Metabolic network destruction: relating topology to robustness

Wynand Winterbach[1,2], Huijuan Wang[2], Marcel Reinders[1], Piet Van Mieghem[2], and Dick de Ridder[1]

[1] The Delft Bioinformatics Lab, Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands,
WWW home page: http://bioinformatics.tudelft.nl
[2] Network Architecture and Services, Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
WWW home page: http://www.nas.ewi.tudelft.nl

Summary. Biological networks exhibit intriguing topological properties such as small-worldness. In this paper, we investigate whether the topology of a metabolic network is related to its robustness. We do so by perturbing a metabolic system *in silico*, one reaction at a time and studying the correlations between growth, as predicted by flux balance analysis, and a number of topological metrics, as computed from three network representations of the metabolic system.

We find that a small number of metrics correlate with growth and that only one of the network representations stands out in terms of correlated metrics. The most correlated metrics point to the importance of hub nodes in this network: so-called "currency metabolites". Since they are responsible for interconnecting distant functional modules in the network, they are important points in the networks for predicting if reaction removal affects growth.

Source code and data are available upon request.

Key words: metabolic networks, flux balance analysis, network topology, robustness

1 Introduction

In the last decade, advances in high-throughput biological measurement systems have made it possible to extract large-scale networks from biological systems. Jeong *et al.* [7] were among the first to study the topologies of metabolic networks, networks of interconversions of small compounds. The metabolic networks of the 43 organisms that they studied gave evidence of a scale-free structure. Characteristic properties of these so-called "small-world" networks are their power-law distributed node degrees and their small average shortest path lengths.

Subsequently, researchers studied the topologies of a number of other types of biological networks [2, 12, 3]. Much of this work confirmed the Jeong *et al.* results: scale-free behavior was everywhere. Even the internet and some power grids are thought to display scale-free behavior [1]. These latter networks have expanded

in a seemingly organic fashion through a process of *preferential attachment* − new nodes are more likely to attach to existing high-degree nodes than to low-degree nodes. This expansion process forms the basis of Barabási and Albert's [1] random network model. They show that the process leads to the characteristic power-law node degree distribution and small-world properties.

Although Kim *et al.* [8] and Lima-Mendez *et al.* [11] argue that biological networks do not develop through simple processes of preferential attachment, the presence of similar topological elements, such as hub nodes, begs the question whether these topological properties confer some benefit or whether certain topologies are inherently suited for particular functionality. In an effort to understand the relationship between the function of a network and its topological properties, Milo *et al.* [13] introduced the concept of *motifs*. A motif is a small sub-network (with at most four nodes) whose over-representation may be indicative of its role in maintaining function at a local level. They found that certain motifs occur more often in biological networks than expected by chance and that they may correspond to certain desired behavior such as response acceleration, signal delay and stability. Prill *et al.* [19] took this idea further and claimed that certain motifs were inherently more prone to display stable behavior than others. By abstracting away from the underlying functionality, they demonstrated that such relations held to some extent over a variety of biological networks. However, Ingram *et al.* [6] considered gene networks and compared the results of a differential equation model of gene expression to specific motif counts in the gene network but found no correlation. Lima-Mendez *et al.* [11] argue that the global topological properties cannot explain the function of networks. According to them, the significance of motif frequencies may have been overestimated (since the frequencies only capture global properties), but they consider a localized approach promising as the key to understanding biological networks lies in understanding local details.

In our work, we take a global approach and investigate to what extent network topology can be related to more systems-level network properties shared by the various network types studied by Barabási *et al.* An interesting property in this respect is that of robustness. Stelling *et al.* [20] and Kitano [10] define robustness as the ability of a system to maintain its function in the face of perturbations or uncertainty. Biological systems are known to be robust [9] to many forms of perturbation while being highly sensitive to other forms, so-called "highly optimized tolerance" [20]. The question is whether there is something in the topology of these networks that confers robustness to the overall system.

In this paper, we study the relationship between the growth of a micro-organism (baker's yeast, *Saccharomyces cerevisiae*) and the topologies of three network representations of its metabolic system. Microbial metabolic systems provide a good test bed, since an often assumed objective − growth − is easily expressed in terms of fluxes through these systems. Furthermore, good quality metabolic datasets are readily available and resulting flux models can be studied computationally with high efficacy.

To study the link between network topology and robustness, we propose an *in silico* metabolic system perturbation experiment. We define robustness as the ability of the yeast cell to maintain growth under reaction removals. First, we show how its metabolic system can be represented by three different networks. Then, through a number of trials, reactions are removed from the metabolic system until growth ceases. This provides a number of snapshots of partially destructed metabolic systems. For each snapshot, growth and a number of network-wide topological metrics can be computed. By calculating correlations between growth and these metrics, we find that most of the topological metrics are not related to function. The strongest correlations point to the importance of hub nodes (so-called "currency metabolites").

