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

A B S T R A C T

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.

© 2011 Elsevier B.V. All rights reserved.

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).

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 γ 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.

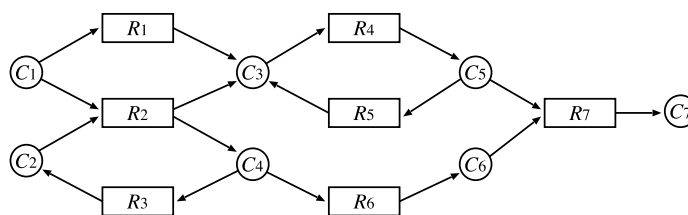[1] $A \wedge B$ means logical AND of A and B.

Fig. 1. Example of a metabolic network. Boxes and circles correspond to reactions and compounds, respectively.

Table 1
Impact degree calculation when R_2 in Fig. 1 is knocked out.

R_1	R_2	R_3	R_4	R_5	R_6	R_7	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	0	1	1	1	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	0	1	0	1	1	1
1	0	0	1	1	0	1	1	0	1	0	1	1	1
1	0	0	1	1	0	1	1	0	1	0	1	0	1
1	0	0	1	1	0	0	1	0	1	0	1	0	1
1	0	0	1	1	0	0	1	0	1	0	1	0	0

Table 2
Impact degree calculation when R_6 in Fig. 1 is knocked out.

R_1	R_2	R_3	R_4	R_5	R_6	R_7	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	1	1	1	1	0	1	1	1	1	1	1	1	1
1	1	1	1	1	0	1	1	1	1	1	1	0	1
1	1	1	1	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	0	1	1	1	1	1	0	0

3. Branching process approximation

We here explain the branching process approximation for estimating the impact degree distributions of metabolic networks.

The branching process is a stochastic process in which each progenitor generates offsprings according to a fixed probability distribution called the offspring distribution. We propose that the branching process approximation is useful for estimating the impact degree distribution because the propagation of an impact on a network is essentially similar to cascading failures, which is a sequence of failures caused by an accident. The branching process model of cascading failure is a standard Galton–Watson branching process [21]. The approximation method using branching process has been already applied to loading-dependent cascading failure, and its relevance has been shown [22,23]. However, the branching process approximation needs the assumption of tree structure of networks; thus, this is a limitation of the branching process approximation because metabolic networks generally have cycles.

3.1. The number of offsprings in metabolic networks

To estimate the impact degree distributions using the branching process approximation, we need to define the notion of offsprings for each reaction node in metabolic networks, and the distribution of the number of offsprings.

For that purpose, we consider the reaction network obtained as the unipartite projection of the metabolic network, where we draw an edge from reaction A to reaction B when at least one product of A is a substrate of B [24–27]. Fig. 2 shows the reaction network obtained from the metabolic network of Fig. 1 by this procedure.

As an easy example, we consider 2 metabolic reactions, A and B. In this case, the edge is drawn from A to B (i.e., A → B) if at least 1 product of reaction A corresponds to at least 1 substrate of reaction B (e.g., the case of $a \rightarrow$ A $\rightarrow b \rightarrow$ B $\rightarrow c$, where a, b, and c are chemical compounds). Through a similar procedure, we obtain a reaction network from a given metabolic network (see Fig. 2).

We now observe that the number of reactions inactivated by the failure of a given reaction, which we want to model in the branching process, does not correspond to its outdegree (i.e., the number of out-going edges from a reaction node) in the reaction network because the spreading of an impact is depressed when there are alternative synthetic pathways. For example, assuming that the reaction R_1 is inactive (or disrupted), the cascading of the impact does not occur because the reaction R_4 remains active due to the chemical compound generated through the reaction R_2. In contrast, the cascading of the impact continues when the reaction R_2 becomes inactive for instance because the reaction R_6 is dependent on this reaction only. In short, the impact spreads though reactions whose substrates are synthesized via unique metabolic reactions (i.e., reaction nodes with the indegree of 1) when assuming tree structures of networks.
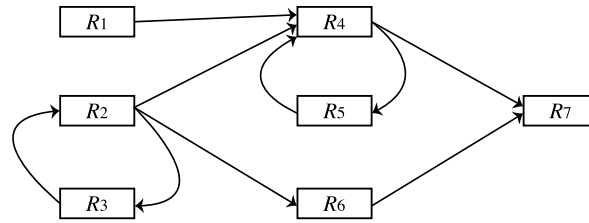
Fig. 2. The reaction network transformed from the metabolic network in Fig. 1.

