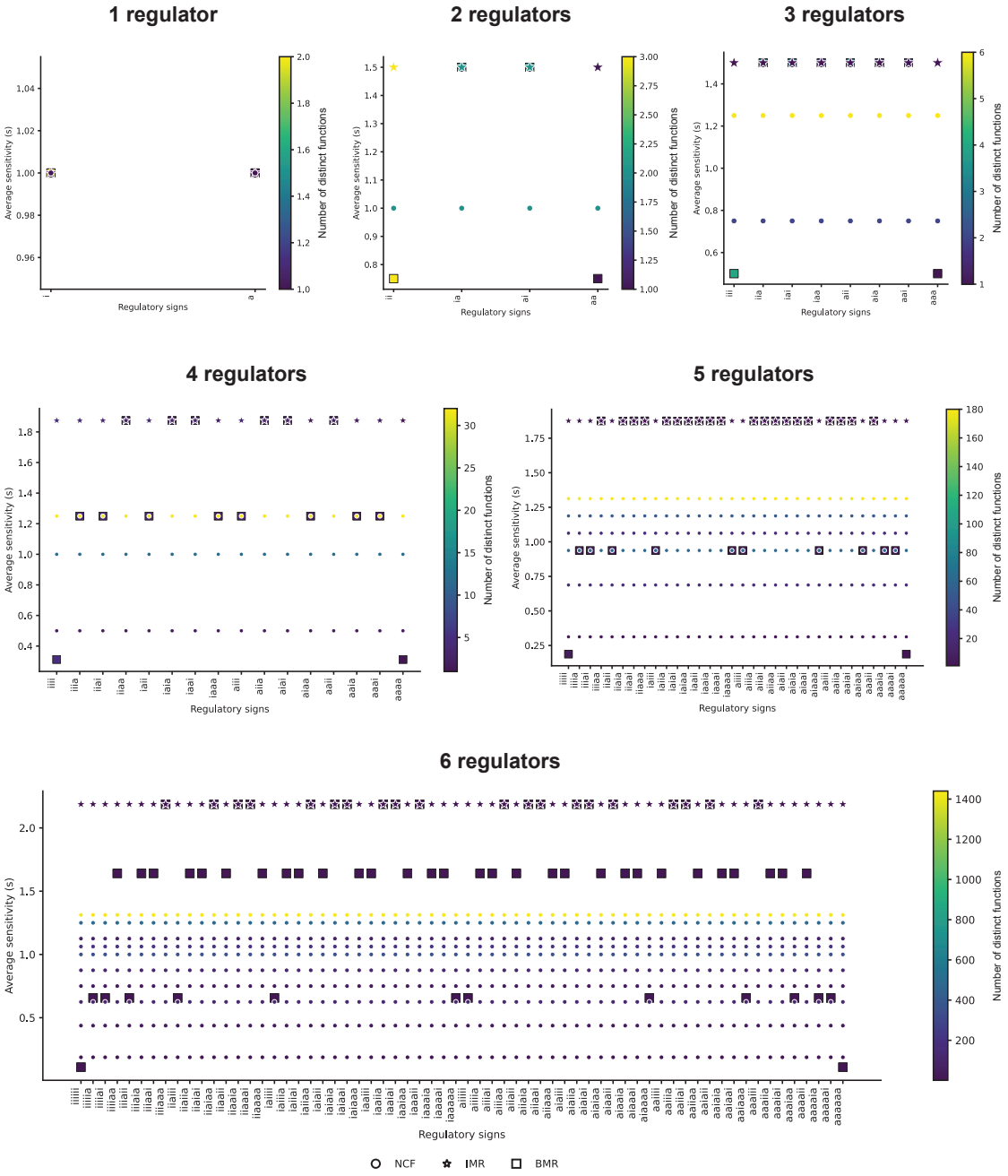


Figure C.2: IMRs attain the highest average sensitivity, and for most sign combinations of regulators there exist NCFs with lower sensitivity than BMRs. Each subplot corresponds to a fixed number of regulators k (≤ 6) and shows the average sensitivity of IMRs, BMRs, and NCFs for different combinations of regulator signs. IMRs attain the maximal average sensitivity for all regulator sign combinations. For a fixed sign combination, IMRs and BMRs each have a unique average sensitivity, whereas NCFs span multiple values and, in most cases, include functions with lower average sensitivity than TMRs. In each subplot, for a given sign combination, the number of distinct functions that have a specific value of the average sensitivity is shown using a color gradient.

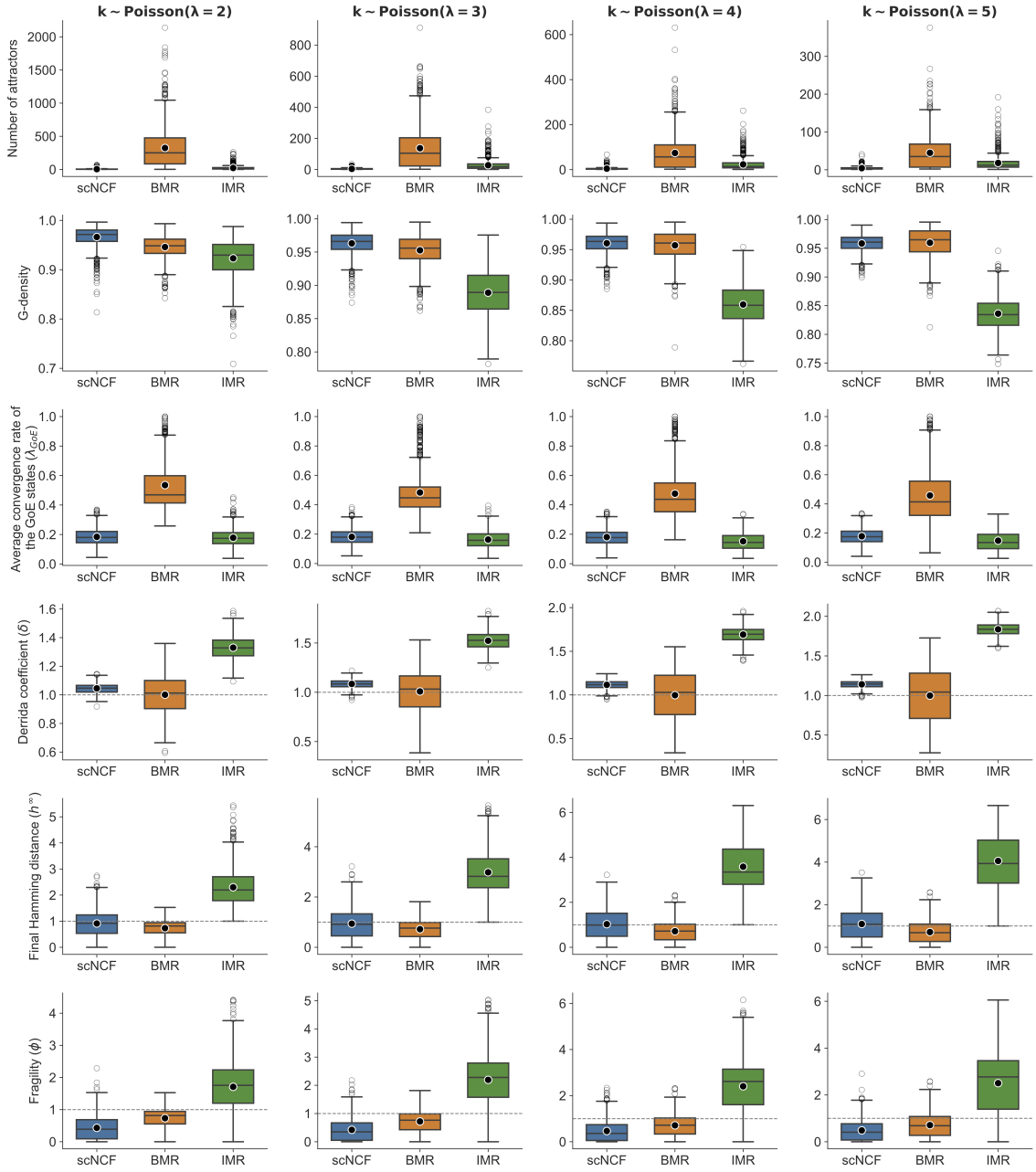


Figure C.3: BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the P-P network topology with $N = 16$. Each row corresponds to one observable and each of the four columns represent different values of the mean in-degree parameter ($\lambda = 2, 3, 4, 5$). In each subplot, the three boxplots present the distributions of the observables across three subtypes of ThFs: scNCFs, BMRs and IMRs. Across all observables and values of λ, IMR-based models consistently display the least stable behavior, while scNCF and BMR models typically cluster closer to the ordered or critical regimes. Outliers in each distribution are shown as grey circles while the black dots outlined in white denote the mean of each group.

