Vol.2009-BIO-18 No.8
2009/9/18

Efficient Computation of Impact Degrees for Multiple Reactions in Metabolic Networks with Cycles

Yang Cong ,[†1] Takeyuki Tamura ,[†2]
Tatsuya Akutsu [†2] and Wai-Ki Ching [†1]

Analysis of the robustness of a metabolic network against deletion of single or multiple reaction(s) is useful for mining important enzymes/genes. For that purpose, the impact degree was proposed by Jiang et al. In this article, we extend the impact degree for metabolic networks containing cycles and develop a simple algorithm for its computation. Furthermore, we propose an improved algorithm for computing impact degrees for deletions of multiple reactions. The results of preliminary computational experiments suggest that the improved algorithm is several tens of times faster than a simple algorithm.

1. Introduction

In bioinformatics, it is very important to extract knowledge from biological networks. In particular, many studies have been done on metabolic networks because many and rather accurate network data are available from such databases as KEGG[6] and EcoCyc[7]. Among various problems on metabolic networks, we focus in this article on analysis of robustness because robustness is an important feature of biological systems.

In order to analyze the robustness of metabolic networks, the *flux balance analysis* (FBA) methods have been extensively studied. Among various approaches based on FBA, *elementary flux modes* (EFMs) play a key role, where an EFM is a minimal set of reactions that can operate at steady state[9]. Based on FBA and/or EFM, several works have been done on finding a minimum set of enzymes/reactions deletion of which leads to prevention of the production of a specified set of compounds, which is called a *minimum reaction cut*[1),8)]. Recently, Behre et al. proposed a measure of structural robustness based on the number of remaining EFMs after knockout vs. the number of EFMs in the unperturbed situation[2]. Deutscher et al. proposed another measure using the Shaply value from the game theory[3]. However, applications of most of the above mentioned methods were limited to the middle-scale metabolic networks. One of the reasons is that EFM based methods are not efficient. Indeed, Klamt and Stelling showed that the number of EFMs grows exponentially with the network size[8], and Acuña et al. showed that finding a minimum reaction cut is NP-hard[1]. Furthermore, stoichiometry parameters, which are required for applying FBA-based methods, are not always easy to obtain. Therefore, other approaches should also be studied.

In order to study larger scale metabolic network data, Boolean models of metabolic networks have recently been studied[4),5),10),11)]. In particular, Jiang et al. introduced the concept of *impact degree*[5]. The impact degree is defined as the number of reactions inactivated by deleting a specified reaction (or a set of specified reactions). Impact degrees are useful both for analyzing the robustness of metabolic networks and for mining influential enzymes/genes (e.g., drug targets) from metabolic networks data. However, cycles are not taken into account in their method. Since cycles are important components of metabolic networks, it would be desirable to take the effects of cycles into account.

In this article, we propose an extension of the impact degree for metabolic networks with cycles. In order to define the impact degree for networks with cycles, we modify the concept of the *maximal valid assignment* and its computation method proposed in Ref. 11). Furthermore, we propose an improved algorithm for computing the impact degrees for deletions of multiple reactions simultaneously. The results of preliminary computational experiments suggest that the improved algorithm is several tens of times faster than the simple algorithm when applied to computation of impact degrees for all reactions pairs.

2. Impact of Single Deletion

We extend the definition of *impact degree* introduced in Ref. 5) so that cycles can be treated. Analysis of metabolic networks including cycles usually becomes

†1 Department of Mathematics, The University of Hong Kong
†2 Bioinformatics Center, Institute for Chemical Research, Kyoto Univerty

harder because there may exist multiple stable global states. In order to uniquely determine the stable global state, the concept of the *maximal valid assignment* was introduced in Ref. 11) using a Boolean model of metabolic networks. Here, we give a new definition of the impact degree by combining these two concepts. Though the impact degree is formally defined using the maximal valid assignment, we give a procedural definition due to the page limit. This definition also provides an algorithm for computing the impact degree, which we call SIMPLE ALGORITHM.

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 $V_c \cap V_r = \{\}$. Let $V = V_c \cup V_r$. It is to be noted that most reactions are catalyzed by enzymes and thus each reaction can be disabled in most cases by disruption of a gene corresponding to the enzyme catalyzing the reaction.

