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 [7] or GINsim [12]. 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 [10], `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.	Regs.	Source
001	SIGNALING IN MACROPHAGE ACTIVATION	321	533	[13, 7]
002	SIGNAL TRANSDUCTION IN FIBROBLASTS			[6, 7]
003	MAMMALIAN CELL CYCLE			[15, 7]
004	ERBB RECEPTOR SIGNALING			[5, 7]
005	FA/BRCA PATHWAY			[14, 7]
006	HGF SIGNALING IN KERATINOCYTES			[16, 7]
007	CORTICAL AREA DEVELOPMENT			[4, 7]
008	DEATH RECEPTOR SIGNALING			[2, 7]
009	YEAST APOPTOSIS			[9, 7]
010	CARDIAC-DEVELOPMENT			[8, 7]
011	GUARD CELL ABSCISIC ACID SIGNALING			[11, 7]
012				
013				
014				
015				
016				
017				
018				
019				
020				
021				
022				

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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.