



HAL
open science

From quantitative SBML to boolean networks

Athénaïs Vaginay, Taha Boukhobza, Malika Smaïl-Tabbone

► **To cite this version:**

Athénaïs Vaginay, Taha Boukhobza, Malika Smaïl-Tabbone. From quantitative SBML to boolean networks. 19th conference on Computational Methods in Systems Biology, CMSB 2021, Sep 2021, Bordeaux, France. hal-03481267

HAL Id: hal-03481267

<https://hal.science/hal-03481267v1>

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.

From Quantitative SBML to Boolean Networks

Athénais Vaginay^(1, 2), Taha Boukhobza⁽¹⁾, Malika Smail-Tabbone⁽²⁾

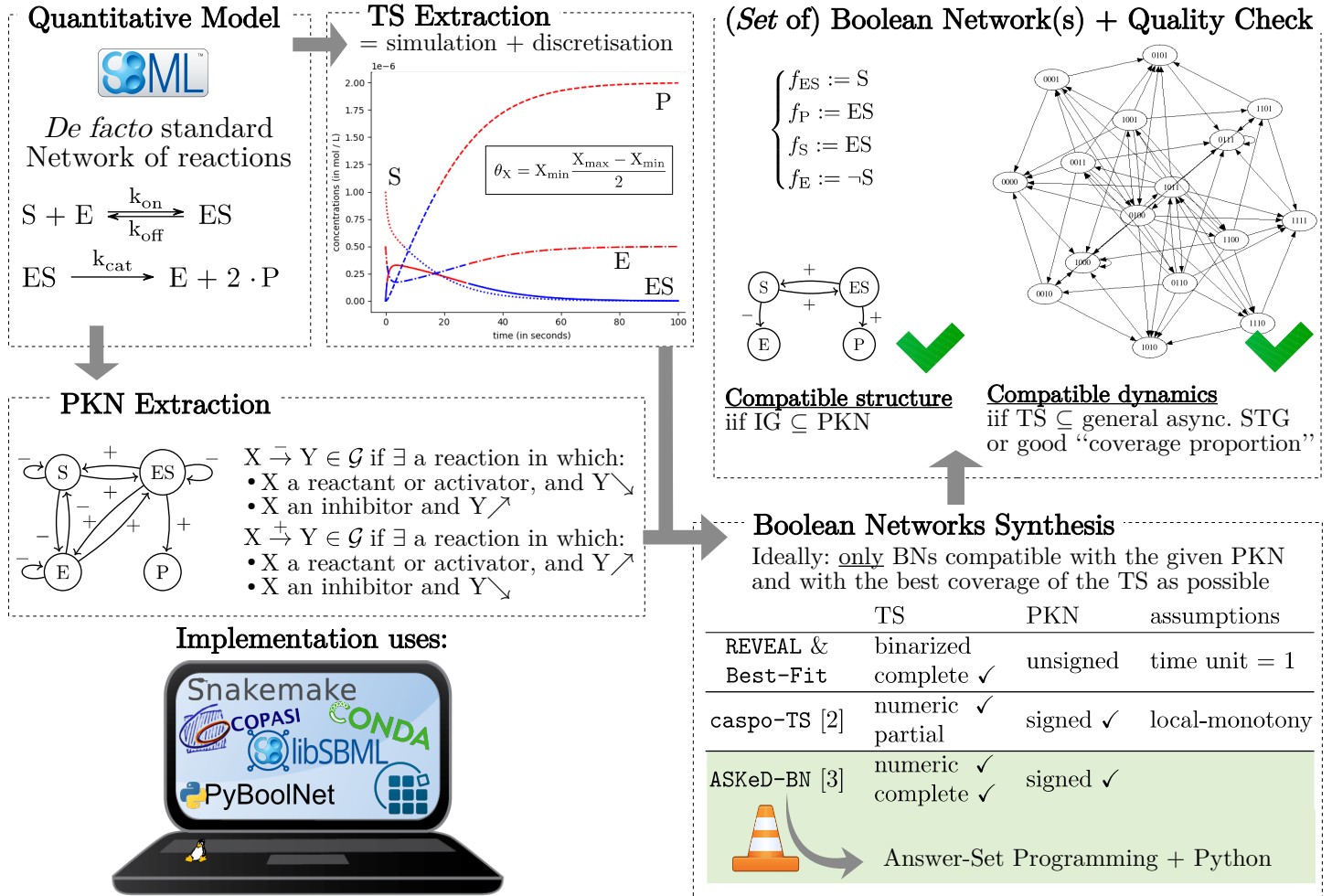
⁽¹⁾ CRAN (Université de Lorraine, CNRS) ⁽²⁾ LORIA (Université de Lorraine, CNRS, Inria)

<athenais.vaginay@loria.fr>



SBML is the *de facto* standard to encode and exchange models of biological systems in various formalisms. More than 300 curated SBML models of the repository Biomodels are quantitative. We aim at automatically converting these SBML models into a *set of compatible* Boolean networks. Here, we present our pipeline SBML2BN, which consists in extracting the model structure from the input SBML model and to retrieve the dynamics of the model by deterministic simulation. The structure and dynamics are then used to synthesise a set of compatible BNs. Finally, the obtained BNs are evaluated. We have tested SBML2BN on 200 models for which we show that it is overall efficient and successful.

SBML2BN Pipeline



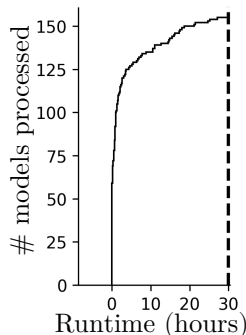
Models for the Evaluation

Curated branch of Biomodels release 31 (June 2017)

- 640 models in total
- 369 **complete** quantitative models = those with a PKN and a TS
- 209 **models evaluated** = models with all the components having less than 10 parent nodes. They have from 1 to 70 components



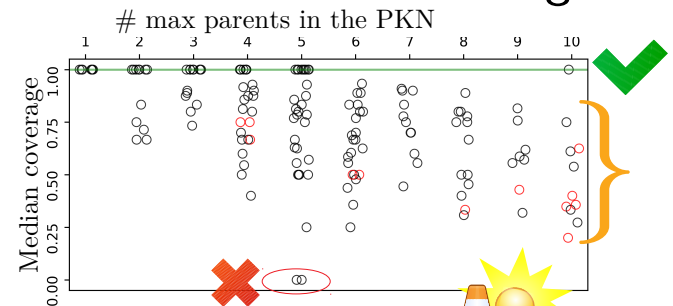
Runtime and # BN Synthesised



155 models processed in ≤ 30 hours
 ~ half processed in ≤ 30 mins
 106 models w/ a unique BN synthesised
 Median of coverage median = 0.77
 Coverage variance of 0, even when > 1 BN synthesised, except for 12 models (red circles in the figure on the right)

→ Overall, the pipeline is good

Evaluation of the Coverage



Models not "well-formed" [3]?

Example: BIOMD n°44 has reaction kinetics using components not listed in the reactants nor modifiers
 → PKN not complete (missing parents)
 → 1 BN generated with a coverage of 0.55

References:

- [1] Fages et al. 2012 hal-00723554
- [2] Ostrowski et al. 2016 doi:10.1016/j.biosystems.2016.07.009
- [3] Vaginay et al. 2021 doi:10.1007/978-3-030-85672-4_12