Physica A 391 (2012) 379–387

Contents lists available at SciVerse ScienceDirect

Physica A

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

# Analysis of the impact degree distribution in metabolic networks using branching process approximation

Kazuhiro Takemoto [a,b], Takeyuki Tamura [c], Yang Cong [d], Wai-Ki Ching [d], Jean-Philippe Vert [e,f,g], Tatsuya Akutsu [c,*]

[a] *PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan*
[b] *Department of Biophysics and Biochemistry, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan*
[c] *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan*
[d] *Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong*
[e] *Centre for Computational Biology, Mines ParisTech, 35 rue Saint-Honoré, 77305 Fontainebleau cedex, France*
[f] *Institut Curie, 75248 Paris, France*
[g] *INSERM, U900, 75248 Paris, France*

ARTICLE INFO

*Article history:*
Received 1 June 2011
Received in revised form 2 August 2011
Available online 25 August 2011

*Keywords:*
Metabolic network
Branching process
Power law
Cascading failure

ABSTRACT

Theoretical frameworks to estimate the tolerance of metabolic networks to various failures are important to evaluate the robustness of biological complex systems in systems biology. In this paper, we focus on a measure for robustness in metabolic networks, namely, the *impact degree*, and propose an approximation method to predict the probability distribution of impact degrees from metabolic network structures using the theory of branching process. We demonstrate the relevance of this method by testing it on real-world metabolic networks. Although the approximation method possesses a few limitations, it may be a powerful tool for evaluating metabolic robustness.



## 1. Introduction

Robustness is a key feature in the analysis of complex systems, especially for complex biological systems. Many organisms have strong adaptability to environmental changes or failures in some of their components, and can live even if some of their genes are mutated. In particular, it is known that cancer cells are very robust [1]. Therefore, understanding the origin of robustness of living cells has become an important research topic.

In particular, extensive studies have focused on the analysis of structural robustness of metabolic networks. Structural robustness refers to the tolerance of the system's behavior to changes in the structure of networks, and most existing studies focus on changes caused by knockout of gene(s) or enzyme(s). One of the reasons why extensive studies have been done on structural robustness of metabolic networks is that rather accurate data of metabolic networks are available via databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [2] and the Encyclopedia of *Escherichia coli* K-12 Genes and Metabolism (EcoCyc) [3], and kinetic parameters, which are not necessarily available, are not required.

In order to analyze the structural robustness of metabolic networks, the *flux balance analysis* (FBA) methods have been widely used. In many of these approaches, *elementary flux modes* (EFMs) play a key role, where an EFM is a minimal set of reactions that can operate at a steady state [4]. Based on FBA and/or EFM, several studies have focused on finding a minimum reaction cut [5–8], that is, a minimum set of reaction (or enzyme) removals which prevent the production of a specified set

* Corresponding author.
*E-mail address:* takutsu@kuicr.kyoto-u.ac.jp (T. Akutsu).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.08.011

of compounds. Other FBA-based measures of robustness have also been proposed. Behre et al. proposed a measure based on the number of remaining EFMs after knockout versus the number of EFMs in the unperturbed situation [9]. Deutscher et al. proposed another measure using the Shapley value from the game theory [10].

Other approaches have been proposed based on Boolean models of metabolic networks in which reactions and compounds are modeled as AND and OR nodes, respectively. Handorf et al. analyzed robustness of metabolic networks by introducing the concept of *scope* [11]. Li et al., Sridhar et al., and Tamura et al. developed integer programming-based methods for finding a minimum reaction cut in Boolean models of metabolic networks [12–15]. Smart et al. defined the *topological flux balance* (TFB) criterion based on a Boolean model of metabolic networks and analyzed the *damage* (number of reactions) caused by knockout of a single reaction under TFB [16]. Jiang et al. defined and analyzed the *impact degree*, which is the number of reactions inactivated by knockout of a specific reaction [17]. Although there are some differences in the treatment of reversible reactions, the damage and the impact degree are very similar concepts. Cong et al. extended the impact degree for knockout of multiple reactions [18].