2 Method
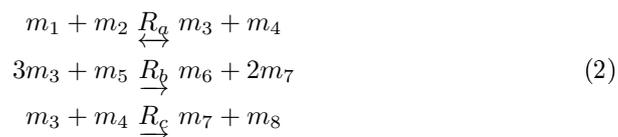
2.1 Computing function

In this work, we define robustness as the maintenance of cell growth under perturbations to the organism's metabolic system when reactions are removed from the metabolic network. A metabolic system with r reactions and m metabolites is modeled by a set of m differential equations:

$$\frac{dX_i}{dt} = s_{syn}v_{syn} - s_{deg}v_{deg} - s_{use}v_{use} + s_{trans}v_{trans} \tag{1}$$

that specify how the concentration X_i of a metabolite i changes in time. v_{syn} is the rate of metabolite synthesis, v_{deg} is the degradation rate, v_{use} is the rate of consumption (by other reactions) and v_{trans} is the rate of transport across the cell boundary (into the cell). s_{syn}, s_{deg}, s_{use} and s_{trans} are *stoichiometric coefficients*[1] (reaction rates are measured in μmol gDW^{-1} h, i.e., micromoles per gram of dry weight per hour).

We assume that $s_{trans}v_{trans}$ is a constant value b_i, allowing (1) to be written in vector form as $\frac{d\mathbf{X}}{dt} = \mathbf{S} \cdot \mathbf{v} + \mathbf{b}$, with $\mathbf{S}$ the $m \times r$ *stoichiometric matrix*, $\mathbf{v}$ an $r \times 1$ vector of reaction rates (fluxes) and $\mathbf{b}$ the vector of boundary transport reaction rates.

Here we show a small example to make the form of $\mathbf{S}$ clear (and later to show how networks are derived from $\mathbf{S}$). Consider the metabolic system:

$$\begin{aligned} m_1 + m_2 &\underset{\longleftrightarrow}{R_a} m_3 + m_4 \\ 3m_3 + m_5 &\underset{\longrightarrow}{R_b} m_6 + 2m_7 \\ m_3 + m_4 &\underset{\longrightarrow}{R_c} m_7 + m_8 \end{aligned} \tag{2}$$

The corresponding $\mathbf{S}$ matrix is shown in Figure 1a. Since each column is labeled by a reaction R_i, we refer to the corresponding flux value in $\mathbf{v}$ as v_i. At steady-state $\frac{d\mathbf{X}}{dt} = \mathbf{0}$, rendering the linear system:

---

[1] These are derived from the chemical mass balance coefficients: e.g. $2H_2 + O_2 \rightarrow 2H_2O$ corresponds to the stoichiometric coefficient vector $[-2 \quad -1 \quad 2]$

$$\mathbf{S} \cdot \mathbf{v} + \mathbf{b} = \mathbf{0}. \tag{3}$$

$$\mathbf{S} = \begin{array}{c} \\ m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \\ m_8 \end{array} \begin{array}{ccc} R_a & R_b & R_c \\ \left(\begin{array}{ccc} -1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & -3 & -1 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{array}\right) \end{array} \qquad \mathbf{S'} = \begin{array}{c} \\ m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \\ m_7 \\ m_8 \end{array} \begin{array}{cccc} R_a^+ & R_a^- & R_b & R_c \\ \left(\begin{array}{cccc} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & -3 & -1 \\ 1 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right) \end{array}$$

(a) The stoichiometric matrix from (4).   (b) The stoichiometric matrix from (6).

Fig. 1: Stoichiometric matrices of the toy problem in Section 2.1.

Due to the small size of the example, $\mathbf{S}$ is overdetermined (i.e., there are fewer reactions than metabolites). In real biological networks however, stoichiometric matrices are under-determined. Such systems generally have infinitely many solutions; but biologists are only interested in biologically significant solutions. A common (biological) assumption is that microbial cells attempt to maximize the rate of their biomass production or in other words, growth. Growth can be expressed as a linear combination $\mathbf{c}^T \cdot \mathbf{v}$ of certain key reaction rates in the metabolic system. The reaction rates can then be computed by a linear program:

$$\text{Maximize} \ \ \mu = \mathbf{c}^T \cdot \mathbf{v} \tag{4}$$
$$\text{subject to} \ \ \mathbf{S} \cdot \mathbf{v} + \mathbf{b} = \mathbf{0}$$