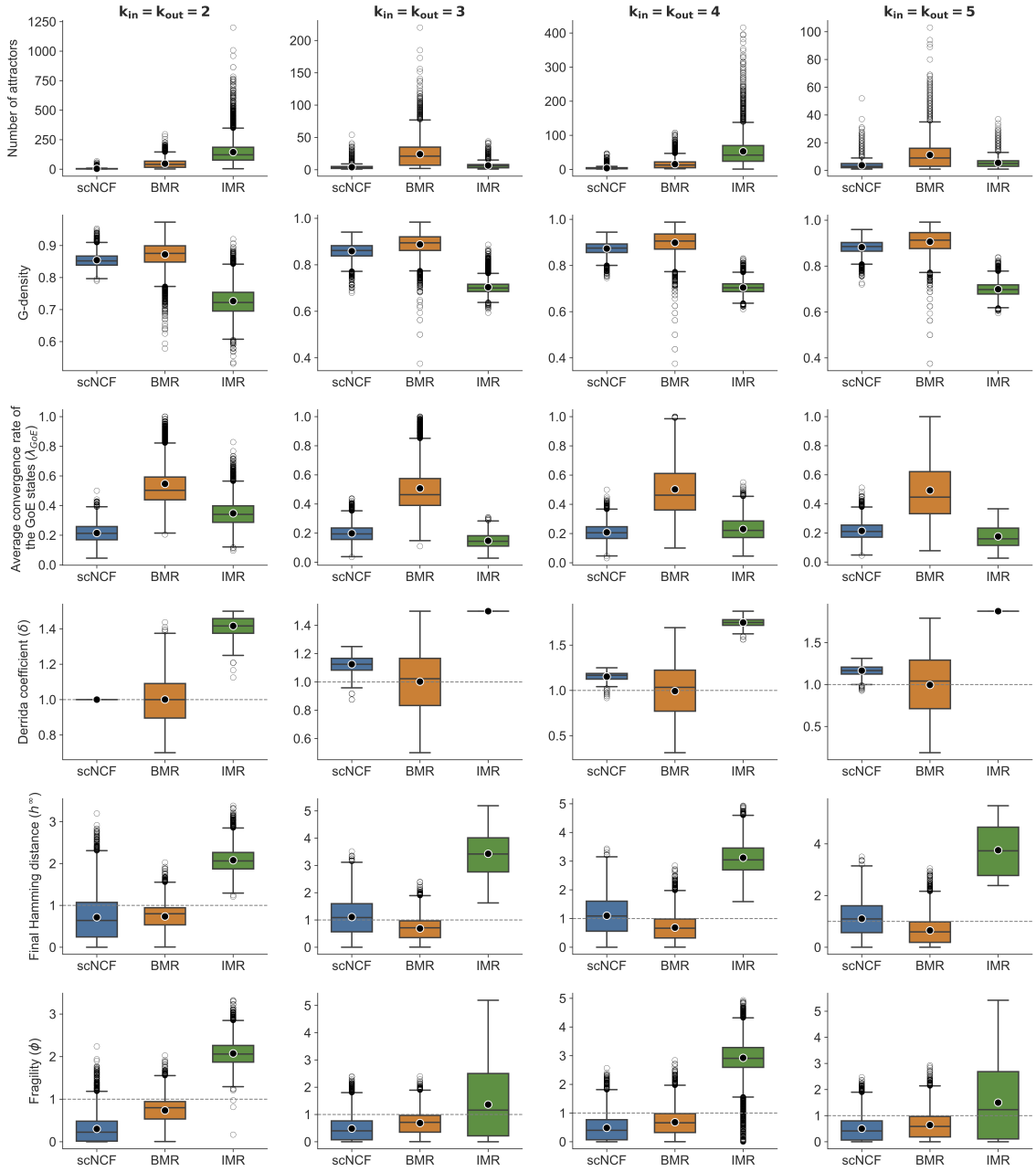


Figure C.4: BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the R-R network topology with $N = 12$. Each row corresponds to one observable and each of the four columns represent different values of the fixed in-degree ($k_{in} = k_{out} = 2, 3, 4, 5$). In each subplot, the three boxplots present the distributions of the observables across three subtypes of ThFs: scNCFs, BMRs and IMRs. Across all observables and values of k_{in}, IMR-based models consistently display the least stable behavior, while scNCF and BMR models typically cluster closer to the ordered or critical regimes. Outliers in each distribution are shown as grey circles while the black dots outlined in white denote the mean of each group.

