Biodivne Boolean Models: A Comprehensive Logical Modelling Benchmark

Samuel Pastva

Abstract

Recent years have seen emergence of a wide variety of powerful tools for computational analysis of logical models represented as Boolean networks. However, assessment of validity, efficiency and scalability of such tools requires a comprehensive benchmark set of Boolean networks that can be used to obtain comparable results for different tools.

At the moment, this need is largely served using databases of biological models such as CellCollective or GINsim database. However, these databases are more focused on human curated, biological aspects of the networks and are therefore limited in scope. Furthermore, the models in these databases are not available as a single dataset and often have to be manually obtained one by one.

Here, we introduce a comprehensive benchmark dataset that has been created by surveying the aforementioned databases, as well as a large body of other literature to obtain as many biologically motivated Boolean networks as possible. To make the dataset useful to a wide range of tool maintainers, we provide the models in different machine-readable formats and ensure all models are valid and consistent using an automated validation procedure. At the moment, the dataset comprises 145 networks.

1 Introduction

Logical models provide a very useful and simple framework for description of complex biological processes. The most common mechanism for describing executable logical models are Boolean networks. In recent years, we have seen a rapid development of new tools and algorithms for analysis of large Boolean networks. However, in many instances, it is hard to assess usefulness and scalability of such tools due to a lack of commonly recognised "benchmark dataset" of networks on which the tools can be compared.

This purpose is often served by models obtained from databases maintained by the authors of some of the larger modelling tools, such as CellCollective [10] or GINsim [24]. However, these models are often hard to obtain in bulk and have to be downloaded one by one. Additionally, authors often modify the models slightly, or assume non-standard values of inputs which prevents comparisons. Finally, these databases are far from comprehensive, so a wide range of models is often omitted.

As a result, most papers develop an ad hoc benchmark set that is often partially proprietary and hard or impossible to replicate and compare to. Here, we propose a standardized comprehensive benchmark set that can be used for this purpose instead. To make the benchmark set as user friendly as possible, we provide the following benefits compared to existing solutions:

- The dataset is open source and available on Github, so that anyone can propose new additions or modifications. Each tracked model is (primarily) referred to using a unique ID as opposed to name or citation. However, we also keep track of the original source (publication) where the model first appeared.
- Every model is provided in three formats that can be consumed by different tools or easily parsed by a new tool. Namely, we consider `bnet`, as popularised by PyBoolNet [16], `aeon` format as used in AEON [1], and the universal SBML-qual [3] format.
- If the model contains inputs (constants), aside from the model as published by the authors, we also generate two variants with all inputs fixed to `true` and `false`, and a variant where the values of all inputs are unspecified.
- For each model format and variant, we provide a single bundle with all available models that can be easily used for batch processing.
- We provide an automated procedure to check the validity and integrity of all included models, as well as generate different model bundles. This minimises possible user errors when adding new models.

This document then serves as a cumulative report of all the models included in the dataset and the sources of these models.

2 Models

ID	Name	Vars.	Inps.	Regs.	Source
001	SIGNALING IN MACROPHAGE ACTIVATION	302	19	533	[27, 10]
002	SIGNAL TRANSDUCTION IN FIBROBLASTS	130	9	557	[9, 10]
003	MAMMALIAN CELL CYCLE	19	1	51	[32, 10]
004	ERBB RECEPTOR SIGNALING	225	22	1100	[8, 10]
005	FA/BRCA PATHWAY	28	0	123	[28, 10]
006	HGF SIGNALING IN KERATINOCYTES	62	6	103	[34, 10]
007	CORTICAL AREA DEVELOPMENT	5	0	14	[6, 10]
008	DEATH RECEPTOR SIGNALING	25	3	45	[2, 10]
009	YEAST APOPTOSIS	60	13	114	[13, 10]
010	CARDIAC-DEVELOPMENT	13	2	37	[11, 10]
011	GUARD CELL ABSCISIC ACID SIGNALING	40	4	78	[17, 10]
012	T-CELL RECEPTOR SIGNALING	94	7	158	[31, 10]
013	CHOLESTEROL REGULATORY PATHWAY	32	2	41	[14, 10]
014	T-LGL SURVIVAL NETWORK 2008	54	7	193	[36, 10]
015	NEUROTRANSMITTER SIGNALING PATHWAY	14	2	20	[7, 10]
016	IL-1 SIGNALING	104	14	218	[29, 10]
017	DIFFERENTIATION OF T-LYMPHOCYTES	41	9	97	[20, 10]
018	EGFR-ERBB SIGNALING	76	28	226	[33, 10]
019	IL-6 SIGNALING	71	15	149	[29, 10]
020	APOPTOSIS NETWORK	39	2	73	[18, 10]
021	BODY SEGMENTATION IN DROSOPHILA 2013	14	3	29	[19, 10]
022	B-CELL DIFFERENTIATION	17	5	39	[22, 10]
023	MAMMALIAN CELL CYCLE 2006	9	1	34	[5, 10]
024	BUDDING YEAST CELL CYCLE	16	4	42	[35, 10]
025	T-LGL SURVIVAL NETWORK 2011	54	6	195	[30, 10]
026	BUDDING YEAST CELL CYCLE 2009	18	0	59	[12, 10]
027	WG PATHWAY OF DROSOPHILA	12	14	29	[21, 10]
028	VEGF PATHWAY OF DROSOPHILA	10	8	18	[21, 10]
029	TOLL PATHWAY OF DROSOPHILA	9	2	11	[21, 10]
030	SPZ NETWORK OF DROSOPHILA	18	6	28	[21, 10]