Positive components of $\mathbf{v}$ correspond to forward-acting reactions, whilst negative components correspond to reactions running in reverse. In (4), the components of $\mathbf{v}$ may assume negative and positive values meaning that any reaction can, in principle, occur in either direction. Due to thermodynamics, some reactions are very unlikely to occur in reverse (in the example, only reaction R_a is reversible). These constraints are modeled by restricting rates of non-reversible reactions to be non-negative. Thus for each non-reversible reaction R, the constraint $v_R \geq 0$ is added, rendering the linear system:

$$\text{Maximize} \ \ \mu = \mathbf{c}^T \cdot \mathbf{v} \tag{5}$$
$$\text{subject to} \ \ \mathbf{S} \cdot \mathbf{v} + \mathbf{b} = \mathbf{0}$$
$$v_{R_i} \geq 0 \text{ for each non-reversible reaction } R_i$$

In addition, biological constraints limit the rates of some reactions. These inequalities are simply added to the list of constraints of the linear program.

This steady-state framework for computing metabolic fluxes by optimizing some criterion is known as *flux balance analysis*. Orth *et al.* [17] give a good overview of flux balance analysis.

Testing robustness We test robustness by iteratively removing reactions and recalculating (5) until growth μ drops below a certain low threshold value (1×10^{-9} μmol gDW^{-1} h). This produces a sequence $T = \{s_1, s_1, s_2, \ldots, s_n\}$ which is referred to as the *trial T*. A *step* is a reaction label index: step s_i corresponds to the removal of reaction R_{s_i}. The steps in a trial are associated with a sequence of linear programs $P_0, P_1, P_2, \ldots, P_n$, where P_0 is the unmodified linear program (from which no reaction has been removed) and P_i is the linear program resulting from the removal of the reactions $R_{s_1}, R_{s_2}, \ldots, R_{s_i}$. Removal of a reaction is modeled by removing its corresponding column from $\mathbf{S}$.
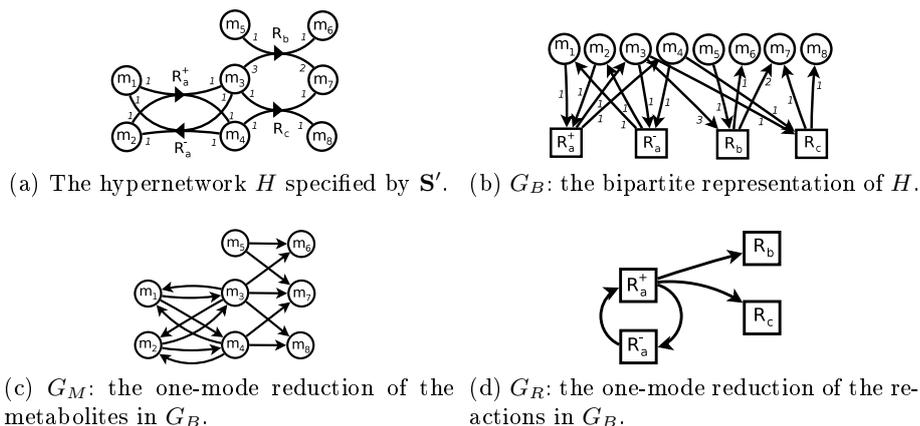
Reactions are removed uniformly (at random) when constructing a trial. Since most reactions are mediated by enzymes, removal of a reaction roughly corresponds to a gene knockout. Some reactions are not mediated by enzymes and occur due to chemical processes such as diffusion. Knocking out a gene will not bring these reactions to a halt and we therefore never remove such reactions.

2.2 Topology

To be able to calculate topological properties of the metabolic system, the stoichiometric matrix $\mathbf{S}$ should be represented as a network. However, $\mathbf{S}$ cannot be directly represented as a network since a reaction may interact with more than two metabolites and a metabolite may interact with more than two reactions. A natural representation of such a system is a *hypernetwork* in which a link may connect more than two nodes. The stoichiometric matrix represents a hypernetwork where the columns are links and the rows are nodes. The links are directed: negative values in a column represent source nodes and positive values represent target nodes. Let u be a node, and let L be a set of links that have u as their source nodes, then the target nodes of L are the out-neighbors of u. The in-neighbors are defined analogously, with u as the target node.