A *metabolic network* is defined as a directed graph $G(V, E)$ satisfying the following conditions: For each edge $(u, v) \in E$, either $(u \in V_c) \wedge (v \in V_r)$ or $(u \in V_r) \wedge (v \in V_c)$ holds. The state of each reaction (or compound) is quantized to two levels: non-disabled (or activated) represented by 1 and disabled (or inactivated) represented by 0.

To calculate the impact degree of reaction R_i, we first only delete reaction R_i ($R_i = 0$, and $R_j = 1$ for all $j \neq i$) and activate all the compounds ($C_k = 1$). Then we deduce the states of reactions and compounds according to 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. Reaction should be inactivated if any consumed compound or produced compound is inactivated.
3. For each compound, there are three different reactions: consuming reactions, producing reactions, and directly unrelated reaction.
4. Compound should be inactivated if all its consuming reactions or all its producing reactions are inactivated.

We repeat the above procedure until the states are stable. The impact degree of the reaction is the number of inactivated reactions (represented by 0).

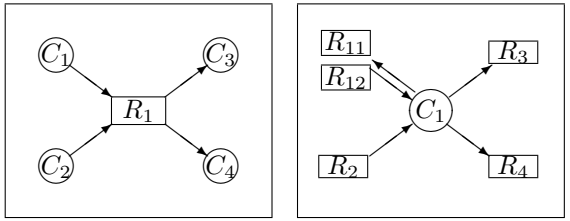

Fig. 1 Examples of reactions and compounds.

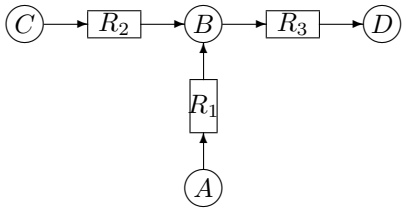

Fig. 2 An example of metabolic network.

The above rules for reaction and compound can be represented by Boolean Functions. In Figure 1, the state of reaction R_1 is determined by $R_1 = (C_1 \wedge C_2) \wedge (C_3 \wedge C_4)$, and the state for compound C_1 is determined by $C_1 = (R_{12} \vee R_2) \wedge (R_{11} \vee R_3 \vee R_4) = (R_1 \vee R_2) \wedge (R_1 \vee R_3 \vee R_4)$. There are two kinds of reactions, reversible reactions and irreversible reactions. We divide a reversible reaction into two irreversible reactions with opposite directions. In Figure 1, a reversible reaction is divided into two irreversible reactions R_{11} and R_{12}.

We can prove that the number of inactivated reactions and compounds increases monotonically and thus converges within $m + n$ repetitions. We can also prove that the impact degree calculated by SIMPLE ALGORITHM is the same as the number of reactions assigned to 0 in the maximal valid assignment[11], where both production pathways and degradation pathways are taken into account here. The proofs are omitted since we focus on experimental parts in this article.

We use Figure 2 to illustrate how to calculate the impact degree. To calculate the impact degree of reaction R_1, we first set $R_1(0) = 0, R_2(0) = R_3(0) = 1, A(0) = B(0) = C(0) = D(0) = 1$. For compounds, we let $A(t + 1) = R_1(t)$,

$B(t+1) = (R_1(t) \vee R_2(t)) \wedge R_3(t)$, $C(t+1) = R_2(t)$ and $D(t+1) = R_3(t)$, then we have $A(1) = 0$, $B(1) = 1$, $C(1) = 1$, $D(1) = 1$, and $R_i(1) = R_i(0)$ for $i = 1, 2, 3$. For reactions, we let $R_1(t+1) = A(t) \wedge B(t)$, $R_2(t+1) = B(t) \wedge C(t)$, and $R_3(t+1) = B(t) \wedge D(t)$, then we have $R_1(2) = 0$, $R_2(2) = R_3(2) = 1$ and $A(2) = A(1)$, ..., $D(2) = D(1)$. Note that we let $R_1(t) = 0$ for all t since R_1 is deleted. Then, the states become stable and thus the impact degree for reaction R_1 is 1. In the same way, we know that the impact degrees for deletion of R_2 and deletion of R_3 are one and three, respectively.

