Physica A 392 (2013) 5525–5535

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Theoretical estimation of metabolic network robustness against multiple reaction knockouts using branching process approximation

Kazuhiro Takemoto [a,*], Takeyuki Tamura [b], Tatsuya Akutsu [b]

[a] Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Kawazu 680-4, Iizuka, Fukuoka 820-8502, Japan
[b] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan

HIGHLIGHTS

- We propose a model based on branching processes for estimating metabolic robustness.
- The model is extended to consider the knockout of multiple reactions.
- The validity of the modeling was demonstrated with real data.

ARTICLE INFO

Article history:
Received 9 May 2013
Received in revised form 28 June 2013
Available online 6 July 2013

Keywords:
Metabolic network
Branching process
Cascading failure

ABSTRACT

In our previous study, we showed that the branching process approximation is useful for estimating metabolic robustness, measured using the *impact degree*. By applying a theory of random family forests, we here extend the branching process approximation to consider the knockout of *multiple* reactions, inspired by the importance of multiple knockouts reported by recent computational and experimental studies. In addition, we propose a better definition of the number of offspring of each reaction node, allowing for an improved estimation of the impact degree distribution obtained as a result of a single knockout. Importantly, our proposed approach is also applicable to multiple knockouts. The comparisons between theoretical predictions and numerical results using real-world metabolic networks demonstrate the validity of the modeling based on random family forests for estimating the impact degree distributions resulting from the knockout of multiple reactions.

1. Introduction

Robustness is an important concept for understanding how living organisms adapt to changing environments, as well as for understanding how they are able to survive when carrying mutated genes [1,2]. In particular, metabolic robustness is of increasing interest not only to researchers in the field of basic biology, but also to those in biotechnology and medical research because metabolic processes are essential for physiological functions, and are responsible for maintaining life.

The development of high-throughput methods has facilitated the collection and compilation of large metabolic network datasets, which are stored in databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [3] and the Encyclopedia of Metabolic Pathways (MetaCyc) [4]. In recent years, numerous methods and measures have been developed for analyzing metabolic robustness in the context of gene/reaction knockout by using available metabolic network data.

* Corresponding author. Tel.: +81 948 29 7822.
E-mail address: takemoto@bio.kyutech.ac.jp (K. Takemoto).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2013.07.003

Many of these methods and measures are based on *flux balance analysis* (FBA). Edwards and Palsson used the change in the optimal objective function value (e.g., growth rate) as a measure of robustness against the change in a particular reaction [5]. Segrè et al. [6] proposed the minimization of metabolic adjustment (MOMA) method, which predicts the flux vectors by minimizing the Euclidean distance between the mutant and wild type. Deutscher et al. [7] proposed another measure by combining FBA with the Shapley value in game theory. With respect to FBA, *elementary flux modes* (EFMs) have also been used to analyze the robustness of metabolic networks, where an EFM is a minimal set of reactions that can operate at the steady state [8]. Wilhelm et al. [9] proposed a measure based on the numbers of EFMs before and after knockout, which was later extended to include the knockout of multiple reactions [10]. In order to evaluate robustness for the production of a specific target compound(s), several studies have used a minimum reaction cut based on FBA and/or EFM [11–14], which involves a minimum set of reactions (or enzymes), the removal of which leads to prevention of the production of a specific set of compounds. Other approaches based on FBA/EFM have also been implemented [15].

Boolean modeling is an alternative way to model metabolic networks, whereby the activity of each reaction or compound is represented by either 0 (inactive) or 1 (active) and reactions and compounds are modeled as AND nodes and OR nodes respectively. Handorf et al. [16] introduced the concept of *scope* based on Boolean modeling and applied it to analyses of the robustness of metabolic networks. Li et al. [17], Sridhar et al. [18], and Tamura et al. [19] developed integer programming-based methods for determining the minimum reaction cut under Boolean models. Lemke et al. [20] defined the *damage* as the number of reactions deactivated by the knockout of a single reaction under a Boolean model. Smart et al. [21] refined the concept of damage by introducing the *topological flux balance* (TFB) criterion. Jiang et al. defined the *impact degree* as the number of reactions deactivated by knockout of a specified reaction [22] under a Boolean model. Although there are some differences in the treatment of reversible reactions, the damage and the impact degree are very similar concepts.