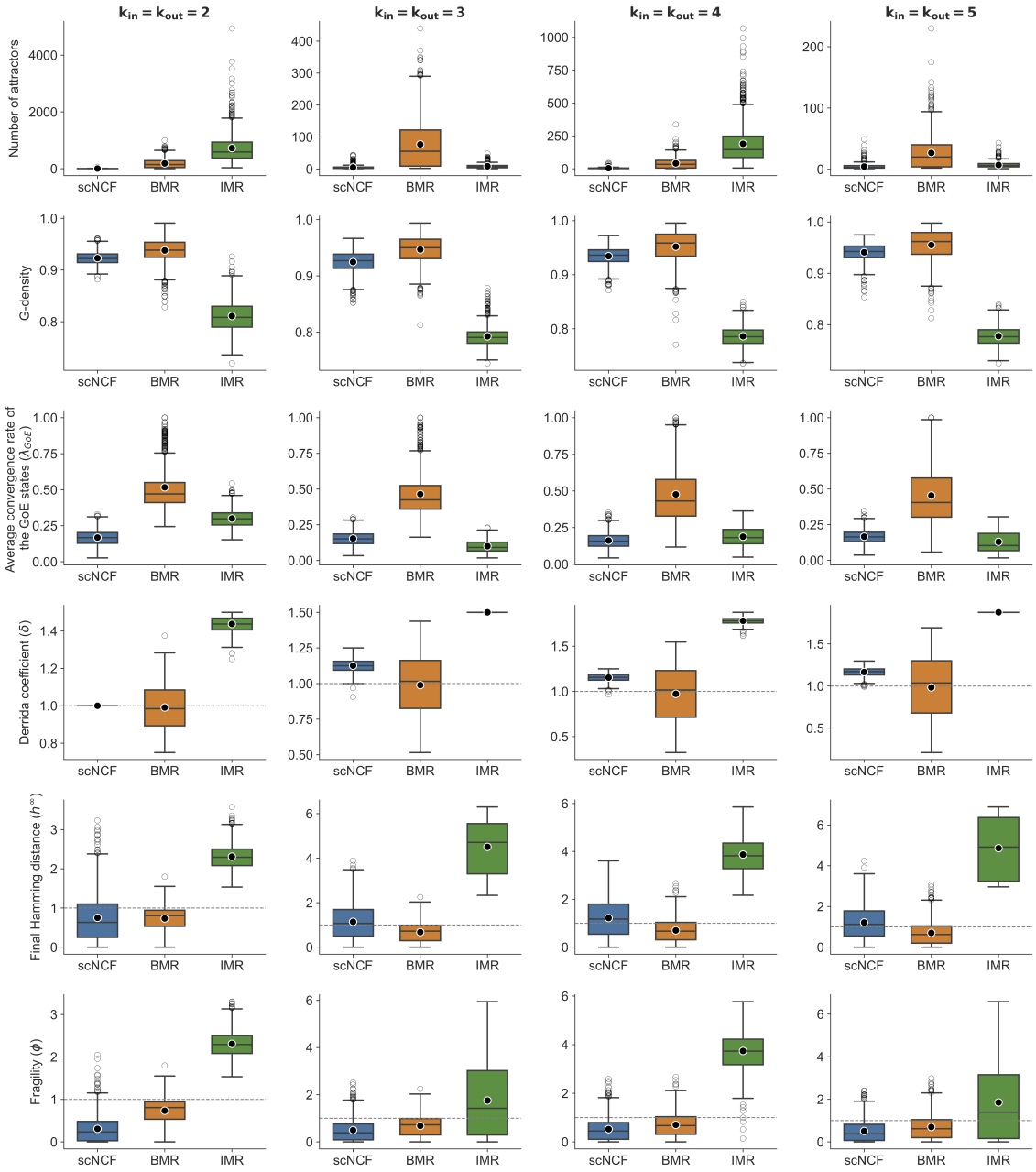


Figure C.5: BMRs and scNCFs typically exhibit more stable dynamics than IMRs across stability measures for the R-R network topology with $N = 16$. Each row corresponds to one observable and each of the four columns represent different values of the fixed in-degree ($k_{in} = k_{out} = 2, 3, 4, 5$). In each subplot, the three boxplots present the distributions of the observables across three subtypes of ThFs: scNCFs, BMRs and IMRs. Across all observables and values of k_{in}, IMR-based models consistently display the least stable behavior, while scNCF and BMR models typically cluster closer to the ordered or critical regimes. Outliers in each distribution are shown as grey circles while the black dots outlined in white denote the mean of each group.

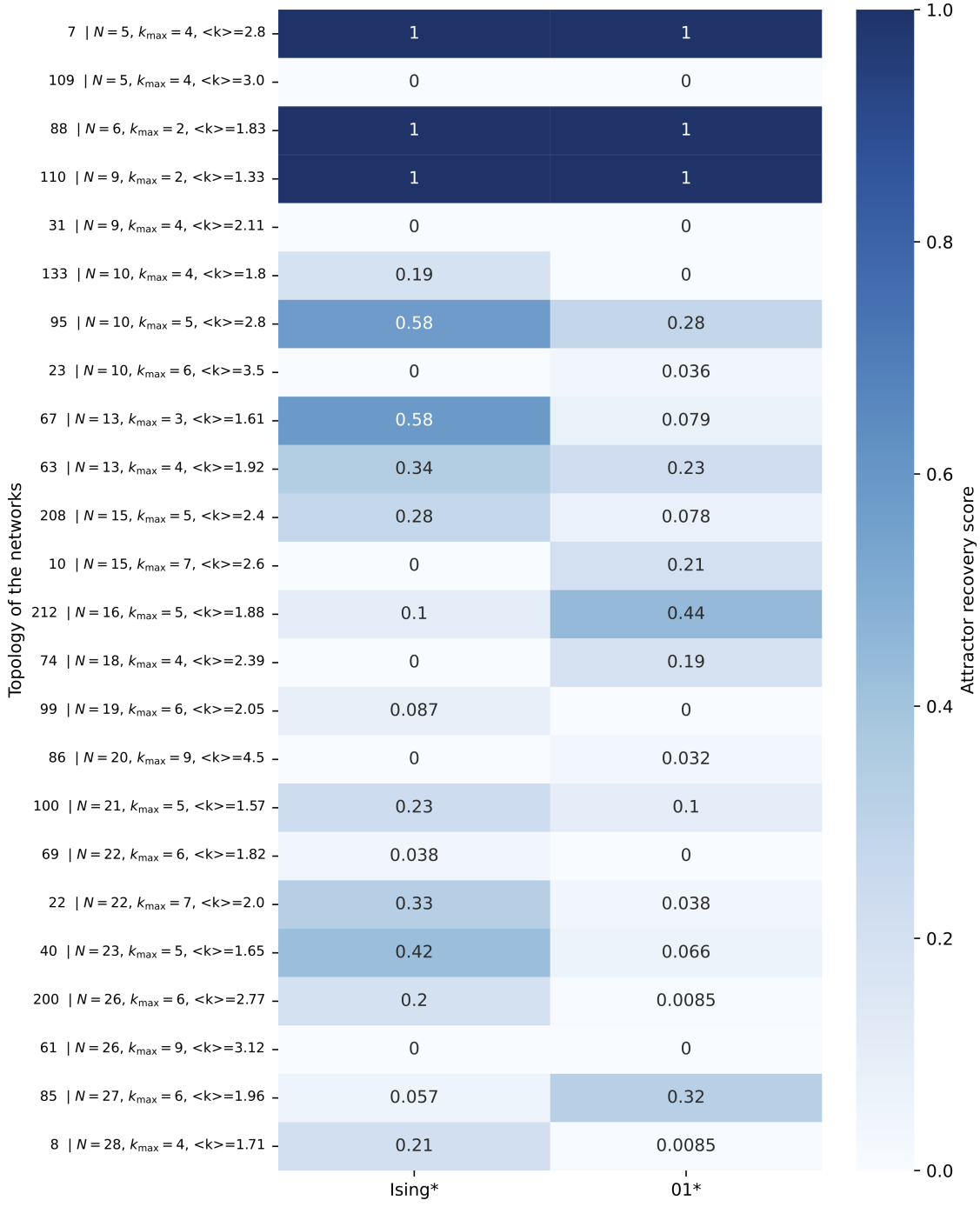


Figure C.6: Modified TMR-based (Ising* and 01*) models generally fail to recover biological attractors. This heatmap displays the attractor recovery scores for the alternative models obtained by replacing all BFs in 24 published BBM models with sign-conforming Ising* or 01* variants (as proposed in [190]). Each row corresponds to a specific network from the BBM dataset. On the far left, the number indicates the network ID as assigned in the BBM dataset. The next columns list the number of nodes (N), the maximum in-degree ($k_{\max}$), and the average in-degree ($\langle k \rangle$) of the network. A score near 0 indicates poor recovery of biological attractors, while higher scores indicate better performance. For most networks, both Ising* and 01*-based models show scores near 0.

