HAL open science

Automatic synthesis of boolean networks from biological knowledge and data

Athénaïs Vaginay, Taha Boukhobza, Malika Smaïl-Tabbone

To cite this version:

Athénaïs Vaginay, Taha Boukhobza, Malika Smaïl-Tabbone. Automatic synthesis of boolean networks from biological knowledge and data. Journées Ouvertes en Biologie, Informatique et Mathématiques, JOBIM 2021, Jul 2021, Paris, France. hal-03481253

HAL Id: hal-03481253

https://hal.science/hal-03481253v1

Submitted on 15 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatic Synthesis of Boolean Networks from Biological Knowledge and Data

Athénaïs Vaginay[1,2], Taha Boukhobza[2] and Malika Smaïl-Tabbone[1]
[1] LORIA (Université de Lorraine, CNRS, Inria)
[2] CRAN (Université de Lorraine, CNRS) Contact: athenais.vaginay@loria.fr

Corresponding author: athenais.vaginay@loria.fr

Reference paper: **Vaginay *et al.* (june 2021) Automatic Synthesis of Boolean Networks from Biological Knowledge and Data *International Conference in Optimization and Learning (OLA)*, long paper accepted for publication in the OLA 2021 Springer CCIS proceedings.**
https://hal.archives-ouvertes.fr/hal-03256693

Boolean Networks (BNs) are a simple formalism used to study complex biological systems when the prediction of exact reaction times is not of interest. They play a key role in understanding the dynamics of the studied systems and predicting their disruption in case of complex human diseases. A BN consists of a set of n Boolean *transition functions* (one per components) giving the successive Boolean states of the components, depending on the previous state of the other components of the system. Here is an example of a BN of three components called A, B and C: $\mathcal{B} = \{f_{\mathrm{A}} := \mathrm{C}\ ;\quad f_{\mathrm{B}} := \mathrm{B} \wedge \neg\mathrm{C}\ ;\quad f_{\mathrm{C}} := \neg\mathrm{C}\}$. It reads like "A will be activated is C was activated", "B will be activated if B was activated but C was not" and "C will be activated if C was not". The dynamics of a BN is obtained by applying iteratively the transition functions starting from all the 2^n possible configurations. The order of application of the transition functions is defined by the *update scheme*. In the *mixed* updated scheme, any number of components can be updated at each step. The dynamics is represented by a directed graph whose nodes are the 2^n configurations and the edges are the transitions according to the chosen update scheme. Such a graph is called *state transition graph* (STG).

BNs are generally built from experimental data and knowledge from the literature, either manually or with the aid of programs. The automatic synthesis of BNs is still a challenge for which several approaches have been proposed, such as `REVEAL` [1], `Best-Fit` [2] and `caspo-TS` [3]. In this paper, we propose `ASKeD-BN`, a new approach based on *Answer-Set Programming* (ASP) to synthesize BNs *constrained* in their dynamics by a multivariate Time Series (TS), and in their structure by a Prior Knowledge Network (PKN). A PKN is a directed graph on the components of the system. It constrains the structure of the synthesized BNs by defining which components can appear as variables in each transition function and the polarity of those variables. The synthesized BNs also have to reproduce as well as possible the sequence of configurations extracted from the given multivariate TS.

We compare `ASKeD-BN` with `REVEAL`, `Best-Fit` and `caspo-TS` according to three criteria: (i) the number of BNs returned by the approaches ran on a PKN and a multivariate TS; (ii) the median of the coverage ratios *i.e.,* the proportion of transitions extracted from the input TS that are present in the mixed STG of the BN; (iii) the standard deviation of the coverage ratios. We ran experiments on two real datasets and more than 300 synthetic datasets according to various settings (synchronous or asynchronous, with or without repetition, with or without noise), and provided empirical evidence that `ASKeD-BN` has the best trade-off on the evaluation criteria: it returns a small set of BNs which comply with the provided structural constraints, cover a good proportion of the dynamical constraints, with a small variance.

References

[1] Shoudan Liang, Stefanie Fuhrman, and Roland Somogyi. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. pages 18–29.

[2] Harri Lähdesmäki, Ilya Shmulevich, and Olli Yli-Harja. On learning gene regulatory networks under the boolean network model. 52(1):147–167.

[3] Max Ostrowski, Loïc Paulevé, Torsten Schaub, Anne Siegel, and Carito Guziolowski. Boolean network identification from perturbation time series data combining dynamics abstraction and logic programming. 149:139–153.