Note that the stoichiometric matrix derived from the linear programming formulation does not capture the reversibility of reactions (such as R_a in the example) because a reaction R_i is considered to act in reverse when its rate v_i in the linear program solution is negative. We therefore reformulate the linear program such that $\mathbf{v} \geq \mathbf{0}$ (i.e., all fluxes are positive). A reversible reaction R_i is converted to a pair of reactions R_i^+ and R_i^-; then if $\mathbf{c}_i$ is the column vector in $\mathbf{S}$ corresponding to R_i, $\mathbf{c}_i$ is replaced by two column vectors $\mathbf{c}_i^+$ and $\mathbf{c}_i^-$ (corresponding to R_i^+ and R_i^- respectively) such that $\mathbf{c}_i^+ = \mathbf{c}_i$ (the forward reaction) and $\mathbf{c}_i^- = -\mathbf{c}_i$ (the reverse reaction). Converting $\mathbf{S}$ leads to the stoichiometric matrix $\mathbf{S}'$ in Figure 1b. The hypernetwork is shown in Figure 2a.

The linear program (5) is modified with the new stoichiometric matrix $\mathbf{S}'$ and non-negative flux constraints, giving:

(a) The hypernetwork H specified by $\mathbf{S}'$.

(b) G_B: the bipartite representation of H.

(c) G_M: the one-mode reduction of the metabolites in G_B.

(d) G_R: the one-mode reduction of the reactions in G_B.

Fig. 2: The hypernetwork and networks derivable from $\mathbf{S}'$ in (6).

$$\text{Maximize} \quad \mu = \mathbf{c}^T \cdot \mathbf{v} \qquad (6)$$
$$\text{subject to} \quad \mathbf{S}' \cdot \mathbf{v} + \mathbf{b} = \mathbf{0}$$
$$\mathbf{v} \geq \mathbf{0}$$

Network theory provides many tools for studying the topological properties of normal networks, whilst there are very few metrics that can be computed on hypernetworks. Thus we considered three possible network representations of the hypernetworks specified by the stoichiometric matrix $\mathbf{S}'$. First, a hypernetwork $H(\mathcal{M}', \mathcal{R}')$ can be modeled as a bipartite network $G_B(\mathcal{M} \cup \mathcal{R}, \mathcal{L})$. The nodes in $\mathcal{M}$ represent the metabolites $\mathcal{M}'$ in H, whilst the nodes in $\mathcal{R}$ represent reaction links $\mathcal{R}'$ in H. Conversion of the hypernetwork H in Figure 2a produces the biparitite network G_B in Figure 2b. We refer to this network as the *metabolite-reaction network*[2] as it contains both metabolite nodes $\mathcal{M}$ and reaction nodes $\mathcal{R}$.

Although standard network theory techniques can be applied to G_B, its bipartite nature makes some metrics difficult or impossible to compute. For example, the clustering coefficient for any node in a bipartite network is 0. For this reason, we also considered one-mode reductions of G_B. An $\mathcal{M}$-node ($\mathcal{R}$-node) *one-mode* reduction $G'(\mathcal{N}, \mathcal{L}')$ of $G_B(\mathcal{M} \cup \mathcal{R}, \mathcal{L})$ is a network that contains only nodes from the set $\mathcal{M}$ (the set $\mathcal{R}$) such that for each directed link $l = (n_1, n_2) \in \mathcal{L}'$ there is a node $n_3 \in \mathcal{R}$ ($n_3 \in \mathcal{M}$) such that $(n_1, n_3) \in \mathcal{L}$ and $(n_3, n_2) \in \mathcal{L}$ (note that there may be many nodes n_3 that satisfy this condition). We call the $\mathcal{M}$-node one-mode reduction simply the *metabolite network* G_M (shown in Figure 2c) and likewise the $\mathcal{R}$-node one-mode reduction simply the *reaction network* G_R (illustrated in Figure 2d).

---

[2] This representation is exactly the Petri-net representation [14, 4] of the metabolic system.

Note that it is possible to represent the link weights of the hypernetwork H in its bipartite representation G_B: such a mapping can be seen in Figure 2b. However, there is no obvious way to map these weights to G_M or G_R. For this paper, we opted to consider only unweighted networks. Furthermore, note that when a reaction is removed from the metabolic system, the corresponding networks G_B, G_M and G_R may become disconnected. For a given network, all metrics are applied to the largest component whilst the small components are ignored.

Topological metrics For every step of each trial, a number of topological metrics were computed for each of the three network representations (where possible). The computed metrics comprise:

Newman's assortativity coefficient [15]: how likely is it for nodes with similar degrees to be connected to each other. Calculated for the out-degrees, in-degrees and undirected degrees of nodes.

Transitivity: the number of triangles in the network divided by the maximum possible number of triangles in the network. This is computed on undirected versions of the networks.