ID	Name	Vars.	Inps.	Regs.	Source
031	CELL CYCLE TRANSCRIPTION	9	0	19	[25, 10]
032	T-CELL SIGNALING 2006	37	3	53	[15, 10]
033	BT474 BREAST CELL LINE LONG TERM	19	5	68	[4, 10]
034	HCC1954 BREAST CELL LINE LONG TERM	19	4	68	[4, 10]
035	BT474 BREAST CELL LINE SHORT TERM	11	5	46	[4, 10]
036	HCC1954 BREAST CELL LINE SHORT TERM	11	5	46	[4, 10]
037	SKBR3 BREAST CELL LINE SHORT TERM	11	5	41	[4, 10]
038	SKBR3 BREAST CELL LINE LONG TERM	21	4	81	[4, 10]
039	HIV-1 INTERACTIONS WITH T-CELL SIGNALING	124	14	368	[26, 10]
040	T-CELL DIFFERENTIATION	19	4	34	[23, 10]

References

[1] Nikola Beneš, Luboš Brim, Jakub Kadlecaj, Samuel Pastva, and David Šafránek. AEON: attractor bifurcation analysis of parametrised boolean networks. In *International Conference on Computer Aided Verification*, pages 569–581. Springer, 2020.

[2] Laurence Calzone, Laurent Tournier, Simon Fourquet, Denis Thieffry, Boris Zhivotovsky, Emmanuel Barillot, and Andrei Zinovyev. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol*, 6(3):e1000702, 2010.

[3] Claudine Chaouiya, Duncan Bérenguier, Sarah M Keating, Aurélien Naldi, Martijn P Van Iersel, Nicolas Rodriguez, Andreas Dräger, Finja Büchel, Thomas Cokelaer, Bryan Kowal, et al. Sbml qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC systems biology*, 7(1):1–15, 2013.

[4] Silvia Von der Heyde, Christian Bender, Frauke Henjes, Johanna Sonntag, Ulrike Korf, and Tim Beissbarth. Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC systems biology*, 8(1):1–22, 2014.

[5] Adrien Fauré, Aurélien Naldi, Claudine Chaouiya, and Denis Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.

[6] Clare E Giacomantonio and Geoffrey J Goodhill. A boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Comput Biol*, 6(9):e1000936, 2010.

[7] Simone Gupta, Siddharth S Bisht, Ritushree Kukreti, Sanjeev Jain, and Samir K Brahmachari. Boolean network analysis of a neurotransmitter signaling pathway. *Journal of theoretical biology*, 244(3):463–469, 2007.

[8] Tomáš Helikar, Naomi Kochi, Bryan Kowal, Manjari Dimri, Mayumi Naramura, Srikumar M Raja, Vimla Band, Hamid Band, and Jim A Rogers. A comprehensive, multi-scale dynamical model of erbb receptor signal transduction in human mammary epithelial cells. *PloS one*, 8(4):e61757, 2013.

[9] Tomáš Helikar, John Konvalina, Jack Heidel, and Jim A Rogers. Emergent decision-making in biological signal transduction networks. *Proceedings of the National Academy of Sciences*, 105(6):1913–1918, 2008.

[10] Tomáš Helikar, Bryan Kowal, Sean McClenathan, Mitchell Bruckner, Thaine Rowley, Alex Madrahimov, Ben Wicks, Manish Shrestha, Kahani Limbu, and Jim A Rogers. The cell collective: toward an open and collaborative approach to systems biology. *BMC systems biology*, 6(1):1–14, 2012.

[11] Franziska Herrmann, Alexander Groß, Dao Zhou, Hans A Kestler, and Michael Kühl. A boolean model of the cardiac gene regulatory network determining first and second heart field identity. *PloS one*, 7(10):e46798, 2012.

[12] David J Irons. Logical analysis of the budding yeast cell cycle. *Journal of theoretical biology*, 257(4):543–559, 2009.

[13] Laleh Kazemzadeh, Marija Cvijovic, and Dina Petranovic. Boolean model of yeast apoptosis as a tool to study yeast and human apoptotic regulations. *Frontiers in physiology*, 3:446, 2012.

[14] Gwenael Kervizic and Laurent Corcos. Dynamical modeling of the cholesterol regulatory pathway with boolean networks. *BMC systems biology*, 2(1):1–14, 2008.

[15] Steffen Klamt, Julio Saez-Rodriguez, Jonathan A Lindquist, Luca Simeoni, and Ernst D Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC bioinformatics*, 7(1):1–26, 2006.