3. Impact of Multiple Deletion

SIMPLE ALGORITHM given in Section 2 can be trivially applied to the computation of impact degree for multiple reactions. Without loss of generality, we consider the deletion of two reactions simultaneously. Suppose that R_g and R_h are deleted. Then, we start with $R_g(0) = R_h(0) = 0$ and $R_i(0) = 1$ for all $i \neq g, h$ and $C_i(0) = 1$ for all i. The corresponding pseudo code is as follows:

SIMPLE ALGORITHM(R_g, R_h):
  Initialize
    For $i = 1$ **to** n **do**
      If $i = g$ or $i = h$, **then** set $R_i = 0$
      else set $R_i = 1$.
    For $j = 1$ **to** m **do**
      set $C_j = 1$.
    set $t = 0$, $M(t) = [R_1, R_2, \ldots, R_n, C_1, C_2, \ldots, C_m]$
  While $M(t) \neq M(t-2)$ **do**
    For $j = 1$ **to** m **do**
      If $C_j \neq 0$, **then**
        set $C_j = (R^1_{pro} \vee \ldots \vee R^{p_j}_{pro}) \wedge (R^1_{con} \vee \ldots \vee R^{q_j}_{con})$.
    set $t = t+1$, $M(t) = [R_1, \ldots, R_n, C_1, \ldots, C_m]$
    For $i = 1$ **to** n **do**
      If $R_i \neq 0$, **then**
        set $R_i = (C^1_{pro} \wedge \ldots \wedge C^{u_j}_{pro}) \wedge (C^1_{con} \wedge \ldots \wedge C^{v_j}_{con})$
    set $t = t+1$, $M(t) = [R_1, \ldots, R_n, C_1, \ldots, C_m]$
    set $impact = 0$
    For $i = 1$ **to** n **do**
      If $R_i = 0$, **then** $impact = impact + 1$.
    Return $M(t)$ and $impact$.

Here for compound C_j: R^k_{pro} (R^k_{con}) is the kth producing (consuming) reaction and p_j (q_j) is the number of producing (consuming) reactions. For reaction R_i: C^k_{pro} (C^k_{con}) is the kth produced (consumed) compound and u_j(v_j) is the number of produced (consumed) compounds.

The *impact vector* $M(t) = [R_1, \ldots, R_n, C_1, \ldots, C_m]$ is a $1 \times (n+m)$ row vector, where the values of disabled reactions and inactivated compounds are 0.

However, it would take long CPU time if the impact degrees for all pairs of reactions should be computed. Therefore, we develop an efficient algorithm (called as IMPROVED ALGORITHM) for computing the impact degrees for all pairs of reactions, where it can be generalized for triplets, quadruplets, ... of reactions.

In order to explain IMPROVED ALGORITHM, we begin with a simple example. In the metabolic network shown by Figure 3, deletion of reaction R_1 impacts reactions R_1, R_2 and compounds B, C. Deletion of reaction R_3 impacts only reaction R_3. The deletion of reaction pair (R_1, R_3) impacts reactions R_1, R_2, R_3 and compounds B, C. In the aspect of reaction and compound, the impact of reaction pair (R_1, R_3) is the sum of impacts of deleting reaction R_1 and reaction R_3 separately. We call this case as *simplified case*.

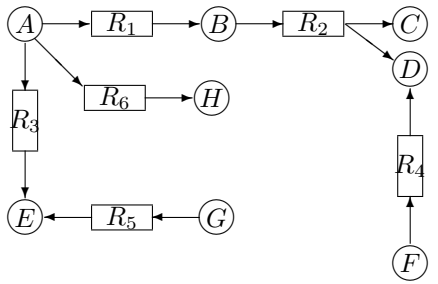

Fig. 3 An example for deletion of multiple reactions.

For reaction R_i, *related reactions* are defined as the reactions disabled by deletion of R_i. We define *inactivated compounds* as all the compounds inactivated. *Related compounds* are defined as all the consumed and produced compounds for all the related reactions. *Remained compounds* of reaction R_i are defined as the compounds related but can not be inactivated by reaction R_i. Table 1 lists the relationship among reactions and compounds in the metabolic network shown in Figure 3.

Overlapped compounds are defined as the compounds that are remained for both reaction R_g and reaction R_h. For reaction pair (R_1, R_3) in Figure 3 , the overlapped compound is compound A. Since $A = R_1 \vee R_3 \vee R_6$, compound A can not be inactivated by reaction pair (R_1, R_3). The impact of R_1 and R_3 can not be extended to any other compound except those inactivated by single deletion of R_1 or R_3. Thus the impact of the reaction pair can not extend to any reaction not related to R_1 and R_3. This is why the reaction pair (R_1, R_3) is a simplified case. For reaction pair (R_1, R_5), there is no overlapped compound. Obviously, the impact of the reaction pair only stays among the reactions related to R_1 and R_5.