Reciprocity: the ratio of reciprocal pairs to all possible reciprocal pairs. A pair of nodes n_1 and n_2 is reciprocal if there are bi-directed links (n_1, n_2) and (n_2, n_1).

Eigenvector & betweenness centrality [16]: these metrics measure the importance of a node in a network; they are applied to undirected versions of the networks.

Maximum eigenvalue of the adjacency matrix: this metric is associated with the efficiency of dynamic processes (such as virus spreading) on networks and is computed on undirected versions of the networks.

Average in- and out-degrees of incoming and outgoing neighbors

Clustering coefficient: for a node n, the number of links spanning n's neighbors divided by the maximum possible number of links that can span n's neighbors.

Dice coefficient: if the neighbors of two nodes are the sets X and Y, the Dice coefficient of the nodes is $2|X \cap Y|/(|X| + |Y|)$, i.e., a measure of how similar their neighbor sets are. Since this coefficient is defined for pairs of nodes, a vector of coefficients is associated with each node. We compute the Dice coefficient for all outgoing neighbors, incoming neighbors and also the combination of the incoming and outgoing neighbors.

Reciprocal hopcount: the hopcount between a pair of nodes is equal to the number of links on a shortest path between them. For each node there is a vector of hop counts to all other nodes − this was reduced to a single value by computing the mean of the vector. Because the networks are directed, there are nodes which are unreachable from other nodes and are thus at an infinite distance. We dealt with the problem by using the reciprocal hop count values, thereby converting infinite distances to zero distances in the reciprocal version.

Since $G_B = G_B(\mathcal{M} \cup \mathcal{R}, \mathcal{L})$ contains two types of nodes, the metrics are applied separately to its reaction nodes $\mathcal{R}$ and metabolite nodes $\mathcal{M}$, giving two sets of results.

The metrics divide into two groups: metrics that associate a value with a network and metrics that associate a value with each node of a network. In order to compare the latter group to growth values (obtained from solving the linear program), node values (for a given metric) have to be reduced to a single value that can be associated with the network. A straightforward way to reduce node values to a single value for the network is to compute their minimum, mean or maximum values (thereby yielding three metrics). This is the approach that we took. Some metrics associate vectors of values with each node. For example, the hopcount associates with a node a list of hopcount values. In this case, we employ the above approach of computing minima, means and maxima to reduce complex node values to simple node values. Once that is done, the node values can be reduced to network values.

In our experiments, many reactions always have zero reaction rates (as predicted by the flux balance linear program). These reactions contribute links and nodes to the network representations whilst their removal does not influence growth. We excluded these reactions when constructing $G_B = G_B(\mathcal{M} \cup \mathcal{R}, \mathcal{L})$ by letting $\mathcal{R}$ be the set of all reactions that have non-zero reaction rates in at least one step of one trial and $\mathcal{M}$ the metabolites that interact with the reactions in $\mathcal{R}$.

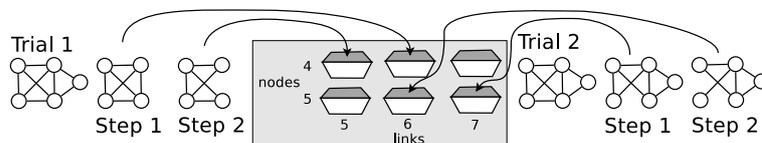### 2.3 Relating growth and topology

Fig. 3: An example of binning for two steps of two trials.

For each trial (i.e., sequence of reaction removals) we compute a sequence of growth values (computed from the linear program discussed in Section 2.1) and three sequences of networks, one for each representation. For each network, a set of topological metrics is calculated. This allows us to relate growth to topology.

An obvious first choice for calculating the relationship is, for each individual trial, to compute correlation coefficients between the growth sequence and each of the sequences of topological metrics. However, apparent correlations found by this method may simply be side-effects of the network size decreasing as we remove reactions. We reduce the impact of this incidental correlation by binning the steps from all of the trials: trial-step pairs whose corresponding networks have similar numbers of nodes and links are placed into the same bin. This process is

illustrated in Figure 3: here one sees network sequences from two trials placed into bins (the bin width here is 1 for both nodes and links). In our experiments, we used a bin width of 2 nodes $\times$ 4 links – i.e., in a bin, node counts can differ by 1 and link counts by 3.

Bin number	1	2	3	4	5
# links	2685-2688	2689-2692	2693-2696	2697-2700	2701-2704
Pearson correlation ρ_i	0.342	0.286	0.322	0.236	0.172
# items in bin n_i	889	935	907	959	936

Table 1: A few bins showing correlations between growth and an unspecified network metric.