In this paper, we study the distribution of the impact degree caused by random knockout of a single reaction using the theory of branching process [19,20]. In order to analyze earthquakes, Saichev et al. proposed a branching process with power-law distributions of offspring $d$ : $P(d) \propto 1/d^{\gamma+1}$, where $\gamma$ is some constant, and approximately derived the distribution of the total number of offsprings [20]. We regard propagation of the impact of knockout of a reaction as a branching process, and apply their method to estimate the impact degree distribution, where the impact degree in our problem corresponds to the total number of offsprings in the branching process. In order to apply this method, we develop a simple method for estimating the offspring distribution in a metabolic network. Although Smart et al. have already applied percolation theory and branching process to the analysis of the size distribution of rigid clusters, defined as clusters of contagion nodes that do not contain any branched metabolite nodes (see Fig. 3A in Ref. [16]), they did not explicitly estimate the damage distribution (i.e., the impact degree distribution). We finally show an estimation method for the damage distributions. The proposed method is applied to the analysis of metabolic networks of four species: *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae* and *Homo sapiens*. The results show good agreement of impact degree distributions between empirical results and theoretical estimates.

## 2. Impact degree

Jiang et al. proposed the *impact degree* as a measure of the importance of each reaction in a metabolic network [17]. The impact degree is defined as the number of inactivated reactions caused by knockout of a single reaction. However, they did not consider the effect of cycles in metabolic networks. Since cycles play an important role in metabolic networks, Cong et al. extended the impact degree so that the effect of cycles is taken into account [18] by using a concept of the maximal valid assignment [15]. Here, we briefly review the definition of this extended impact degree [18].

As in other works, we regard each metabolic network as a bipartite directed graph. 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 directed graph $G(V_c \cup V_r, E)$ in which either $(u \in V_c) \wedge (v \in V_r)$ or $(u \in V_r) \wedge (v \in V_c)$ holds for each edge $(u, v) \in E$.[1] Each reaction and compound take one of the two states: 0 or 1, where 0 and 1 correspond to inactive and active reactions (compounds), respectively. Reverse reactions are treated as two irreversible reactions.

In order to define the impact degree of a reaction, we proceed as follows. Suppose that reaction $R_i$ is knocked out. Then, we start with the global state where all compounds are active (i.e., $C_k = 1$ for all $C_k \in V_c$) and all reactions but $R_i$ are active (i.e., $R_j = 1$ for all $R_j \in V_r \setminus \{R_i\}$ and $R_i = 0$). Then, we alternatively update the states of reactions and compounds by the following rules.

(1) For each reaction, there are three different compounds: consumed compounds (i.e., substrates), produced compounds (i.e., products), and directly unrelated compounds.
(2) A reaction is inactivated if any of its consumed or produced compound is inactivated.
(3) For each compound, there are three different reactions: consuming reactions, producing reactions, and directly unrelated reaction.
(4) A compound is inactivated if all of its consuming reactions or all of its producing reactions are inactivated.

Since no activation is possible in this process, the procedure converges to a stable state in a finite number of iterations. The impact degree of reaction $R_i$ is defined as the number of inactivated reactions in the stable state. This procedure simultaneously gives the definition of the impact degree and an algorithm to compute it.

Let us illustrate the above process with the metabolic network shown in Fig. 1. Suppose that reaction $R_2$ is knocked out. Then, the states of nodes change as shown in Table 1. Since four reactions (including $R_2$) are inactivated in the stable state, the resulting impact degree is four. Next, suppose that reaction $R_6$ is knocked out. In this case, the states of nodes change as shown in Table 2 and the resulting impact degree is two. It is to be noted that $C_4$ is not inactivated because $R_2$ is still active in this case.

[1] $A \wedge B$ means logical AND of $A$ and $B$.