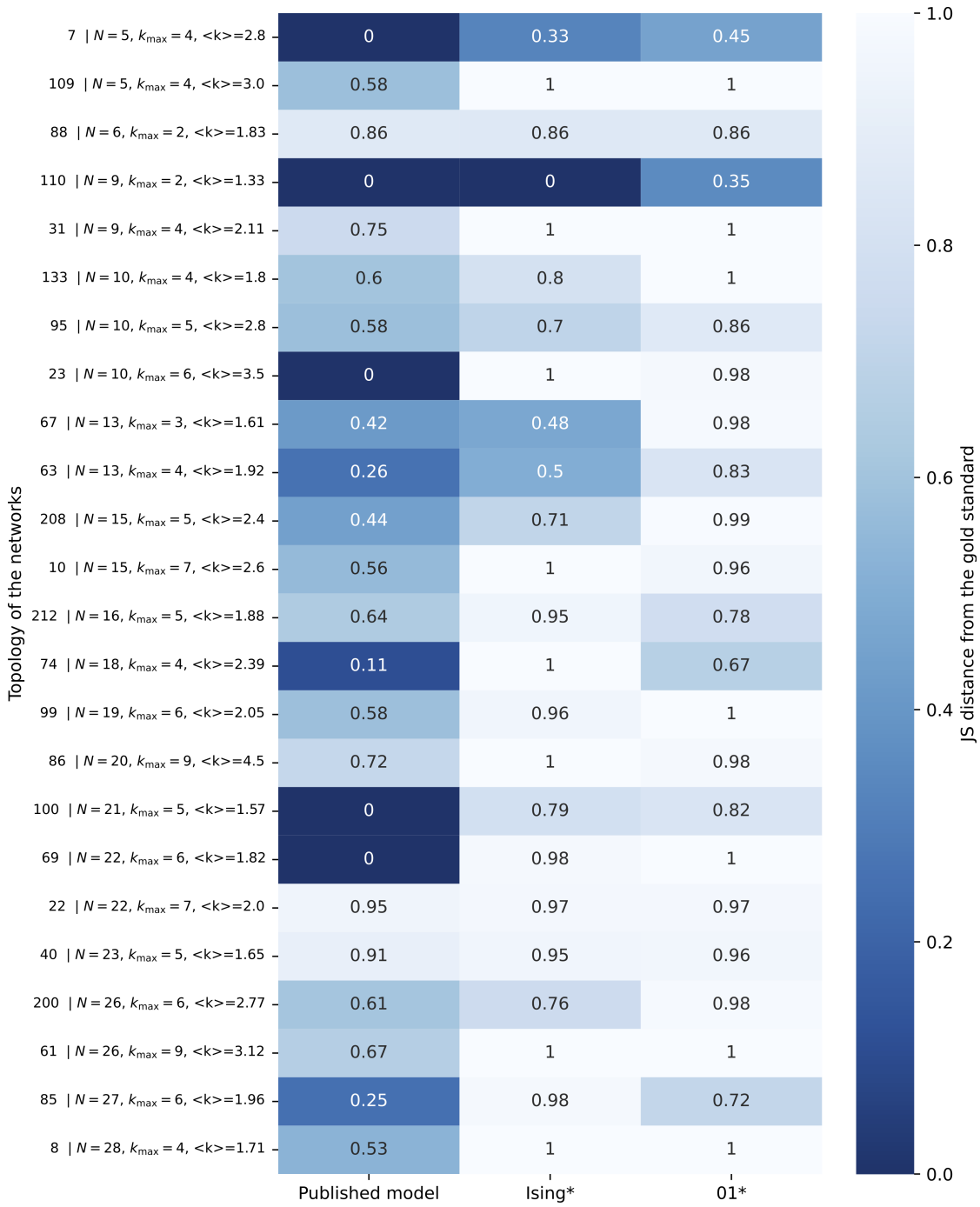


Figure C.7: JS distance from gold standard reveals poor preservation of basin size distributions of biological attractors by modified TMR-based (Ising* and 01*) models. The heatmap shows the Jensen–Shannon (JS) distance between the basin fraction distributions of the 24 published BBM models and their Ising* and 01* variants (as proposed in [190]) relative to the corresponding gold standards. Each row corresponds to a specific network from the BBM dataset. On the far left, the number indicates the network ID as assigned in the BBM dataset. The next columns list the number of nodes (N), the maximum in-degree ($k_{\max}$), and the average in-degree ($\langle k \rangle$) of the network. A JS distance of 0 indicates perfect overlap with the gold standard, while values closer to 1 reflect increasingly divergent distributions. Across most networks, Ising*- and 01*-based models typically exhibit high JS distances, indicating poor preservation of the expected basin size distribution of the biological attractors.

References

[1] Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC., 2006).

[2] McAdams, H. H. & Shapiro, L. Circuit Simulation of Genetic Networks. *Science* **269**, 650–656 (1995).

[3] Kitano, H. Biological robustness. *Nature Reviews Genetics* **5**, 826–837 (2004).

[4] Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).

[5] Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).

[6] Maslov, S. & Sneppen, K. Specificity and Stability in Topology of Protein Networks. *Science* **296**, 910–913 (2002).

[7] Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences* **100**, 14796–14799 (2003).

[8] Barabási, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101–113 (2004).

[9] Albert, R. Scale-free networks in cell biology. *Journal of Cell Science* **118**, 4947–4957 (2005).

[10] Ciliberti, S., Martin, O. C. & Wagner, A. Innovation and robustness in complex regulatory gene networks. *Proceedings of the National Academy of Sciences* **104**, 13591–13596 (2007).

[11] Macneil, L. T. & Walhout, A. J. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* **21**, 645–657 (2011).

[12] Heltberg, M. L., Krishna, S. & Jensen, M. H. On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. *Nature Communications* **10**, 71 (2019).

[13] Nagata, S. & Kikuchi, M. Emergence of cooperative bistability and robustness of gene regulatory networks. *PLOS Computational Biology* **16**, 1–24 (2020).

[14] de Jong, H. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology* **9**, 67–103 (2002).

[15] Tyson, J. J., Chen, K. & Novak, B. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology* **2**, 908–916 (2001).

[16] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340–2361 (1977).

[17] Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186 (2002).

[18] Paulsson, J. Models of stochastic gene expression. *Physics of Life Review* **2**, 157–175 (2005).

[19] Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22**, 437–467 (1969).

[20] Thomas, R. Regulatory networks seen as asynchronous automata: A logical description. *Journal of Theoretical Biology* **153**, 1–23 (1991).

[21] Thieffry, D. & Romero, D. The modularity of biological regulatory networks. *Biosystems* **50**, 49–59 (1999).

[22] Chaouiya, C. Petri net modelling of biological networks. *Briefings in Bioinformatics* **8**, 210–219 (2007).

[23] Fauré, A., Naldi, A., Chaouiya, C. & Thieffry, D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22**, e124–e131 (2006).

[24] Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318–356 (1961).

[25] Sugita, M. Functional analysis of chemical systems in vivo using a logical circuit equivalent. II. The idea of a molecular automaton. *Journal of Theoretical Biology* **4**, 179–192 (1963).

[26] Sugita, M. Functional analysis of chemical systems in vivo using a logical circuit equivalent IV. Simulation of cellular control systems using a hybrid computing system. *Journal of Theoretical Biology* **13**, 330–356 (1966).

[27] Kauffman, S. Homeostasis and Differentiation in Random Genetic Control Networks. *Nature* **224**, 177–178 (1969).

[28] Thomas, R. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology* **42**, 563–585 (1973).

[29] Thomas, R. *Kinetic logic: a Boolean approach to the analysis of complex regulatory systems, Proceedings of the EMBO course "Formal analysis of genetic regulation", held in Brussels, September 6–16, 1977, Lecture notes in Biomathematics* (Springer-Verlag, 1979).

[30] Thomas, R. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In *Numerical Methods in the Study of Critical Phenomena: Proceedings of a Colloquium, Carry-le-Rouet, France, June 2–4, 1980*, 180–193 (Springer, 1981).

[31] Mendoza, L., Thieffry, D. & Alvarez-Buylla, E. R. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics* **15**, 593–606 (1999).

[32] Albert, R. & Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *Journal of Theoretical Biology* **223**, 1–18 (2003).

[33] Samal, A. & Jain, S. The regulatory network of E. coli metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Systems Biology* **2**, 1–18 (2008).

[34] Davidich, M. I. & Bornholdt, S. Boolean Network Model Predicts Cell Cycle Sequence of Fission Yeast. *PLOS ONE* **3**, 1–8 (2008).

[35] Sahin, O. *et al.* Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC systems biology* **3**, 1–20 (2009).

[36] Saadatpour, A., Wang, R. S., Liao, A., Liu, X., Loughran, T. P., Albert, I. & Albert, R. Dynamical and Structural Analysis of a T Cell Survival Network Identifies Novel Candidate Therapeutic Targets for Large Granular Lymphocyte Leukemia. *PLOS Computational Biology* **7**, 1–15 (2011).

[37] Krumsiek, J., Marr, C., Schroeder, T. & Theis, F. J. Hierarchical Differentiation of Myeloid Progenitors Is Encoded in the Transcription Factor Network. *PLOS ONE* **6**, 1–10 (2011).

[38] García-Gómez, M. L., Azpeitia, E. & Álvarez-Buylla, E. R. A dynamic genetic-hormonal regulatory network model explains multiple cellular behaviors of the root apical meristem of Arabidopsis thaliana. *PLOS Computational Biology* **13**, 1–36 (2017).

[39] Hérault, L., Poplineau, M., Duprez, E. & Remy, É. A novel Boolean network inference strategy to model early hematopoiesis aging. *Computational and Structural Biotechnology Journal* **21**, 21–33 (2023).

[40] Bornholdt, S. Boolean network models of cellular regulation: prospects and limitations. *Journal of the Royal Society Interface* **5**, S85–S94 (2008).