Since a bin contains numerous steps, it is possible to correlate growth with any of the topological metrics. We used the Pearson correlation coefficient to compute, for a given topological metric, a correlation value for every bin. An example of bin correlations is shown in Table 1 (here binning is only performed using link counts, with a bin width of 4 links). The per-bin results for each metric were then averaged, weighted by the number n_i of items in each bin: $\bar{\rho} = \frac{\sum_i n_i \rho_i}{\sum_i n_i} = 0.27$ in our example. For each topological metric, this yields one value $\bar{\rho}$ indicating the strength of its correlation with growth.

2.4 Experimental setup

We used the genome-scale metabolic data set which is available from the UCSD Systems Biology Research Group website [18]. The website provides a minimal aerobic growth environment which was used for our experiments. In this experiment,

- the rate of the ATP maintenance reaction (ATPM) is limited to 1 μmol gDW^{-1} h whilst the acetyl-CoA hydrolase (ACOAH) and the glutamate synthase for NADH (GLUSx) reactions are disabled;
- the reaction rates of reactions that transport O_2, NH_4, SO_4, P_i, H_2O, K, Na and CO_2 are unconstrained.

3 Results and discussion

We performed one thousand *in silico* reaction removal trials, calculating growth and all topological metrics (on G_B, G_M and G_R) for each step of every trial. Taking care to reduce the effect of a decrease in the number of links during network destruction (through binning), we then calculated the correlations $\bar{\rho}$ between growth and each of these metrics. We then filtered our results by considering only those metrics for which $\bar{\rho} \geq 0.25$ (this threshold was chosen to be small, but the choice is somewhat arbitrary). The thresholded $\bar{\rho}$ correlations for G_B,
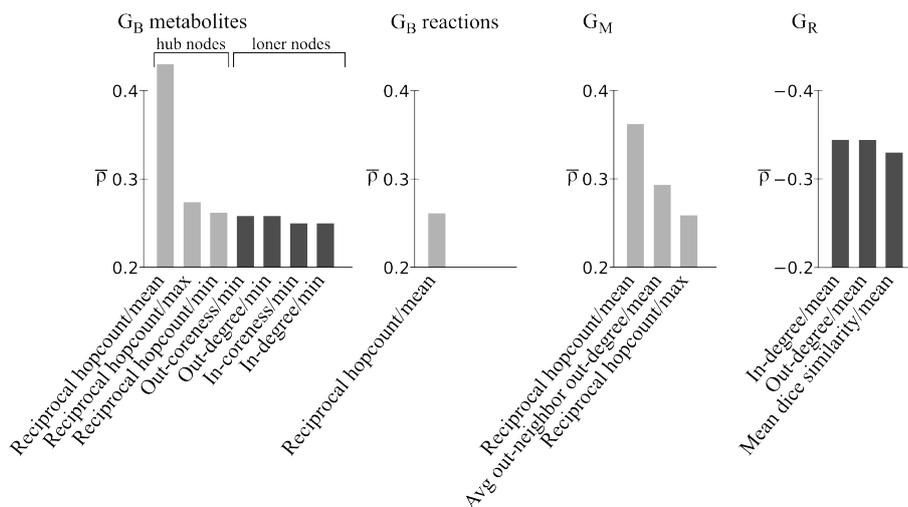
Fig. 4: $\bar{\rho}$ measures between growth and topological metrics for G_B, G_M and G_R. The postfixes "/min", "/mean" and "/max" indicate how node values were reduced to single values as discussed in Section 2.2. The mean dice similarity is computed by taking the mean of a vector (of dice similarity values) that is associated with a node (as discussed in Section 2.2).

G_M and G_R are shown in Figure 4 (recall that there are two sets of results for $G_B = G_B(\mathcal{M} \cup \mathcal{R}, \mathcal{L})$: one for the metabolite nodes $\mathcal{M}$ and one for the reaction nodes $\mathcal{R}$). First we discuss these results from a purely topological perspective and then we interpret the biological aspects.

3.1 Topology is weakly correlated with function

It is interesting that most metrics do not correlate well with growth: only a few $\bar{\rho}$ values are above the 0.25 threshold. An obvious first explanation for this lack of correlation is that it is possible to remove a reaction without affecting growth (since the reaction may be part of a bypass that is not used when the cell is functioning normally). However, at a deeper level, the low correlations may be explained by the indirect relationship between the flux balance analysis framework (which measures function) and the network (on which topological metrics are measured). In flux balance analysis, growth is the objective function of a linear program in terms of metabolic fluxes, whilst the topologies of the metabolic networks are only functions of the stoichiometric matrix. While the objective function may be changed (perhaps to study a scenario other than growth maximization) the topology remains unchanged. Thus, correlations between the objective function and topological metrics depend to some extent on the objective function.

