

Stability cannot be derived from local structure in biochemical networks

Domenico Bellomo
Inf. & Comm. Theory Group/
Bioprocess Tech. Group, Delft
University of Technology, The
Netherlands; and Kluyver
Centre for Genomics of
Industrial Fermentation
d.bellomo@tudelft.nl

Peter van Nes
Inf. & Comm. Theory Group,
Delft University of Technology,
The Netherlands;
pvannes@gmail.com

Marcel J.T. Reinders
Inf. & Comm. Theory Group,
Delft University of Technology,
The Netherlands; and Kluyver
Centre for Genomics of
Industrial Fermentation
m.j.t.reinders@tudelft.nl

Dick de Ridder
Inf. & Comm. Theory Group,
Delft University of Technology,
The Netherlands; and Kluyver
Centre for Genomics of
Industrial Fermentation
d.deridder@tudelft.nl

ABSTRACT

In recent literature, the relationship between structure and dynamics in biochemical networks has been intensively investigated. In fact, the scarcity of information about such networks has led to attempts to predict some of their dynamic features based exclusively on more easily available structural information. A recent finding relating structure to dynamics is that *network motifs* (a structural feature) that are *structurally stable* (a dynamic feature) are enriched in some biochemical networks [21].

In this work, we systematically investigate the method in [21] and the assumptions it relies on. Our findings suggest that the conclusions drawn on the considered biological networks (over-representation of structurally stable motifs) cannot be generalized, as they critically depend on a user-defined choice of null model in the motif enrichment analysis. We have further applied the method in [21] to metabolic networks, which provide an excellent test-bed, as a relatively large amount of information is available about the type, strength and activity of metabolic interactions. For metabolic networks we arrive at the same conclusion: stability cannot be derived from local structure.

Keywords

Biochemical networks, structure, dynamics, network motifs, structural stability.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Bionetics '08, November 25-28, 2008, Hyogo, Japan
Copyright 2008 ICST 978-963-9799-35-6.

1. INTRODUCTION

Experimental techniques in molecular biology have witnessed impressive developments over the last few years, and the genomes of many micro-organisms have been extensively annotated. Despite this progress, the available experimental data (e.g. data generated in knock-out studies or in steady-state perturbation experiments) provide mainly structural information about the biochemical networks in the cell (e.g. the transcriptional, signalling and metabolic networks). The limited availability of time-series measurements of good quality makes it difficult to reliably estimate the kinetic parameters characterizing the dynamic behaviour of such networks *in vivo*.

However, a better understanding of the complex dynamics of living cells is a pre-requisite for boosting the emerging fields of *metabolic engineering* and *synthetic biology* [11]. Therefore, different methods have been investigated to find out to what extent the dynamic behaviour of different biochemical networks is determined/constrained by their structure, and to develop methods elucidating such possible relationships between structure and dynamics. The starting point of all these methods is to associate a (directed) graph to the considered biochemical network, in which the nodes represent its constituent building blocks (e.g. genes, proteins, metabolites, etc.), and the edges represent the interactions between them. One of the earliest attempts to relate structure to dynamics was *chemical reaction network theory*: the existence of multiple steady states is predicted exclusively based on the analysis of topological properties of the network graph [5, 6]. More recent studies have investigated the interplay between global topological properties of the network graph (i.e. node degree distribution) and robustness against random or targeted perturbations (see for example [2, 4]). For some regulatory networks, for which the sign of the edges in the associated graph is also known (i.e. activation or inhibition), the over-abundance of sign-consistent feedback and feed-forward loops (compared

to negative loops) has been related to *monotonicity* and to stability [19, 22]. Moreover, a number of quite successful methods have been specifically developed for the analysis and re-engineering of metabolic networks, such as flux balance analysis, elementary fluxes modes, and structural kinetic modeling [9, 16, 23].

Another very active research area is the investigation of the relationship between local topological properties and the (global) dynamic behaviour of the network. A network graph can be decomposed into small subgraphs (*network motifs*), and the frequencies of occurrence of these motifs (typically three- or four-node motifs are considered) can be used to characterize the local connectivity patterns of the network. It has been shown in [3] that certain motifs (e.g. the feed-forward loop) are enriched in many different biochemical networks; that is, they are over-represented compared to random networks of the same size. However, it is not straightforward to determine what mechanisms govern the relative abundance of the different motifs, or what dynamic property is carried by motifs that may have been positively selected for by natural selection.

A recent paper has investigated the relationship between the stability of a motif, referred to as *structural stability*, and its frequency of occurrence in a number of transcriptional regulatory networks and other biological networks [21]. The method entails two main steps for each motif: (i) calculation of the over- or under-representation of the motif in the considered biological network, compared to a reference random network (the so-called *null model*); (ii) calculation of the *structural stability score* (SSS) of the motif, the fraction of parameter settings for which the motif is stable (see Section 2.4 for more details). The key finding in [21] is that there is a positive correlation between the SSS and the frequency of occurrence of a motif: structurally stable motifs are enriched in biochemical networks.

The idea to relate motifs to dynamic properties of the overall network, though very attractive, has recently met some serious criticism. In [13], it is argued that some specific motifs (e.g. the feed-forward loop) can exhibit a broad range of dynamic behaviour, making it very hard to clearly establish a link between structure and dynamics. Furthermore, in [7] it has been pointed out that a network made up of stable motifs is not necessarily stable itself: stability is a global property of an interconnected system and cannot be directly inferred from the properties of its constituent sub-systems. These controversies call for a more systematic investigation of the results presented in [21].