Based on this analysis, we define the number of offsprings for each reaction node in metabolic networks as follows:

$$d_i = \begin{cases} k_i^{\text{out}} & (\text{if } k_i^{\text{in}} = 1) \\ 0 & (\text{otherwise}), \end{cases} \tag{1}$$

where k_i^{out} and k_i^{in} are the outdegree and indegree of reaction node i in the reaction network, respectively.

3.2. Branching process models

To analytically estimate the impact degree distributions, it is useful to assume that the number of offsprings for each node follows a probability distribution. We here consider two types of distributions, which are frequently observed in real worlds.

3.2.1. Poisson model

The simplest case of branching processes is a Poisson branching process (hereafter called the *Poisson model*) in which the number of offsprings d for each progenitor follows the Poisson distribution: $\mu^d e^{-\mu}/d!$, where μ corresponds to the mean of this distribution. In this model, the total number of offsprings (i.e., impact degree) r is distributed according to the Borel distribution [28]:

$$P(r) = (\mu r)^{r-1} \frac{e^{-\mu r}}{r!}. \tag{2}$$

Using Stirling's formula (i.e., $r! \approx \sqrt{2\pi r} r^r e^{-r}$), the above equation leads to the approximation

$$P(r) \propto r^{-3/2} e^{-r(\ln \mu - \mu + 1)}. \tag{3}$$

In particular, when $\mu = 1$ (i.e., the critical case), the impact degree follows a power-law distribution with exponent $-3/2$.

3.2.2. Power-law model

In addition to the Poisson case, we consider the case where the number of offsprings is determined based on a power-law distribution (hereafter called the *power-law model*). Indeed, the number of offsprings for each reaction node is based on the outdegree in metabolic networks. Since real-world complex networks including metabolic networks have power-law degree distributions [24,29,30], the number of offsprings for each reaction node may also obey a power-law distribution.

Saichev et al. [20,31] showed analytical asymptotic approximations for the distribution $P(r)$ of the total number of offsprings (i.e., impact degree) r in the case where the number of offsprings d for each progenitor follows asymptotically a power-law distribution $P(d) \propto 1/d^{\gamma+1}$ and the mean number of offsprings has a given value μ. In particular, they derive the following approximation for large impact degree r, in the case $1 < \gamma < 2$ where the variance of the number of offsprings is infinite:

$$P(r) \simeq \frac{\mu}{\nu r^{1+1/\gamma}} \varphi_\gamma \left(\frac{(1-\mu)r - \mu - 1}{\nu r^{1/\gamma}} \right), \tag{4}$$

where

$$\varphi_\gamma(x) = \int_0^\infty \exp\left[u^\gamma \cos\left(\frac{\pi \gamma}{2} \right) \right] \cos\left[u^\gamma \sin\left(\frac{\pi \gamma}{2} \right) + ux \right] du, \tag{5}$$

and $\nu = \mu(\gamma - 1)\gamma^{1/\gamma-1}\Gamma(-\gamma)^{1/\gamma}$, where $\Gamma(x)$ is the Gamma function.

When $\mu = 1$ (i.e., the critical case), in particular, the distribution of the total number of offsprings obeys the power-law distribution: $P(r) \propto 1/r^{1+1/\gamma}$.

In addition, $P(r)$ in the case of $\gamma > 2$ is approximately similar to that in the case of the Poisson model (i.e., Eq. (2)) [20].

