Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ••• (••••) •••–•••

FUZZY sets and systems

www.elsevier.com/locate/fss

Modeling multi-valued biological interaction networks using fuzzy answer set programming

Mushthofa Mushthofa [a,b,e], Steven Schockaert [c], Ling-Hong Hung [d], Kathleen Marchal [f], Martine De Cock [a,d]

[a] *Department of Applied Mathematics, Statistics and Informatics, Ghent University, Ghent, Belgium*
[b] *Department of Computer Science, Bogor Agricultural University, Bogor, Indonesia*
[c] *School of Computer Science & Informatics, Cardiff University, UK*
[d] *Institute of Technology, University of Washington Tacoma, USA*
[e] *Bioinformatics Institute Ghent, Ghent University, 9052 Ghent, Belgium*
[f] *Department of Information Technology (INTEC, iMINDS), Ghent University, 9052 Ghent, Belgium*

Received 2 February 2017; received in revised form 29 December 2017; accepted 1 January 2018

Abstract

Fuzzy Answer Set Programming (FASP) is an extension of the popular Answer Set Programming (ASP) paradigm that allows for modeling and solving combinatorial search problems in continuous domains. The recent development of practical solvers for FASP has enabled its applicability to real-world problems. In this paper, we investigate the application of FASP in modeling the dynamics of Gene Regulatory Networks (GRNs). A commonly used simplifying assumption to model the dynamics of GRNs is to assume only Boolean levels of activation of each node. Our work extends this Boolean network formalism by allowing multi-valued activation levels. We show how FASP can be used to model the dynamics of such networks. We experimentally assess the efficiency of our method using real biological networks found in the literature, as well as on randomly-generated synthetic networks. The experiments demonstrate the applicability and usefulness of our proposed method to find network attractors.

1. Introduction

In biological systems, genes are known to interact with each other in a complex and dynamic way. Briefly, each gene's activation state can influence the activation states of other genes, either positively or negatively. These interactions can be modeled using a graph structure, which is usually called a Gene Regulatory Network (GRN). It determines the patterns of activation states of the genes, which in turn affects the phenotypic behavior of the system.

E-mail addresses: Mushthofa.Mushthofa@UGent.be, mush@ipb.ac.id (M. Mushthofa), SchockaertS1@cardiff.ac.uk (S. Schockaert), lhhung@uw.edu (L.-H. Hung), Martine.DeCock@UGent.be, mdecock@uw.edu (M. De Cock).

https://doi.org/10.1016/j.fss.2018.01.003
0165-0114/

One of the most important concepts in modeling the dynamics of GRNs are the so-called *attractors*, which are the sets of states to which the system converges. An attractor usually corresponds to the observed characteristics/phenotypes of the biological system [1]. For example, the attractors of a GRN usually correspond to the expression patterns of the genes in the network for specific types of cells [2,3]. In studying the dynamics of such networks it is therefore of importance to be able to identify their attractors.

In systems biology, one of the most popular approaches to formalise a GRN is to use a so-called Boolean Network (BN) [4–6]. Boolean networks represent genes as nodes that can take on Boolean values (intuitively representing the activation levels of the genes), while interactions between the genes are represented as Boolean functions that determine the value of each node at a certain time, depending on the current values of the other genes. The state transitions of a GRN and their attractors can be readily represented using such a formalism.

There have been numerous works about computational tools to simulate the dynamics of Boolean networks and to compute their attractors, mostly using logic-based techniques such as Binary Decision Diagrams (BDDs) or Boolean SAT solvers [7–12]. More recently, Answer Set Programming (ASP) has become a particularly interesting framework for modeling GRNs and Boolean networks [13–16].

ASP is a popular declarative programming paradigm which allows for an easy and intuitive encoding of many combinatorial search and optimisation problems [17,18]. The availability of fast and efficient solvers for ASP, such as clasp [19] and DLV [20], allows for the application of ASP in various fields [21,22]. Despite its flexibility and expressive power, however, ASP lacks the ability to directly encode problems in continuous domains.

Having only two levels of activation is sometimes not always enough to fully understand the dynamics of real biological systems. For example, in [3,23–26], examples of systems are given whose dynamics can only be modeled by considering more than two activation levels. One classic example is the *lac operon* regulatory system, which is a set of genes that controls the production of the proteins needed to metabolise lactose in enteric bacterias, such as *Escherichia coli* (see e.g., [27]). In this case, it has been shown that one of the key attractors cannot be characterized using a Boolean encoding (because of the so-called "leaky-expression"). Despite the importance of multi-valued activation levels for modeling gene regulatory networks, only limited progress has been made on developing simulation tools that can support them. To the best of our knowledge, only one tool has been developed that supports multi-valued activation levels [24].

In this paper, we propose the use of Fuzzy Answer Set Programming (FASP) [28] as a computational framework to simulate the dynamics of multi-valued regulatory networks. FASP is a form of declarative programming that extends ASP by allowing graded truth values in atomic propositions and using fuzzy logic connectives to aggregate these truth values. Recent work on the implementation of a FASP solver, such as [29–33], has opened the door to the application of FASP for solving real-world applications. Other frameworks dealing with the extension of ASP, or more generally, logic programming into the fuzzy domains have been proposed in the literature, e.g., [34–39]. While we have specifically chosen to use the FASP framework and the corresponding solver from [32], other multi-valued extensions of ASP might also be suitable for the purpose of modeling the dynamics of multi-valued regulatory networks.

Here, we propose an encoding of the dynamics of multi-valued biological interaction networks that can be executed/solved using the FASP solver proposed in [32], and we prove the correctness of this encoding. We then perform an extensive benchmark test using synthetic networks as well as real biological networks found in the literature to show the efficiency and applicability of this method. The results indicate that the method is efficient for the size of the networks typically used in the Boolean/discrete modeling of regulatory networks (up to around a few dozen genes in the network).

This paper extends our previous work [40] with the following contributions: (1) we provide complete formal definitions of multi-valued networks and their dynamics, (2) we provide detailed proofs of the correctness of the encoding, (3) we extend the framework to address the problems with the encoding of cyclic attractors, in particular, in the case of asynchronous updates, (4) we describe a method to perform automatic encoding of the network structure into fuzzy propositions and the implementation of a tool to perform this (FASPG), and (5) we extend the experiments to include synthetic networks and show the performance of our methods for increasingly large networks, including the computation of cyclic attractors of synthetic networks under different schemes of updates.

The remainder of this paper is structured as follows: we first describe related work in Section 2, and present the preliminaries on Boolean networks and the theoretical background on (F)ASP in Section 3. We then formally define the multi-valued networks and present our FASP-based encoding in Section 4. Section 5 describes the FASPG tool that implements the proposed method, as well as providing an automatic encoding for the network. Section 6 contains

Please cite this article in press as: M. Mushthofa et al., Modeling multi-valued biological interaction networks using fuzzy answer set programming, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.01.003