[16] Hannes Klarner, Adam Streck, and Heike Siebert. PyBoolNet: a python package for the generation, analysis and visualization of boolean networks. *Bioinformatics*, 33(5):770–772, 2017.

[17] Song Li, Sarah M Assmann, and Réka Albert. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol*, 4(10):e312, 2006.

[18] Zhongxing Mai and Haiyan Liu. Boolean network-based analysis of the apoptosis network: irreversible apoptosis and stable surviving. *Journal of theoretical biology*, 259(4):760–769, 2009.

[19] Manuel Marques-Pita and Luis M Rocha. Canalization and control in automata networks: body segmentation in drosophila melanogaster. *PloS one*, 8(3):e55946, 2013.

[20] Pablo Martínez-Sosa and Luis Mendoza. The regulatory network that controls the differentiation of T lymphocytes. *Biosystems*, 113(2):96–103, 2013.

[21] Abibatou Mbodj, Guillaume Junion, Christine Brun, Eileen EM Furlong, and Denis Thieffry. Logical modelling of drosophila signalling pathways. *Molecular BioSystems*, 9(9):2248–2258, 2013.

[22] Akram Méndez and Luis Mendoza. A network model to describe the terminal differentiation of b cells. *PLoS computational biology*, 12(1):e1004696, 2016.

[23] Luis Mendoza and Ioannis Xenarios. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theoretical Biology and Medical Modelling*, 3(1):1–18, 2006.

[24] Aurélien Naldi, Céline Hernandez, Wassim Abou-Jaoudé, Pedro T Monteiro, Claudine Chaouiya, and Denis Thieffry. Logical modeling and analysis of cellular regulatory networks with ginsim 3.0. *Frontiers in physiology*, 9:646, 2018.

[25] David A Orlando, Charles Y Lin, Allister Bernard, Jean Y Wang, Joshua ES Socolar, Edwin S Iversen, Alexander J Hartemink, and Steven B Haase. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, 453(7197):944–947, 2008.

[26] Oyebode J Oyeyemi, Oluwafemi Davies, David L Robertson, and Jean-Marc Schwartz. A logical model of hiv-1 interactions with the t-cell activation signalling pathway. *Bioinformatics*, 31(7):1075–1083, 2015.

[27] Sobia Raza, Kevin A Robertson, Paul A Lacaze, David Page, Anton J Enright, Peter Ghazal, and Tom C Freeman. A logic-based diagram of signalling pathways central to macrophage activation. *BMC systems biology*, 2(1):1–15, 2008.

[28] Alfredo Rodriguez, David Sosa, Leda Torres, Bertha Molina, Sara Frias, and Luis Mendoza. A boolean network model of the fa/brca pathway. *Bioinformatics*, 28(6):858–866, 2012.

[29] Anke Ryll, Regina Samaga, Fred Schaper, Leonidas G Alexopoulos, and Steffen Klamt. Large-scale network models of IL-1 and IL-6 signalling and their hepatocellular specification. *Molecular Biosystems*, 7(12):3253–3270, 2011.

[30] Assieh Saadatpour, Rui-Sheng Wang, Aijun Liao, Xin Liu, Thomas P Loughran, István Albert, and Réka Albert. Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. *PLoS computational biology*, 7(11):e1002267, 2011.

[31] Julio Saez-Rodriguez, Luca Simeoni, Jonathan A Lindquist, Rebecca Hemenway, Ursula Bommhardt, Boerge Arndt, Utz-Uwe Haus, Robert Weismantel, Ernst D Gilles, Steffen Klamt, et al. A logical model provides insights into T cell receptor signaling. *PLoS computational biology*, 3(8):e163, 2007.

[32] Özgür Sahin, Holger Fröhlich, Christian Löbke, Ulrike Korf, Sara Burmester, Meher Majety, Jens Mattern, Ingo Schupp, Claudine Chaouiya, Denis Thieffry, et al. Modeling erbb receptor-regulated g1/s transition to find novel targets for de novo trastuzumab resistance. *BMC systems biology*, 3(1):1–20, 2009.

[33] Regina Samaga, Julio Saez-Rodriguez, Leonidas G Alexopoulos, Peter K Sorger, and Steffen Klamt. The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS computational biology*, 5(8):e1000438, 2009.

[34] Amit Singh, Juliana M Nascimento, Silke Kowar, Hauke Busch, and Melanie Boerries. Boolean approach to signalling pathway modelling in hgf-induced keratinocyte migration. *Bioinformatics*, 28(18):i495–i501, 2012.

[35] Robert G Todd and Tomáš Helikar. Ergodic sets as cell phenotype of budding yeast cell cycle. 2012.

[36] Ranran Zhang, Mithun Vinod Shah, Jun Yang, Susan B Nyland, Xin Liu, Jong K Yun, Réka Albert, and Thomas P Loughran. Network model of survival signaling in large granular lymphocyte leukemia. *Proceedings of the National Academy of Sciences*, 105(42):16308–16313, 2008.