Table 3
Model parameters extracted from real metabolic networks. The character # indicates "the number of". The number of reaction nodes corresponds to the number of parents in branching processes.

Species	#Reaction nodes	Mean μ	Exponent γ
Escherichia coli	1467	0.68	1.52
Bacillus subtilis	1279	0.60	1.65
Saccharomyces cerevisiae	1172	0.54	1.73
Homo sapiens	1982	0.58	1.50

3.2.3. Empirical model

The above probability distributions may be unsuitable to approximate real-world offspring distributions. In addition, we also consider a branching process model using empirical offspring distributions (hereafter called the *Empirical model*). This way, we can estimate the distribution $P(r)$ of the total number of offsprings (i.e., impact degree distributions) without the approximation of offspring distributions, although the model is analytically intractable. Moreover, we can evaluate whether the prediction accuracy of the proposed method is influenced by the approximation of offspring distributions or the fidelity of branching processes.

Let $F(s)$ be the probability generating function of the impact degree r (i.e., the total number of offsprings), the function $F(s)$ satisfies the recursive relation [20,21]:

$$F(s) = f(sF(s)), \tag{6}$$

where $f(s)$ denotes the probability generating function of the number d of offsprings of each node.

Using the Lagrange expansion and the relation of $P(r) = (1/r!)\mathrm{d}^r F(s)/\mathrm{d}s^r|_{s=0}$, the distribution $P(r)$ (i.e., impact degree distribution) is derived from the above implicit equation as the following explicit equation [20,21]:

$$P(r) = \frac{1}{r!}\frac{\mathrm{d}^{r-1}}{\mathrm{d}s^{r-1}}\left[f^r(s)\frac{\mathrm{d}f(s)}{\mathrm{d}s}\right]\Bigg|_{s=0} \quad (r > 0), \tag{7}$$

where $f(s) = \sum_{d=0}^{d_{max}} P(d)s^d$. The value d_{max} indicates the maximum of d, and the function $P(d)$ corresponds to the probability density function of empirical d (i.e., empirical offspring distribution). In addition, $P(r) = f(0)$ when $r = 0$.

3.2.4. Parameter extraction

To apply the Poisson and the power-law models, we need to estimate the model parameters, namely the mean μ and the exponent γ, from real metabolic networks.

We estimate the mean of the number of offsprings for each reaction node by the empirical average:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i, \tag{8}$$

where N is the total number of reaction nodes in a metabolic network.

We estimate the exponent of a power-law offspring distribution using the maximum likelihood estimation method [32]:

$$\gamma = |N^*|\left[\sum_{i \in N^*} \ln \frac{d_i}{d_{min}}\right]^{-1}, \tag{9}$$

where N^* is the set of reaction nodes with $d_i > 0$, and $|N^*|$ indicates the total number of such reaction nodes. d_{min} is the minimum of d_i in the set of N^*.

4. Evaluation of the branching process approximation

We evaluated the above estimation methods for the impact degree distributions on several real metabolic networks.

We selected two bacteria (*Escherichia coli* (eco) and *Bacillus subtilis* (bsu)) and two eukaryotes (*Saccharomyces cerevisiae* (sce) and *Homo sapiens* (hsa)) whose metabolic pathways have been well-identified. We downloaded the data of their metabolic networks, represented as bipartite networks as shown in Fig. 1, from the KEGG database (version 0.7.1) [2,33]. The parenthetic three-letter codes correspond to KEGG organism identifiers [34].

Based on the KEGG metabolic network data, the impact degree distributions in the metabolic networks were calculated using the method explained in Section 2. Moreover, we constructed the reaction networks of these species, and obtained the offspring distributions. Using Eqs. (8) and (9), the model parameters μ and γ were extracted (see Table 3). All metabolic networks show $1 < \gamma < 2$, implying that the assumption of the power-law model is suitable.