[41] Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Genetic networks with canalyzing Boolean rules are always stable. *Proceedings of the National Academy of Sciences* **101**, 17102–17107 (2004).

[42] Murrugarra, D. & Laubenbacher, R. Regulatory patterns in molecular interaction networks. *Journal of Theoretical Biology* **288**, 66–72 (2011).

[43] Zhao, Y., Dong, H. & Ni, H. Categorizing Attractor-Effective Canalyzing Functions in Boolean Networks. *IFAC-PapersOnLine* **50**, 14746–14751 (2017).

[44] Palsson, B. Ø. *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, 2006).

[45] Wagner, A. *Robustness and evolvability in living systems* (Princeton University Press, Princeton, 2013).

[46] Marku, M. & Pancaldi, V. From time-series transcriptomics to gene regulatory networks: A review on inference methods. *PLOS Computational Biology* **19**, 1–26 (2023).

[47] Pušnik, Z., Mraz, M., Zimic, N. & Moškon, M. Review and assessment of Boolean approaches for inference of gene regulatory networks. *Heliyon* **8**, e10222 (2022).

[48] Galindez, G., Sadegh, S., Baumbach, J., Kacprowski, T. & List, M. Network-based approaches for modeling disease regulation and progression. *Computational and Structural Biotechnology Journal* **21**, 780–795 (2023).

[49] Helikar, T. *et al.* The Cell Collective: Toward an open and collaborative approach to systems biology. *BMC Systems Biology* **6**, 96 (2012).

[50] Gonzalez, A. G., Naldi, A., Sánchez, L., Thieffry, D. & Chaouiya, C. GINsim: A software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems* **84**, 91–100 (2006).

[51] Subbaroyan, A., Martin, O. C. & Samal, A. Minimum complexity drives regulatory logic in Boolean models of living systems. *PNAS Nexus* **1**, pgac017 (2022).

[52] Pastva, S., Šafránek, D., Beneš, N., Brim, L. & Henzinger, T. Repository of logically consistent real-world Boolean network models. *bioRxiv 2023.06.12.544361* (2023).

[53] Daniels, B. C., Kim, H., Moore, D., Zhou, S., Smith, H. B., Karas, B., Kauffman, S. A. & Walker, S. I. Criticality Distinguishes the Ensemble of Biological Regulatory Networks. *Physical Review Letters* **121**, 138102 (2018).

[54] Tripathi, S., Kessler, D. A. & Levine, H. Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proceedings of the National Academy of Sciences* **120**, e2216109120 (2023).

[55] Aracena, J. Maximum Number of Fixed Points in Regulatory Boolean Networks. *Bulletin of Mathematical Biology* **70**, 1398–1409 (2008).

[56] Kadelka, C., Butrie, T. M., Hilton, E., Kinseth, J., Schmidt, A. & Serdarevic, H. A meta-analysis of Boolean network models reveals design principles of gene regulatory networks. *Science Advances* **10**, eadj0822 (2024).

[57] Feldman, J. Minimization of Boolean complexity in human concept learning. *Nature* **407**, 630–633 (2000).

[58] Shmulevich, I. & Kauffman, S. A. Activities and Sensitivities in Boolean Network Models. *Physical Review Letters* **93**, 048701 (2004).

[59] Derrida, B. & Stauffer, D. Phase Transitions in Two-Dimensional Kauffman Cellular Automata. *Europhysics Letters* **2**, 739 (1986).

[60] Park, K. H., Costa, F. X., Rocha, L. M., Albert, R. & Rozum, J. C. Models of Cell Processes are Far from the Edge of Chaos. *PRX Life* **1**, 023009 (2023).

[61] Wuensche, A. The Ghost in the Machine: Basins of Attraction of Random Boolean Networks. vol. 17, 465–501 (Addison-Wesley, 1994).

[62] Wuensche, A. Complexity in One-D Cellular Automata: Gliders, Basins of Attraction and the Z Parameter. Working Paper 94-04-025, Santa Fe Institute (1994).

[63] Wuensche, A. *Attractor Basins of Discrete Networks*. D.Phil, The University of Sussex (1997).

[64] Subbaroyan, A., Sil, P., Martin, O. C. & Samal, A. Leveraging developmental landscapes for model selection in Boolean gene regulatory networks. *Briefings in Bioinformatics* **24**, bbad160 (2023).

[65] Font-Clos, F., Zapperi, S. & La Porta, C. A. M. Topography of epithelial–mesenchymal plasticity. *Proceedings of the National Academy of Sciences* **115**, 5902–5907 (2018).

[66] Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A. & Jolly, M. K. Landscape of epithelial–mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. *eLife* **11**, e76535 (2022).

[67] Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences* **101**, 4781–4786 (2004).

[68] Jarrah, A. S., Raposa, B. & Laubenbacher, R. Nested canalyzing, unate cascade, and polynomial functions. *Physica D: Nonlinear Phenomena* **233**, 167–174 (2007).

[69] Li, Y., Adeyeye, J. O., Murrugarra, D., Aguilar, B. & Laubenbacher, R. Boolean nested canalizing functions: A comprehensive analysis. *Theoretical Computer Science* **481**, 24–36 (2013).

[70] McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* **5**, 115–133 (1943).

[71] Chen, F. & Chen, G. REALIZATION AND BIFURCATION OF BOOLEAN FUNCTIONS VIA CELLULAR NEURAL NETWORKS. *International Journal of Bifurcation and Chaos* **15**, 2109–2129 (2005).

[72] Ruz, G. A., Goles, E. & Sené, S. Reconstruction of Boolean Regulatory Models of Flower Development Exploiting an Evolution Strategy. In *2018 IEEE congress on Evolutionary Computation (CEC)*, 1–8 (IEEE, 2018).

[73] Wagner, A. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proceedings of the National Academy of Sciences* **91**, 4387–4391 (1994).

[74] Wagner, A. DOES EVOLUTIONARY PLASTICITY EVOLVE? *Evolution* **50**, 1008–1023 (1996).

[75] Mendoza, L. & Alvarez-Buylla, E. R. Dynamics of the Genetic Regulatory Network forArabidopsis thalianaFlower Morphogenesis. *Journal of Theoretical Biology* **193**, 307–319 (1998).

[76] Manicka, S., Johnson, K., Levin, M. & Murrugarra, D. The nonlinearity of regulation in biological networks. *npj Systems Biology and Applications* **9**, 10 (2023).

[77] Sil, P., Mitra, S., Martin, O. C. & Samal, A. Choices of regulatory logic class modulate the dynamical regime in random Boolean networks. *Chaos, Solitons & Fractals* **195**, 116231 (2025).

[78] Wuensche, A. Classifying cellular automata automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter. *Complexity* **4**, 47–66 (1999).

[79] Sil, P., Subbaroyan, A., Kulkarni, S., Martin, O. C. & Samal, A. Biologically meaningful regulatory logic enhances the convergence rate in Boolean networks and bushiness of their state transition graph. *Briefings in Bioinformatics* **25**, bbae150 (2024).

[80] Sil, P., Martin, O. C. & Samal, A. A systematic evaluation of simple threshold-based Boolean rules for modeling gene regulatory networks. *Journal of Statistical Mechanics: Theory and Experiment* 023502 (2026).

[81] Raeymaekers, L. Dynamics of Boolean Networks Controlled by Biologically Meaningful Functions. *Journal of Theoretical Biology* **218**, 331–341 (2002).

[82] Garg, A., Di Cara, A., Xenarios, I., Mendoza, L. & De Micheli, G. Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* **24**, 1917–1925 (2008).

[83] Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraula, G. & Bonabeau, E. *Self-Organization in Biological Systems* (Princeton University Press, Princeton, 2001).

[84] Kitano, H. Computational systems biology. *Nature* **420**, 206–210 (2002).

[85] Kaneko, K. *Life: An Introduction to Complex Systems Biology*. Understanding Complex Systems (Springer Berlin Heidelberg, 2006).

[86] Von Neumann, J. & Burks, A. W. *Theory of self-reproducing automata* (University of Illinois press Urbana, 1966).