Table 1 Relationship among Reaction and Compound

R	R_{relate}	$C_{inactivate}$	C_{relate}	C_{remain}
R_1	R_1, R_2	B, C	A, B, C, D	A, D
R_2	R_1, R_2	B, C	A, B, C, D	A, D
R_3	R_3	$-$	A, E	A, E
R_4	R_4	F	D, F	D
R_5	R_5	G	E, G	E
R_6	R_6	H	A, H	A

For reaction pair (R_g, R_h), if any of the following two conditions is satisfied, then we have the simplified case. One condition is that there is no overlapped compound, e.g. reaction pair (R_1, R_5). The other is, after setting all the related reactions to R_g and R_h disabled, no overlapped compound can be inactivated, e.g. reaction pair (R_1, R_3) or (R_1, R_2). Then, the impact of the reaction pair is computed from the bitwise AND of the impact vectors for R_g and R_h.

On the other hand, if there exists at least one overlapped compound that can be inactivated, then we need to check the impact for the reaction pair, e.g. reaction pair (R_2, R_4).

Based on the above ideas, we develop IMPROVED ALGORITHM as follows. We utilize the impact vector of single deletion, where we assume that a single impact vector $\mathbf{v}_g$ (i.e., 0-1 vector representing reactions and compounds impacted by deletion of R_g) is already computed for every reaction R_g.

IMPROVED ALGORITHM(R_g, R_h)
For $i = 1$ **to** n **do** $R_i := \mathbf{v}_g(i) \wedge \mathbf{v}_h(i)$.
For $j = 1$ **to** m **do** $C_j := \mathbf{v}_g(n+j) \wedge \mathbf{v}_h(n+j)$.
$t := 0$, $M(t) := [R_1, R_2, \ldots, R_n, C_1, C_2, \ldots, C_m]$.
If there exist overlapped compounds $(C_{1'}, \ldots, C_{s'})$ **then**
  $flag := 0$.
  For $k = 1$ **to** s **do**
    $C_{k'} := (R^1_{pro} \vee \ldots \vee R^{p_{k'}}_{pro}) \wedge (R^1_{con} \vee \ldots \vee R^{q_{k'}}_{con})$.
    If $C_{k'} = 0$ **then** $flag := 1$.
If $s = 0$ or $flag = 0$ **then return** $\sum_{i=1}^{n}(1 - R_i)$.
  /* simplified case */

If $flag = 1$ **then**
  While $M(t) \neq M(t-1)$ **do**
    For $j = 1$ **to** m **do**
      If $C_j \neq 0$ **then**
        $C_j := (R^1_{pro} \vee \ldots \vee R^{p_j}_{pro}) \wedge (R^1_{con} \vee \ldots \vee R^{q_j}_{con})$.
        If $C_j = 0$ **then** $R^1_{pro} := 0, \ldots, R^{p_j}_{pro} := 0, R^1_{con} := 0, \ldots, R^{q_j}_{con} := 0$.
    $t := t + 1$, $M(t) := [R_1, \ldots, R_n, C_1, \ldots, C_m]$.
  Return $\sum_{i=1}^{n}(1 - R_i)$.

In the above, $R^1_{pro}, \ldots, R^{p_j}_{pro}$ and $R^1_{con}, \ldots, R^{q_j}_{con}$ denote producing reactions and consuming reactions for C_j respectively, and $\mathbf{v}_k(i)$ denotes the value of the i-th position in vector $\mathbf{v}_k$.

4. Computational Experiments

We extracted 253 reactions and 261 compounds of E. coli metabolic network

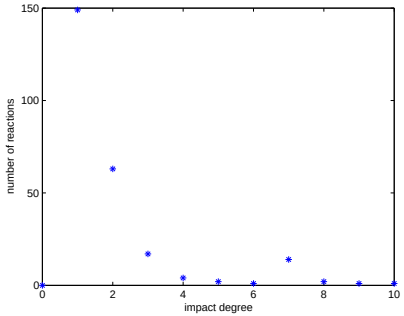

Fig. 4 Impact degree for single deletion.

from the KEGG database[6], among which 150 reactions are reversible. All experiments were conducted on this extracted subnetwork.