To test the validity of the offspring distribution models, we compared the fitting results on empirical offspring distributions in real metabolic networks between the power-law distribution and the Poisson distribution. Fig. 3 shows
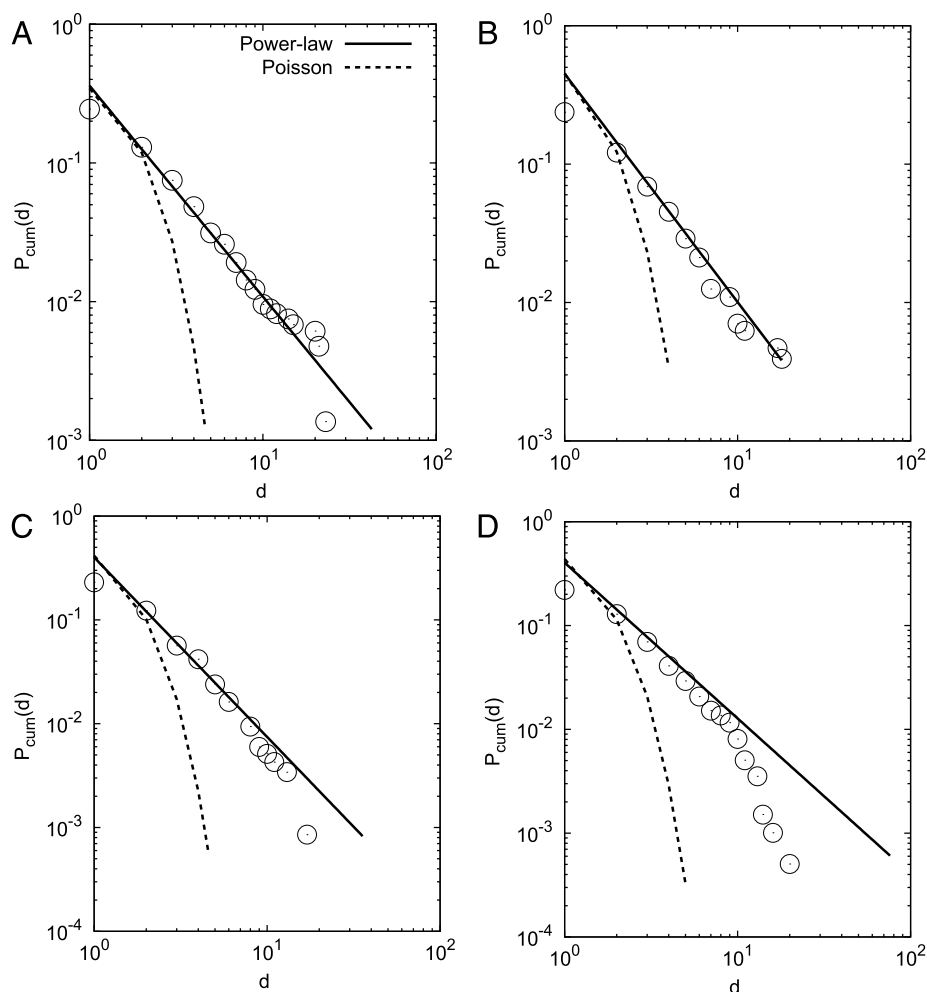
Fig. 3. Cumulative offspring distributions of *Escherichia coli* (A), *Bacillus subtilis* (B), *Saccharomyces cerevisiae* (C), and *Homo sapiens* (D). $P_{cum}(x)$ is defined as $P(X \geq x)$. Note that $P_{cum}(1) < 1$ because the cumulative distributions also consider the case of $d = 0$. $P_{cum}(0)$ is not shown due to the logarithmic display. The symbols indicate observed data. The black solid lines and dashed lines correspond to the cumulative representations of the power-law distribution with the exponent estimated by Eq. (9) and the Poisson distribution with the mean obtained from Eq. (8), respectively.

the cumulative offspring distributions from the real metabolic networks and the cumulative representation of theoretical distributions. The figure clearly indicates that the power-law distributions are more appropriate for modeling offspring distributions than the Poisson distributions. However, the power-law distributions may not be the best model because of the poor fittings for the larger *d* in the case of *H. sapiens* (Fig. 3(D)).