To date, most studies have focused on the prediction and/or accuracy of robustness measures but have given less consideration to the distribution of such measures. Lemke et al. [20] analyzed the distribution of damage using computer simulation. Smart et al. [21] performed a similar analysis. In addition, they applied percolation theory and branching processes to the analysis of the distribution of cluster sizes of damaged subnetworks [21]; however, they did not explicitly estimate damage distribution (i.e., impact degree distribution).

Until recently, theoretical frameworks for estimating the tolerance of metabolic networks to various failures were poorly established. Motivated by this, in our previous study [23], we analyzed the distribution of impact degree triggered by random knockout of a single reaction using a branching process theory [24,25]. By treating the propagation of the impact triggered by the knockout of a reaction as a branching process approximation, the relevance of which had been shown in the context of loading-dependent cascading failure [26–28], we demonstrated that the branching process model (or theory) reflects the observed impact degree distributions. In addition, Lee et al. [29] also recently demonstrated the use of a Boolean model and a theory of branching process in this context.

As above, most previous studies focused on the impact of a single knockout. In recent years, however, multiple-knockout experiments have been actively performed, and have shown new interesting results on metabolic robustness, such as synergetic effects resulting from multiple knockouts [30–32]. Therefore, computational and theoretical frameworks need to be extended to include multiple knockouts. For example, Deutscher et al. [7] discussed the impact of multiple knockouts in yeast metabolism based on the Shapley value from game theory. Tamura et al. [33] proposed an efficient method for computing metabolic robustness in the context of impact degree. However, theoretical approaches remain incomplete.

In this study, by extending the branching process approximation proposed in our previous study [23], we show that the branching process approximation (specifically, the assumption of a random family forest, which is a collection of family trees) is also useful for estimating the distribution of the impact degree triggered by the random knockout (or disruption) of *multiple* reactions.

2. Impact degree

Here, we briefly review the impact degree [22] and its extensions [19,33]. The *impact degree* was originally proposed by Jiang et al. as a measure of the importance of each reaction in a metabolic network [22], and is defined as the number of deactivated reactions caused by the knockout of a single reaction. Since the effect of cycles was not considered in their study, Tamura et al. extended the impact degree so that the effect of cycles is taken into account by introducing the maximal valid assignment concept [19]. Furthermore, they extended it to cope with the knockout of multiple reactions [33].

Let $V_c = \{C_1, \ldots, C_m\}$ and $V_r = \{R_1, \ldots, R_n\}$ be a set of *compound nodes* and a set of *reaction nodes* respectively, where $V_c \cap V_r = \{\}$. A *metabolic network* is defined as a bipartite directed graph $G(V_c \cup V_r, E)$ in which each edge is directed either from a node in V_c to a node in V_r, or from a node in V_r to a node in V_c. Each of the included reactions and compounds takes 1 of 2 states: 0 (inactive) or 1 (active).

The impact degree for the knockout of multiple reactions is computed as follows [33]. Let $V_{ko} = \{R_{i_1}, \ldots, R_{i_D}\}$ be a set of reactions that have been knocked out. We start with the global state, such that all compounds are active (i.e., $C_i = 1$ for all $C_i \in V_c$) and all reactions except for those in V_{ko} are active (i.e., $R_i = 1$ for all $R_i \in V_r - V_{ko}$ and $R_i = 0$ for all $R_i \in V_{ko}$). Then, we update the states of reactions and compounds using the following rules.

(1) A reaction is deactivated if any predecessor (i.e., substrate) or successor (i.e., product) is inactive.
(2) A compound is deactivated if all predecessors or all successors are inactive.