3.2 The metabolite-reaction network G_B is the best representation

Metabolite-reaction network G_B As mentioned in Section 2.1, there are more reactions than metabolites in metabolic systems. A reaction ties together a small number of metabolites, while there are metabolites that are involved in many reactions. In other words, metabolites tie the network together at a high level and are responsible for the global connectivity of G_B.

The $\bar{\rho}$ correlations for G_B are shown in the first two bar charts of Figure 4. The results for the metabolite nodes fall into two groups: those caused by hub metabolite nodes (light gray bars) and those caused by "loner" metabolite nodes (dark gray bars), so-called because they are consumed and/or produced by a single reaction (and therefore they have out- and/or in-degrees of one).

For both the metabolite nodes and the reaction nodes, hub nodes provide shortcuts through which shortest paths are routed. Removal of a reaction node that interacts with a hub node may therefore remove a shortcut through which some shortest paths are routed. Thus the mean reciprocal hopcount is decreased (and the mean hopcount is increased). This explains the hopcount correlations for the metabolite and reaction bar charts of G_B.

The in- and out-degree correlations of the metabolites are the result of "loner" metabolites. Removal of a reaction that is the only in- or out-neighbor of a metabolite reduces the minimum in- or out-degree in the network from one to zero. More than half of the metabolites in G_B are "loner" metabolites, so that many reaction removals cause such decreases. The out- or in-coreness of a node is decreased when its in- or out-degree is decreased, which explains the coreness correlations.

The degree and coreness correlations do not appear for the reaction nodes, because a reaction removal cannot change the in- or out-degrees of any other reactions (since those reactions are only connected to metabolite nodes).

Metabolite network G_M The third bar chart in Figure 4 shows the results for G_M. The hub nodes in G_M have high out-degrees and are out-neighbors of many nodes. Therefore, they are implicated in the average out-degree correlation of out-neighbors. Many reaction removals decrease out-degrees of hub nodes in G_M or remove incoming links to them. In the former case, the average out-degrees of out-neighbors of all nodes that have such a hub nodes as out-neighbors, are decreased. In the latter case, a node loses a hub node as an out-neighbor, which lowers the average of the out-degrees of its out-neighbors.

Reaction network G_R The out- and in-degrees of G_R are negatively correlated with growth. This is due to a skewed degree distribution: only 104 out of 764 nodes in G_R have in- and out-degrees above the mean in- and out-degrees of G_R, while 393 nodes have in- and out-degrees below the means. Removal of a random node is therefore likely to increase the mean in- and/or out-degree of G_R. The negatively correlated mean dice coefficient is a result of the negative correlations of the in- and out-degrees: the average dice coefficient of a node is increased if it loses a neighbor with which it shares fewer neighbors than it does, on average, with its other neighbors.

G_R is much denser than G_B and G_M because the high degrees of metabolites in G_B cause many reactions in its projection G_R to be connected. As a result, much of the connectivity information is lost. This effect is not as pronounced in G_M, since the relatively low degrees of the reaction nodes in G_B do not add as many links to metabolites in the metabolite projection G_M.

G_B is the preferred network and metabolite nodes are the preferred nodes G_B is the preferred representation of the metabolic system, since it retains most of the original information and has the highest correlations with growth. Henceforth we focus only on G_B. Some applications, such as clustering, would have to be applied to a one-mode network. For these cases, the results suggest that the metabolite network is to be preferred.

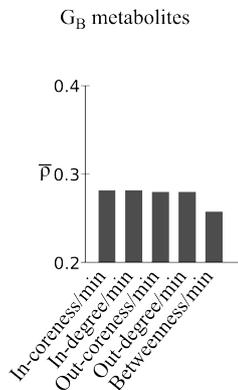### 3.3 Correlations point to currency metabolites

Fig. 5: $\bar{\rho}$ measures between growth and topological metrics for G_B with currency metabolites removed.

Many hub metabolite nodes in G_B correspond to so-called *currency metabolites*. We know from biology that currency metabolites play a crucial role in metabolism: they are energy carriers or cofactors that are used by many reactions. Holme *et al.* [5] found the currency metabolites of *S. cerevisiae* to be Hydrogen, Water, ATP, ADP, AMP, NAD, NADH, NADP, NADPH, Coenzyme A, CO_2, Oxygen, Phosphate, Diphosphate and Ammonium (for this set they used the undirected version of G_M with information taken from the BiGG database).