Using Eqs. (2), (4) and (7), we obtained the estimated impact degree distributions using the branching process approximation. Fig. 4 shows the comparison between the observed cumulative impact degree distributions and estimated ones. The theoretical predictions (lines) are in good agreement with the real impact degree distributions (symbols), suggesting the relevance of branching process approximations. Note that the impact degree distributions does not follow a clear power law and show an exponential cut-off for larger impact degrees because $\mu < 1$ (i.e., not the critical case). Eq. (3) or (4) can explain this distributional tendency.

To evaluate the prediction accuracy for impact degree distributions between the power-law model and the Poisson model, we evaluated the distributional distance (i.e., Kolmogorov–Smirnov (KS) statistics) between the observed distributions and theoretical distributions (Table 4). The power-law model is better than the Poisson model on all networks. The empirical model outperforms the power law model on both bacterial networks. Surprisingly, the power law model outperforms the empirical model on both eukaryote's networks.

5. Discussion and conclusion

We proposed a model to estimate the impact degree distributions in metabolic networks, using a branching process approximation, and demonstrated its validity on real data.

The power-law model could more accurately estimate the impact degree distributions in real metabolic networks than the Poisson model because the number of offsprings for each reaction node is assumed to follow the power-law distribution. Especially, the power-law model showed the significant agreements between the predicted distribution and the observed
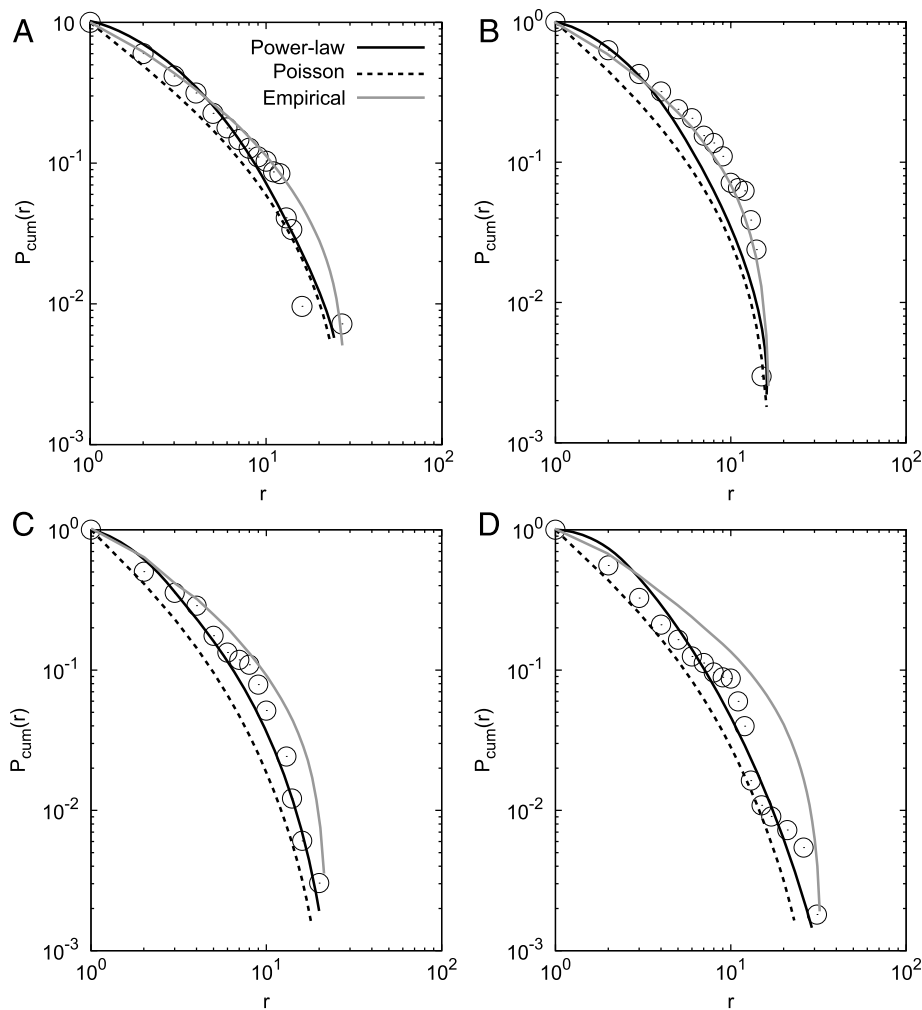
Fig. 4. Cumulative impact degree r distributions of *Escherichia coli* (A), *Bacillus subtilis* (B), *Saccharomyces cerevisiae* (C), and *Homo sapiens* (D). The symbols indicate observed data. The black solid lines and dashed lines correspond to the cumulative representations of theoretical distributions of the power-law model and the Poisson model, respectively. The gray solid lines are the cumulative representations of theoretical distributions from the empirical model. Note that $P_{cum}(x)$ is defined as $P(X \geq x)$.