[87] Wolfram, S. Statistical mechanics of cellular automata. *Reviews of Modern Physics* **55**, 601–644 (1983).

[88] Langton, C. G. Studying artificial life with cellular automata. *Physica D: Nonlinear Phenomena* **22**, 120–149 (1986).

[89] Ermentrout, G. B. & Edelstein-Keshet, L. Cellular Automata Approaches to Biological Modeling. *Journal of Theoretical Biology* **160**, 97–133 (1993).

[90] Martin, O., Odlyzko, A. M. & Wolfram, S. Algebraic properties of cellular automata. *Communications in Mathematical Physics* **93**, 219–258 (1984).

[91] Wuensche, A. & Lesser, M. *The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata* (Addison-Wesley, Reading, MA, 1992).

[92] Brooks, R. A. & Maes, P. *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (The MIT Press, 1994).

[93] Mortveit, H. S. & Reidys, C. M. *An Introduction to Sequential Dynamical Systems* (Springer-Verlag, Berlin, Heidelberg, 2007).

[94] Bilke, S. & Sjunnesson, F. Stability of the Kauffman model. *Physical Review E* **65**, 016129 (2001).

[95] Samuelsson, B. & Troein, C. Superpolynomial Growth in the Number of Attractors in Kauffman Networks. *Physical Review Letters* **90**, 098701 (2003).

[96] Mihaljev, T. & Drossel, B. Scaling in a general class of critical random Boolean networks. *Physical Review E* **74**, 046101 (2006).

[97] Saadatpour, A. & Albert, R. Boolean modeling of biological regulatory networks: a methodology tutorial. *Methods* **62**, 3–12 (2013).

[98] Han, H. *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research* **46**, D380–D386 (2017).

[99] Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **52**, D174–D182 (2023).

[100] Wingender, E. *et al.* TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Research* **28**, 316–319 (2000).

[101] Liang, S., Fuhrman, S. & Somogyi, R. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. In *Pacific Symposium on Biocomputing*, vol. 3, 18–29 (1998).

[102] Lähdesmäki, H., Shmulevich, I. & Yli-Harja, O. On Learning Gene Regulatory Networks Under the Boolean Network Model. *Machine learning* **52**, 147–167 (2003).

[103] Barman, S. & Kwon, Y. K. A Boolean network inference from time-series gene expression data using a genetic algorithm. *Bioinformatics* **34**, i927–i933 (2018).

[104] Guziolowski, C., Videla, S., Eduati, F., Thiele, S., Cokelaer, T., Siegel, A. & Saez-Rodriguez, J. Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming. *Bioinformatics* **29**, 2320–2326 (2013).

[105] Gjerga, E. *et al.* Converting networks to predictive logic models from perturbation signalling data with CellNOpt. *Bioinformatics* **36**, 4523–4524 (2020).

[106] Albert, R. & Thakar, J. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *WIREs Systems Biology and Medicine* **6**, 353–369 (2014).

[107] Calzone, L., Tournier, L., Fourquet, S., Thieffry, D., Zhivotovsky, B., Barillot, E. & Zinovyev, A. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLOS Computational Biology* **6**, 1–15 (2010).

[108] Golumbic, M. C., Gurvich, V., Crama, Y. & Hammer, P. L. Read-once functions. In *Boolean Functions: Theory, Algorithms, and Applications*, 448–486 (Cambridge University Press, 2011).

[109] Kadelka, C., Kuipers, J. & Laubenbacher, R. The influence of canalization on the robustness of Boolean networks. *Physica D: Nonlinear Phenomena* **353-354**, 39–47 (2017).

[110] Mendoza, L. & Xenarios, I. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theoretical Biology and Medical Modelling* **3**, 13 (2006).

[111] Subbaroyan, A., Martin, O. C. & Samal, A. A preference for link operator functions can drive Boolean biological networks towards critical dynamics. *Journal of Biosciences* **47**, 17 (2022).

[112] Zobolas, J., Monteiro, P. T., Kuiper, M. & Flobak, Å. Boolean function metrics can assist modelers to check and choose logical rules. *Journal of Theoretical Biology* **538**, 111025 (2022).

[113] Wegener, I. *The Complexity of Boolean Functions* (John Wiley & Sons, Inc., New York, 1987).

[114] Givone, D. D. *Introduction to Switching Circuit Theory*. McGraw-Hill Computer Science Series (McGraw-Hill, New York, 1970).

[115] Garey, M. R. & Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-completeness* (W. H. Freeman & Co., New York, 1979).

[116] Buchfuhrer, D. & Umans, C. The complexity of Boolean formula minimization. *Journal of Computer and System Sciences* **77**, 142–153 (2011).

[117] Quine, W. V. A Way to Simplify Truth Functions. *The American Mathematical Monthly* **62**, 627–631 (1955).

[118] McCluskey Jr., E. J. Minimization of Boolean functions. *Bell System Technical Journal* **35**, 1417–1444 (1956).

[119] Brayton, R. & Mishchenko, A. ABC: An Academic Industrial-Strength Verification Tool. In *Computer Aided Verification*, 24–40. Springer (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).

[120] Schwab, J. D., Kühlwein, S. D., Ikonomi, N., Kühl, M. & Kestler, H. A. Concepts in Boolean network modeling: What do they all mean? *Computational and Structural Biotechnology Journal* **18**, 571–582 (2020).

[121] Derrida, B. & Pomeau, Y. Random networks of automata: A simple annealed approximation. *Europhysics Letters* **1**, 45 (1986).

[122] Derrida, B. & Weisbuch, G. Evolution of overlaps between configurations in random Boolean networks. *Journal de physique* **47**, 1297–1303 (1986).

[123] Müssel, C., Hopfensitz, M. & Kestler, H. A. BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* **26**, 1378–1380 (2010).

[124] Akutsu, T., Kuhara, S., Maruyama, O. & Miyano, S. A System for Identifying Genetic Networks from Gene Expression Patterns Produced by Gene Disruptions and Overexpressions. *Genome Informatics* **9**, 151–160 (1998).

[125] Glauber, R. J. Time-Dependent Statistics of the Ising Model. *Journal of Mathematical Physics* **4**, 294–307 (1963).

[126] Wuensche, A. Cellular Automata Encryption: the reverse algorithm, Z-parameter and chain-rules. *Parallel Processing Letters* **19**, 283–297 (2009).

[127] Walker, C. C. & Ashby, W. R. On temporal characteristics of behavior in certain complex systems. *Kybernetik* **3**, 100–108 (1966).

[128] Langton, C. G. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: Nonlinear Phenomena* **42**, 12–37 (1990).

[129] Paul, E., Pogudin, G., Qin, W. & Laubenbacher, R. The Dynamics of Canalizing Boolean Networks. *Complexity* **2020**, 3687961 (2020).

[130] Bollobás, B. *Random graphs* (Springer, New York, 1998).

[131] Harary, F. *Graph Theory (on Demand Printing of 02787)* (CRC Press, 2018).

[132] Kullback, S. & Leibler, R. A. On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86 (1951).

[133] Rao, C. R. Diversity: Its Measurement, Decomposition, Apportionment and Analysis. *Sankhyā: The Indian Journal of Statistics* **44**, 1–21 (1982).

[134] Lin, J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* **37**, 145–151 (1991).

[135] Reichhardt, C. J. O. & Bassler, K. E. Canalization and symmetry in Boolean models for genetic regulatory networks. *Journal of Physics A: Mathematical and Theoretical* **40**, 4339–4350 (2007).

[136] Kadelka, C. & Murrugarra, D. Canalization reduces the nonlinearity of regulation in biological networks. *npj Systems Biology and Applications* **10**, 67 (2024).

[137] Huang, S. & Ingber, D. E. Shape-Dependent Control of Cell Growth, Differentiation, and Apoptosis: Switching between Attractors in Cell Regulatory Networks. *Experimental Cell Research* **261**, 91–103 (2000).

[138] Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters* **94**, 128701 (2005).

[139] Wolfram, S. *A New Kind of Science* (Wolfram Media, 2002).

[140] Aldana, M. Boolean dynamics of networks with scale-free topology. *Physica D: Nonlinear Phenomena* **185**, 45–66 (2003).