Figure 4 shows the distribution of impact degree by single deletion. The average impact degree among all the 253 reactions was 1.9331. In Ref. 5), the average impact among all the 3377 reactions in KEGG database was ~ 1.98. Though our network is a subnetwork of Ref. 5), similar results were obtained. In Figure 4, we can observe a peak at the impact degree 7. This is because there are two groups of 7 reactions joining together in a chain shape. In each chain, the only producing compound of one reaction is the only consuming compound of the other reaction. The genes with high impact degrees are listed in Table 2, where GO (Gene Ontology) ID numbers are also shown if they are available, and we could not identify genes for some reactions.

Table 2 Genes with High Impact Degrees

impact	genes
9	fabD(GO:0004314)
8	ubiG, ubiC (0008813)
7	ispD, ispE, ispF, ispG, ispH dxr (GO:0008661), dxs, ubiB

Figure 5 provides the distribution of the impact degrees of all the 32131 two-reaction pairs. The average impact degree is 3.8461. It is interesting that a peak is found at the impact degree 8. The existence of seven-reaction chains is a possible explanation (e.g., seven + one from a deleted pair).

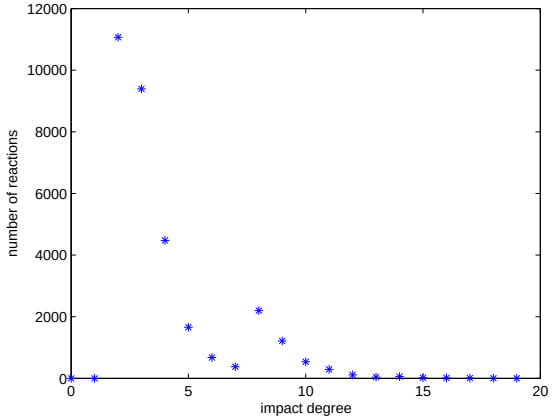

Fig. 5 Impact degree for two-reaction deletion.

For our metabolic network, there are 32045 simplified cases (99.73%) against 32131 reaction pairs in total. For computation of the impact degrees for all pairs of two-reaction, SIMPLE ALGORITHM took 3427.7 seconds, whereas IMPROVED ALGORITHM took 88.9 seconds. It means that IMPROVED ALGORITHM is 38.5 times faster than SIMPLE ALGOITHM (in this case). All experiments were performed via MATLAB 7.0 in Windows XP using an Intel 1.86 GHz processor with 512 MB RAM.

5. Concluding Remarks

In this article, we have proposed algorithms for computing the impact degrees of deletions of single or multiple reaction(s) in a metabolic network including cycles. Though we examined the cases of deletions of single reaction and two reactions, our algorithms can be extended for deletions of more than two reactions. Analysis of the results from a biological viewpoint is left as future work.

Vol.2009-BIO-18 No.8
2009/9/18

References

1) V. Acuña, F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M. F. Sagot and L. Stougie. Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems*, 95(1):51–60, 2009.
2) J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin and S. Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of Theoretical Biology*, 252(3):433–441, 2008.
3) D. Deutscher, I. Meilijson, S. Schuster and E. Ruppin. Can single knockouts accurately single out gene functions? *BMC Bioinformatics*, 2:50, 2008.
4) T. Handorf, N. Christian, O. Ebenhöh and D. Kahn. An environmental perspective on metabolism. *Journal of Theoretical Biology*, 252(3):530–537, 2008.
5) D. Jiang, S. Zhou and Y-P. P. Chen. Compensatory ability to null mutation in metabolic networks. *Biotechnology and Bioengineering*, 103(2):361–369, 2009.
6) M. Kanehisa *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36:D480–D484, 2008.
7) P. D. Karp *et al.* Multidimensional annotation of the Escherichia coli K-12 genome. *Nucleic Acids Research*, 35:7577–7590, 2007.
8) S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Molecular Biology Reports*, 29:233–236, 2002.
9) J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22(8), 400–405, 2004.
10) P. Sridhar, B. Song, T. Kahveci and S. Ranka. Mining metabolic networks for optimal drug targets. *Proc. Pacific Symposium on Biocomputing 2008*, 291–302, 2008.
11) T. Tamura, K. Takemoto and T. Akutsu. Finding minimum reaction cuts of metabolic networks under a Boolean model using integer programming and feedback vertex sets. *International Journal of Knowledge Discovery in Bioinformatics*, in press.