Table 4
Prediction accuracy for the impact degree distributions: Kolmogorov–Smirnov (KS) distance, defined as $\sup_x |R(x) - M(x)|$, where $R(x)$ and $M(x)$ are empirical distributions and theoretical distributions, respectively. The parenthetic values indicate the logarithmic P-values p from the KS test, defined as $-\log_{10}(p)$. The emphasized values correspond to the best accuracy.

Species	Poisson model	Power-law model	Empirical model
Escherichia coli	0.11 (2.06)	0.08 (0.86)	**0.05** (**0.27**)
Bacillus subtilis	0.18 (4.65)	0.08 (0.75)	**0.05** (**0.15**)
Saccharomyces cerevisiae	0.14 (2.76)	**0.07** (**0.40**)	0.07 (0.54)
Homo sapiens	0.12 (3.12)	**0.09** (**1.86**)	0.14 (5.09)

distribution although the case of *H. sapiens* represented the small P-value for the KS test (i.e., the low probability that the distribution is similar between models and observed data).

However, there is no great difference of the prediction accuracy for estimating the impact degree distributions between the power-law model and the Poisson model; thus, the Poisson model may be useful for a rough estimate of the impact degree distributions.

Intrinsically, the distribution of the total number of offsprings (i.e., $P(r)$) is not significantly different between the power-law model and the Poisson model in the case of smaller r. As a simple example, we here consider the case of $\mu = 1$ (i.e., the critical case). In this case, the impact degree distributions $P(r)$ of the power-law model and the Poisson model correspond to $\propto r^{-(1+1/\gamma)}$ and $\propto r^{-3/2}$, respectively. Especially, $P(r)$ of the power-law model is the power-law distribution with the exponent ranging between -2 and -1.5 because of $1 < \gamma < 2$; thus it is not critically different from $r^{-3/2}$ of the Poisson model for smaller r. Since the impact degrees of real metabolic networks (i.e., r) were relatively small ($r < 60$), there might be no great difference of the prediction accuracy between these models.

The similarity of predicted distributions of the total number of offsprings between the Poisson model and the power-law model is also explained using Otter's theorem (Theorem 13.1 in Ref. [21]). This theorem indicates that the distribution $P(r)$ of the total number of offsprings has the universal property of a power-law tail with the exponent $-3/2$ as $r \to \infty$ under mild conditions on offspring distributions, and it implies that the types of offspring distributions hardly influence the distribution of the total number of offsprings (i.e., impact degree distributions). Note that Otter's theorem does not contradict with the analytical distribution $P(r)$ of the power-law model because this theorem is not directly applicable to the power-law model due to the different assumptions in the derivation of $P(r)$ between the power-law model and Otter's theorem.