[141] Serra, R., Villani, M., Barbieri, A., Kauffman, S. & Colacci, A. On the dynamics of random Boolean networks subject to noise: Attractors, ergodic sets and cell types. *Journal of Theoretical Biology* **265**, 185–193 (2010).

[142] Henry, A., Monéger, F., Samal, A. & Martin, O. C. Network function shapes network structure: the case of the *Arabidopsis* flower organ specification genetic network. *Molecular BioSystems* **9**, 1726–1735 (2013).

[143] Zañudo, J. G. T. & Albert, R. An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**, 025111 (2013).

[144] Gates, A. J., Correia, R. B., Wang, X. & Rocha, L. M. The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation and signaling. *Proceedings of the National Academy of Sciences* **118** (2021).

[145] Drossel, B., Mihaljev, T. & Greil, F. Number and Length of Attractors in a Critical Kauffman Model with Connectivity One. *Physical Review Letters* **94**, 088701 (2005).

[146] Klemm, K. & Bornholdt, S. Stable and unstable attractors in Boolean networks. *Physical Review E* **72**, 055101 (2005).

[147] Roli, A., Villani, M., Filisetti, A. & Serra, R. Dynamical Criticality: Overview and Open Questions. *Journal of Systems Science and Complexity* **31**, 647–663 (2018).

[148] Zhou, J. X., Samal, A., d'Hérouël, A. F., Price, N. D. & Huang, S. Relative stability of network states in Boolean network models of gene regulation in development. *Biosystems* **142-143**, 15–24 (2016).

[149] Azpeitia, E., Benítez, M., Vega, I., Villarreal, C. & Alvarez-Buylla, E. R. Single-cell and coupled GRN models of cell patterning in the Arabidopsis thaliana root stem cell niche. *BMC Systems Biology* **4**, 134 (2010).

[150] Sullivan, E., Harris, M., Bhatnagar, A., Guberman, E., Zonfa, I. & Ravasz Regan, E. Boolean modeling of mechanosensitive epithelial to mesenchymal transition and its reversal. *iScience* **26**, 106321 (2023).

[151] Giacomantonio, C. E. & Goodhill, G. J. A Boolean Model of the Gene Regulatory Network Underlying Mammalian Cortical Area Development. *PLOS Computational Biology* **6**, 1–13 (2010).

[152] Herrmann, F., Groß, A., Zhou, D., Kestler, H. A. & Kühl, M. A Boolean Model of the Cardiac Gene Regulatory Network Determining First and Second Heart Field Identity. *PLOS ONE* **7**, 1–10 (2012).

[153] Biane, C. & Delaplace, F. Causal reasoning on boolean control networks based on abduction: theory and application to cancer drug discovery. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **16**, 1574–1585 (2019).

[154] Narváez-Chávez, J. E., Álvarez Buylla, E. R. & Martínez-García, J. C. Uncovering the role of mutations in Epithelial-to-Mesenchymal transition through computational analysis of the underlying gene regulatory network. In Al-Mubaid, H., Aldwairi, T. & Eulenstein, O. (eds.) *Proceedings of International Conference on Bioinformatics and Computational Biology (BICOB-2023)*, vol. 92, 92–101 (EasyChair, 2023).

[155] Sánchez-Corrales, Y., Álvarez Buylla, E. R. & Mendoza, L. The Arabidopsis thaliana flower organ specification gene regulatory network determines a robust differentiation process. *Journal of Theoretical Biology* **264**, 971–983 (2010).

[156] Ríos, O., Frias, S., Rodríguez, A., Kofman, S., Merchant, H., Torres, L. & Mendoza, L. A Boolean network model of human gonadal sex determination. *Theoretical Biology and Medical Modelling* **12**, 26 (2015).

[157] Wuensche, A. Basins of attraction in network dynamics: A conceptual framework for biomolecular networks. In *Modularity in Development and Evolution*, 288–311 (Chicago University Press, 2004).

[158] Wuensche, A. Discrete dynamical networks and their attractor basins. In *Complex Systems '98* (University of New South Wales, Sydney, Australia, 1998).

[159] Gershenson, C. Guiding the self-organization of random Boolean networks. *Theory in Biosciences* **131**, 181–191 (2012).

[160] Sutner, K. On the Computational Complexity of Finite Cellular Automata. *Journal of Computer and System Sciences* **50**, 87–97 (1995).

[161] Tošić, P. T. & Ordonez, C. Boolean Network Models of Collective Dynamics of Open and Closed Large-Scale Multi-agent Systems. In Mařík, V., Wahlster, W., Strasser, T. & Kadera, P. (eds.) *Industrial Applications of Holonic and Multi-Agent Systems*, 95–110 (Springer International Publishing, Cham, 2017).

[162] Muroga, S. *Threshold Logic and Its Applications* (Wiley, New York, 1971).

[163] Kürten, K. Correspondence between neural threshold networks and Kauffman Boolean cellular automata. *Journal of Physics A: Mathematical and General* **21**, L615 (1988).

[164] Kürten, K. E. Critical phenomena in model neural networks. *Physics Letters A* **129**, 157–160 (1988).

[165] Kohavi, Z. & Jha, N. K. *Switching and Finite Automata Theory* (Cambridge University Press, Cambridge, 2009).

[166] Bornholdt, S. & Sneppen, K. Robustness as an evolutionary principle. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 2281–2286 (2000).

[167] Sevim, V. & Rikvold, P. A. Chaotic gene regulatory networks can be robust against mutations and noise. *Journal of Theoretical Biology* **253**, 323–332 (2008).

[168] Li, L., Sun, L., Chen, G., Wong, C. W., Ching, W. K. & Liu, Z. P. LogBTF: gene regulatory network inference using Boolean threshold network model from single-cell gene expression data. *Bioinformatics* **39**, btad256 (2023).

[169] Chauhan, L., Ram, U., Hari, K. & Jolly, M. K. Topological signatures in regulatory network enable phenotypic heterogeneity in small cell lung cancer. *eLife* **10**, e64522 (2021).

[170] Tripathi, S., Kessler, D. A. & Levine, H. Biological Networks Regulating Cell Fate Choice Are Minimally Frustrated. *Physical Review Letters* **125**, 088101 (2020).

[171] Claussen, J. C. *et al.* Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. *PLOS Computational Biology* **13**, 1–21 (2017).

[172] Jyoti, J. & Hütt, M. T. Evaluating changes in attractor sets under small network perturbations to infer reliable microbial interaction networks from abundance patterns. *Bioinformatics* **41**, btaf095 (2025).

[173] Derrida, B. Dynamical phase transition in nonsymmetric spin glasses. *Journal of Physics A: Mathematical and General* **20**, L721 (1987).

[174] Yao, Y., Huang, Z. G. & Pei, D. Diversified dynamic effects and their order origins in Boolean functions. *Chaos, Solitons & Fractals* **191**, 115830 (2025).

[175] He, Q., Chen, F. & Jin, W. Topological Equivalence Classification of Balanced Linearly Separable Boolean Functions on n-Dimensional Hypercube. *International Journal of Bifurcation and Chaos* **31**, 2150031 (2021).

[176] Muroga, S., Tsuboi, T. & Baugh, C. R. Enumeration of Threshold Functions of Eight Variables. *IEEE Transactions on Computers* **C-19**, 818–825 (1970).

[177] Coates, C. L. & Lewis II, P. M. Linearly separable switching functions. *Journal of the Franklin Institute* **272**, 366–410 (1961).

[178] Gowda, T., Vrudhula, S. & Kim, S. Modeling of gene regulatory network dynamics using threshold logic. *Annals of the New York Academy of Sciences* **1158**, 71–81 (2009).

[179] He, Q., Chen, F. & Jin, W. Topological equivalence classification and enumeration of n-input linearly separable Boolean functions. *Information Sciences* **658**, 120020 (2024).

[180] Lozin, V., Razgon, I., Zamaraev, V., Zamaraeva, E. & Zolotykh, N. Linear read-once and related Boolean functions. *Discrete Applied Mathematics* **250**, 16–27 (2018).