To validate the role of these metabolites, we repeated our experiments for G_B with currency metabolites removed (the metabolites were not removed from the flux balance linear program, as this would lead to incorrect chemical equations and it would change the computed growth). Again, we only considered metrics for which $\bar{\rho} \geq 0.25$. The results are shown in Figure 5.

As expected, currency metabolite removal causes the correlations due to hub nodes (involving the mean reciprocal hopcount values) to vanish, leaving the correlations caused by "loner" metabolites (the degree and coreness correlations). The betweenness correlation is new and merits an explanation. The absence of hub nodes (currency metabolites) forces shortest paths to be spread out through more of the network. A reaction removal is therefore more likely to "break" a shortest path and to reduce the betweenness of metabolite node on that path (since fewer shortest paths are then routed through the node).

4 Conclusions

The goal of this study was to determine whether topology and robustness of biological systems are related. Starting with robustness defined as the ability of a system to maintain its function under perturbations, we focused on metabolic systems and defined (a) function as the growth achievable by the metabolic system and (b) perturbations as reaction removals from the system.

We showed that a metabolic system is representable as a hypernetwork but not directly as a normal network. In order to make use of the tools available from network theory, we considered three network representations of a hypernetwork: the metabolite-reaction network G_B, the metabolite network G_M and the reaction network G_R.

Then we generated a number of reaction removal sequences or trials, each of which resulted in the cessation of growth of our metabolic system. Each step in a trial provided a snapshot of the metabolic system from which growth and topological metrics of G_B, G_M and G_R could be computed. This allowed us to calculate a measure of correlation between growth and each of the metrics.

Our results showed that most topological metrics do not correlate with growth. We classified the correlated metrics into two groups: those caused by hub metabolite nodes and those caused by "loner" metabolite nodes. This points to the importance of (a) global connectivity (by hub nodes that tie the network together by connecting to many reactions) and (b) local connectivity (by metabolites that are produced and/or consumed by few reactions). By identifying a group of hub metabolites as currency metabolites, we proceeded to show that the removal of these nodes indeed knocks out the correlations that we associated with hub nodes, verifying their topological role and their biological importance. We found that the correlations were the strongest in the metabolite-reaction network and in particular with the metabolite nodes of this network. We also found correlations related to global connectivity in the metabolite network. This suggests that metabolites are important for studying the structure of a metabolic system and that the bipartite metabolite-reaction representation is to be preferred over the one-mode reduced metabolite representation.

In this work, we have studied the relation between topology and growth. Using our framework, one can investigate whether other functions of the metabolic network are related to topology or whether topology has a larger influence in other biological networks (e.g., gene regulation or protein interaction networks).

Another future research direction would be to consider how the topologies of subnetworks of interest (such as, for example, the subnetwork containing the citric acid cycle) change in relation to changes in growth.

References

1. A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.
2. D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucl. Acids Res.*, 31(9):2443–2450, May 2003.
3. N. Guelzim, S. Bottani, P. Bourgine, and F. Kepes. Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet*, 31(1):60–63, May 2002.
4. Simon Hardy and Pierre N. Robillard. Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. *Journal of bioinformatics and computational biology*, 2(4):595–613, December 2004.
5. P. Holme and M. Huss. Substance graphs are optimal simple-graph representations of metabolism. *arXiv:0806.2763v2 [q-bio.MN]*, Dec 2008.
6. P. J. Ingram, M. P. Stumpf, and J. Stark. Network motifs: structure does not determine function. *BMC genomics*, 7(1):108+, 2006.
7. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000.
8. Wan Kyu K. Kim and Edward M. Marcotte. Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence. *PLoS computational biology*, 4(11):e1000232+, November 2008.
9. H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, November 2002.
10. H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, November 2004.
11. Gipsi Lima-Mendez and Jacques Helden. The powerful law of the power law and other myths in network biology. *Mol. BioSyst.*, 5(12):1482–1493, December 2009.
12. S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, May 2002.
13. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, October 2002.
14. T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, April 1989.
15. M. E. J. Newman. Assortative mixing in networks. *arXiv:cond-mat/0205405v1 [cond-mat.dis-nn]*, May 2002.
16. M. E. J. Newman. The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 2007.
17. Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, March 2010.
18. Organisms - systems biology research group, November 2009. http://systemsbiology.ucsd.edu.

19. R. J. Prill, P. A. Iglesias, and A. Levchenko. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol*, 3(11):e343+, October 2005.
20. J. Stelling, U. Sauer, Z. Szallasi, F.J. Doyle 3rd, and J. Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, September 2004.