The prediction performance is influenced by the assumption of offspring distributions and the fidelity of branching processes. To purely evaluate the validity of branching process approximation, the empirical model is useful because of no assumption of offspring distributions. It is expected that the empirical model show the best prediction accuracy because of using empirical offspring distributions. In the case of bacteria (i.e., *E. coli* and *B. subtilis*), this expectation is true, suggesting that the branching process approximation is useful for estimating the impact degree distributions. In the case of eukaryotes (i.e., *S. cerevisiae* and *H. sapiens*), on the other hand, we observed the unexpected results: the empirical model shows the relatively-low prediction accuracy. Especially, the prediction accuracy of the empirical model is lowest in the case of *H. sapiens* (human). This result implies limitations to the estimation of impact degree distributions based on the branching process approximation in the case of metabolic networks of eukaryotes (i.e., higher organisms).

A limitation of the branching process approximation is that we need to assume tree structures of networks (i.e., no cycles). The presence of cycles may lead to an overestimation of the number of offsprings d_i for each reaction, because some offsprings of a progenitor (reaction node) may have already been inactivated due to cycle structures. From this reason, the models may overestimate the impact degree distributions. On the other hand, however, some reactions with more than one incoming edge may be inactivated in the presence of cycles, if all their parents are inactivated through different paths. In this case, the number of offsprings is underestimated. The estimation of the number of offsprings depends on the relative importance of these two effects, and it may be not simple. The difficulty in this estimation is also a limitation of the model.

The branching process model needs to be improved by considering additional assumptions other than metabolic network structures in order to obtain better predictions. For example, the assumption of variable propagations in the branching process [35], in which the mean of offspring distributions differs at each propagation stage, may be useful because the degree of propagation may depend on metabolic dynamics such as gene expressions and metabolite concentrations. To apply this modified branching process model, time-series data on metabolic dynamics after gene disruptions, which are obtained by metabolomic analysis, are necessary for estimating the mean of offspring distribution at each propagation stage (i.e., time after gene disruptions). Since such data are unavailable at present, however, it is difficult to apply this modified branching process model. Similarly, other biologically-suitable assumptions are hardly determined because of few observed data on metabolic dynamics. Although there are above constraints on observed data on metabolic dynamics, we believe that the consideration of the additional information improves the prediction of impact degree distributions. In the future, the improvement of the prediction of the impact degree distributions using the branching process approximation may be possible with the available data on metabolic dynamics.

The branching process approximation is useful for estimating the impact degree distributions in metabolic networks although it has the above limitations; thus, it may be a powerful tool for evaluating a robustness of biological systems.

Acknowledgments

References

[1] H. Kitano, Cancer as a robust system: implications for anticancer therapy, Nat. Rev. Cancer 4 (2004) 227–235.

[2] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. 38 (2010) D355–D360.

[3] I.M. Keseler, et al., EcoCyc: a comprehensive database of Escherichia coli biology, Nucleic Acids Res. 39 (2011) D583–D590.

[4] J.A. Papin, J. Stelling, N.D. Price, S. Klamt, S. Schuster, B.O. Palsson, Comparison of network-based pathway analysis methods, Trends Biotechnol. 22 (2004) 400–405.

[5] V. Acuña, F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M.F. Sagot, L. Stougie, Modes and cuts in metabolic networks: complexity and algorithms, BioSystems 95 (2009) 51–60.

[6] A.P. Burgard, P. Pharkya, C.D. Maranas, OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization, Biotechnol. Bioeng. 84 (2003) 647–657.

[7] U.U. Haus, S. Klamt, T. Stephen, Computing knock-out strategies in metabolic networks, J. Comput. Biol. 15 (2008) 259–268.

[8] S. Klamt, E.D. Gilles, Minimal cut sets in biochemical reaction networks, Bioinformatics 20 (2004) 226–234.

[9] J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, S. Schuster, Structural robustness of metabolic networks with respect to multiple knockouts, J. Theoret. Biol. 252 (2008) 433–441.

[10] D. Deutscher, I. Meilijson, S. Schuster, E. Ruppin, Can single knockouts accurately single out gene functions? BMC Syst. Biol. 2 (2008) 50.

[11] T. Handorf, O. Ebenhöh, R. Heinrich, Expanding metabolic networks: scopes of compounds, robustness, and evolution, J. Mol. Evol. 61 (2005) 498–512.