[181] Martínez-Antonio, A., Janga, S. C. & Thieffry, D. Functional organisation of Escherichia coli transcriptional regulatory network. *Journal of Molecular Biology* **381**, 238–247 (2008).

[182] Hermsen, R., Ursem, B. & ten Wolde P. R. Combinatorial gene regulation using auto-regulation. *PLOS Computational Biology* **6**, 1–13 (2010).

[183] Feldman, J. A catalog of Boolean concepts. *Journal of Mathematical Psychology* **47**, 75–89 (2003).

[184] Naldi, A., Hernandez, C., Abou-Jaoudé, W., Monteiro, P. T., Chaouiya, C. & Thieffry, D. Logical Modeling and Analysis of Cellular Regulatory Networks With GINsim 3.0. *Frontiers in physiology* **9**, 646 (2018).

[185] Malik-Sheriff, R. S. *et al.* BioModels—15 years of sharing computational models in life science. *Nucleic Acids Research* **48**, D407–D415 (2019).

[186] Ostaszewski, M. *et al.* COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Scientific Data* **7**, 136 (2020).

[187] Mitra, S., Sil, P., Subbaroyan, A., Martin, O. C. & Samal, A. Preponderance of generalized chain functions in reconstructed Boolean models of biological networks. *Scientific Reports* **14**, 6734 (2024).

[188] Szejka, A., Mihaljev, T. & Drossel, B. The phase diagram of random threshold networks. *New Journal of Physics* **10**, 063009 (2008).

[189] Zañudo, J. G. T., Aldana, M. & Martínez-Mekler, G. *Boolean Threshold Networks: Virtues and Limitations for Biological Modeling*, 113–151 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011).

[190] Kadelka, C. & Hari, K. Critical assessment of the ability of Boolean threshold models to describe gene regulatory network dynamics. *bioRxiv 2025.03.06.641948* (2025).

[191] Kadelka, C. & Hari, K. Critical assessment of the ability of Boolean threshold models to describe gene regulatory network dynamics. *PNAS Nexus* **4**, pgaf228 (2025).

[192] Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences* **100**, 5136–5141 (2003).

[193] Badia-i Mompel, P., Wessels, L., Müller-Dott, S., Trimbour, R., Ramirez Flores, R. O., Argelaguet, R. & Saez-Rodriguez, J. Gene regulatory network inference in the era of single-cell multi-omics. *Nature Reviews Genetics* **24**, 739–754 (2023).

[194] Dong, J., Li, J. & Wang, F. Deep learning in gene regulatory network inference: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **21**, 2089–2101 (2024).

[195] Davila-Velderrain, J., Caldu-Primo, J. L., Martinez-Garcia, J. C. & Alvarez-Buylla, E. R. Modeling the Epigenetic Landscape in Plant Development. In *Computational Cell Biology: Methods and Protocols*, 357–383 (Springer New York, New York, NY, 2018).

[196] Nikolajewa, S., Friedel, M. & Wilhelm, T. Boolean networks with biologically relevant rules show ordered behavior. *Biosystems* **90**, 40–47 (2007).

[197] Gallian, J. *Contemporary Abstract Algebra* (Chapman and Hall/CRC, 2021).

[198] Clote, P. & Kranakis, E. Boolean Functions, Invariance Groups, and Parallel Complexity. *SIAM Journal on Computing* **20**, 553–590 (1991).

[199] Lucas, É. Sur les congruences des nombres eulériens et des coefficients différentiels des fonctions trigonométriques suivant un module premier. *Bulletin de la Société mathématique de France* **6**, 49–54 (1878).

[200] Wu, C. W. Sums of products of binomial coefficients mod 2 and run length transforms of sequences. *Integers* **22**, A81 (2022).

[201] Zhang, S. Note on the average sensitivity of monotone Boolean functions. *preprint* 4 (2011).

[202] Biswas, A. & Sarkar, P. Counting Unate and Monotone Boolean Functions Under Restrictions of Balancedness and Non-Degeneracy. *Journal of Integer Sequences* **28**, 1–19 (2025).

[203] Kummer, E. E. Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen. *Journal für die reine und angewandte Mathematik* **1852**, 93–146 (1852).

[204] Straub, A., Moll, V. H. & T., A. The p-adic valuation of k-central binomial coefficients. *Acta Arithmetica* **140**, 31–42 (2009).

[205] Méndez, A. & Mendoza, L. A Network Model to Describe the Terminal Differentiation of B Cells. *PLOS Computational Biology* **12**, 1–26 (2016).

[206] Orlando, D. A., Lin, C. Y., Bernard, A., Wang, J. Y., Socolar, J. E. S., Iversen, E. S., Hartemink, A. J. & Haase, S. B. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* **453**, 944–947 (2008).

[207] Enciso, J., Mayani, H., Mendoza, L. & Pelayo, R. Modeling the Pro-inflammatory Tumor Microenvironment in Acute Lymphoblastic Leukemia Predicts a Breakdown of Hematopoietic-Mesenchymal Communication Networks. *Frontiers in Physiology* **7**, 349 (2016).

[208] Veliz-Cuba, A. & Stigler, B. Boolean Models Can Explain Bistability in the *lac* Operon. *Journal of Computational Biology* **18**, 783–794 (2011).

[209] Jenkins, A. & Macauley, M. Bistability and Asynchrony in a Boolean Model of the L-arabinose Operon in Escherichia coli. *Bulletin of Mathematical Biology* **79**, 1778–1795 (2017).

[210] Brandon, M., Howard, B., Lawrence, C. & Laubenbacher, R. Iron acquisition and oxidative stress response in aspergillus fumigatus. *BMC Systems Biology* **9**, 19 (2015).

[211] Guttula, P. K., Monteiro, P. T. & Gupta, M. K. A Boolean Logical model for Reprogramming of Testes-derived male Germline Stem Cells into Germline pluripotent stem cells. *Computer Methods and Programs in Biomedicine* **192**, 105473 (2020).

[212] Cohen, D. P. A., Martignetti, L., Robine, S., Barillot, E., Zinovyev, A. & Calzone, L. Mathematical Modelling of Molecular Pathways Enabling Tumour Cell Invasion and Migration. *PLOS Computational Biology* **11**, 1–29 (2015).

[213] Coolen, M., Thieffry, D., Drivenes, Ø., Becker, T. S. & Bally-Cuif, L. miR-9 Controls the Timing of Neurogenesis through the Direct Inhibition of Antagonistic Factors. *Developmental Cell* **22**, 1052–1064 (2012).

[214] Fauré, A. & Thieffry, D. Logical modelling of cell cycle control in eukaryotes: a comparative study. *Molecular BioSystems* **5**, 1569–1581 (2009).

[215] Wooten, D. J., Zañudo, J. G. T., Murrugarra, D., Perry, A. M., Dongari-Bagtzoglou, A., Laubenbacher, R., Nobile, C. J. & Albert, R. Mathematical modeling of the Candida albicans yeast to hyphal transition reveals novel control strategies. *PLOS Computational Biology* **17**, 1–28 (2021).

[216] Palma, A. *et al.* Integrating Patient-Specific Information into Logic Models of Complex Diseases: Application to Acute Myeloid Leukemia. *Journal of Personalized Medicine* **11**, 117 (2021).

[217] Sánchez-Osorio, I., Hernández-Martínez, C. A. & Martínez-Antonio, A. *Modeling Asymmetric Cell Division in Caulobacter crescentus Using a Boolean Logic Approach*, 1–21 (Springer International Publishing, Cham, 2017).

[218] Gupta, S., Silveira, D. A. & Mombach, J. C. M. ATM/miR-34a-5p axis regulates a p21-dependent senescence-apoptosis switch in non-small cell lung cancer: a Boolean model of G1/S checkpoint regulation. *FEBS Letters* **594**, 227–239 (2020).

[219] Deal, I., Macauley, M. & Davies, R. Boolean Models of the Transport, Synthesis, and Metabolism of Tryptophan in Escherichia coli. *Bulletin of Mathematical Biology* **85**, 29 (2023).