[12] N. Lemke, F. Herédia, C.K. Barcellos, A.N. dos Reis, J.C.M. Mombach, Essentiality and damage in metabolic networks, Bioinformatics 20 (2004) 115–119.

[13] Z. Li, R.-S. Wang, X.-S. Zhang, L. Chen, Detecting drug targets with minimum side effects in metabolic networks, IET Syst. Biol. 3 (2009) 523–533.

[14] P. Sridhar, B. Song, T. Kahveci, S. Ranka, Mining metabolic networks for optimal drug targets, in: Proc. Pacific Symposium on Biocomputing 2008, 2008, pp. 291–302.

[15] T. Tamura, K. Takemoto, T. Akutsu, Finding minimum reaction cuts of metabolic networks under a Boolean model using integer programming and feedback vertex sets, Int. J. Knowl. Discov. Bioinform. 1 (2009) 14–31.

[16] A.G. Smart, L.A.N. Amaral, J.M. Ottino, Cascading failure and robustness in metabolic networks, Proc. Natl. Acad. Sci. USA 105 (2008) 13223–13228.

[17] D. Jiang, S. Zhou, Y.-P.P. Chen, Compensatory ability to null mutation in metabolic networks, Biotechnol. Bioeng. 103 (2009) 361–369.

[18] Y. Cong, T. Tamura, T. Akutsu, W.-K. Ching, Efficient computation of impact degrees for multiple reactions in metabolic networks with cycles, in: Proc. ACM Third International Workshop on Data and Text Mining in Bioinformatics, 2009, pp. 67–70.

[19] D.-S. Lee, K.-I. Goh, B. Khang, D. Kim, Branching process approach to avalanche dynamics on complex networks, J. Korean Phys. Soc. 44 (2004) 633–637.

[20] A. Saichev, A. Helmstetter, D. Sornette, Power-law distributions of offspring and generation numbers in branching models of earthquake triggering, Pure Appl. Geophys. 162 (2005) 1113–1134.

[21] T.R. Harris, Theory of Branching Processes, Dover, New York, 1989.

[22] I. Dobson, B.A. Carreras, D.E. Newman, A loading-dependent model of probabilistic cascading failure, Probab. Engrg. Inform. Sci. 19 (2005) 15–31.

[23] J. Kim, I. Dobson, Approximating a loading-dependent cascading failure model with a branching process, IEEE Trans. Reliab. 59 (2010) 691–699.

[24] A. Wagner, D.A. Fell, The small world inside large metabolic networks, Proc. R. Soc. Lond. Ser. B 268 (2001) 1803–1810.

[25] S. Light, P. Kraulis, A. Elofsson, Preferential attachment in the evolution of metabolic networks, BMC Genomics 6 (2005) 159.

[26] C. Pál, B. Papp, M.J. Lercher, Adaptive evolution of bacterial metabolic networks by horizontal gene transfer, Nat. Genet. 37 (2005) 1372–1375.

[27] D. Vitkup, P. Kharchenko, A. Wagner, Influence of metabolic network structure and function on enzyme evolution, Genome Biol. 7 (2006) R39.

[28] I. Dobson, B.A. Carreras, D.E. Newman, A branching process approximation to cascading load-dependent system failure, in: Proc. 37th Annual Hawaii International Conference on System Sciences, 2004.

[29] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, Nature 407 (2000) 651–654.

[30] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.

[31] A. Saichev, D. Sornette, Anomalous power law distribution of total lifetimes of branching processes: application to earthquake aftershock sequences, Phys. Rev. E 70 (2004) 046123.

[32] M.E.J. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. 46 (2005) 323–351.

[33] http://www.genome.jp/kegg/pathway.html.

[34] http://www.genome.jp/kegg/catalog/org_list.html.

[35] I. Dobson, B.A. Carreras, Number and propagation of line outages in cascading events in electric power transmission systems, in: Proc. 48th Annual Allerton Conference on Communication, Control & Computing, 2010.