The goal of this paper is systematically investigate the findings in [21]. Our analysis clearly shows that the choice of null model in the motif enrichment calculations critically affects the results in [21]. Since the chosen null model, Erdős-Rényi (ER) random networks, is not generally considered appropriate for the considered biochemical networks (see Section 2.3 for more details), the claim that structurally stable motifs are over-represented cannot be confirmed.

For the biochemical networks studied in [21], only binary interaction information is available (i.e. we know whether an edge is present or not in the network graph). For metabolic networks however, more information is available about the type, strength and activity of many of the interactions. These additional data provide an excellent test-bed for a systematic evaluation of the method in [21], since we can build a network model that is more biologically meaningful. In

this paper, we have considered the metabolic network in [8]. We have determined interaction types based on the stoichiometric matrix of the metabolic network (see Section 2.5.1) and retrieved information about the strength [1] and activity [17] of interactions (see Sections 2.5.2 and 2.5.3). However, for metabolic networks, the conclusion does not change: structurally stable motifs are not over-represented, and hence stability cannot be derived from local structure.

2. METHODS

2.1 Data-sets

2.1.1 Biological networks in Prill's work

The data in [21] describe the transcriptional regulatory networks of *Escherichia coli* and *Saccharomyces cerevisiae*, the developmental transcriptional network of *Drosophila melanogaster*, the signal transduction knowledge environment (STKE) network, and a neural connection map of *Caenorabditis Elegans*.

2.1.2 Metabolic network

We have considered the *S. Cerevisiae* metabolic model presented in [8]. This model contains a quite complete list of chemical reactions (1223 metabolites, 929 irreversible reactions and 477 reversible reactions) including compartmentalization information. The great majority of reactions are associated to annotated genes. The authors kindly provided an updated version of the data in [8], which is still unpublished (*S. Cerevisiae IMM904*).

In order to construct the parameter distribution necessary for the determination of interaction strengths, we used the enzyme database BRENDA [1]. BRENDA is indexed by *enzyme commission* (EC) numbers corresponding to enzyme-catalyzed reactions, thus allowing us to retrieve the kinetic parameters of each reaction.

Information on the activity of enzyme-catalyzed reactions (not all reactions are active at any given time) has been retrieved from a micro-array compendium [17], a large collection of micro-array experiments on *S. Cerevisiae*, consisting of 9335 gene expression levels collected in 165 experiments performed under a number of different growth conditions.

2.2 Network motifs

A compact representation of a biochemical network can be obtained by associating a (directed) graph to the network: the nodes represent its constituent building blocks (genes, proteins, metabolites, etc.), whereas the edges represent the interactions between them. A *network motif* is a very small (directed) subgraph (see Fig. 1). This paper only deals with motifs consisting of three nodes with two to six directed edges. In three-node motifs, the length of the shortest path between any of the three nodes is at most two. A node can be a member of multiple motifs, and the same holds for interactions. Although sets of three nodes might have very different interaction strengths, motif counts only look at structure, i.e. the interaction between the nodes is binarized (either there is an interaction or not). Therefore, there are only 13 possible network motifs, which are shown in Fig. 1.

Each motif is uniquely identified by its *adjacency matrix*. It is a 3×3 matrix, and a nonzero entry at position (i, j) means that component i is influenced by component j . An adjacency matrix contains binarized interaction information

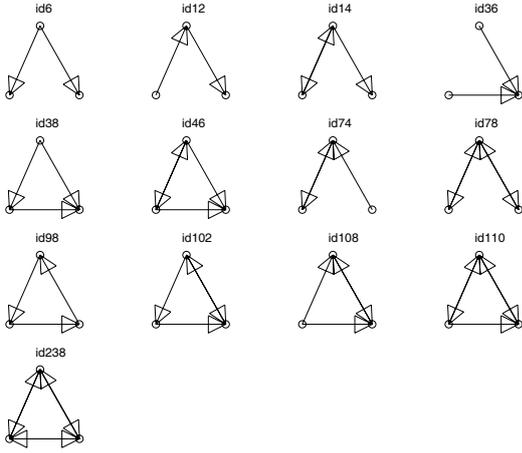


Figure 1: All possible network motifs consisting of 3 nodes. The motif ID is the decimal form of the binary integer obtained by concatenating the rows of the adjacency matrix.

at these positions; the Jacobian corresponding to the kinetic model associated to the motif contains the actual interaction strengths.

2.3 Motif enrichment analysis

Motif enrichment is determined by generating a large number of randomized versions of the original biochemical network, and calculating a z -score for each motif m as

$$Z_m = \frac{N_{real_m} - \text{mean}(N_{rand_m})}{\text{std}(N_{rand_m})}, \quad (1)$$

where N_{real_m} is the number of occurrences of motif m in the real network and N_{rand_m} the set of occurrences of motif m in the random networks. A high z -score for motif m indicates that the probability of finding motif m as often as in the biochemical network by chance is low. Conversely, a low z -score indicates that the probability of finding motif m as little as in the biochemical network is low. Z -scores are normalized to unit length (*normalized z-scores* or NZSes), by using

$$NZS_m = \frac{Z_m}{\sqrt{\sum_{i=1}^M Z_i^2}}, \quad (2)$$

in which M is the number of motifs. These NZSes can be compared across different networks.

The calculation of NZSes requires the generation of random networks and counting motif frequencies. Both of these tasks can be equivalently performed with the software tools *Mfinder* and *FANMOD* [15, 24]. However, *FANMOD* has the advantage of being able to handle colored edges (see below).

In this work, we generate the random networks used as null model according to the *Erdős-Rényi* (ER) method as originally done in [21], and according to the *switching method* [20], which is the default method used by both *Mfinder* and *FANMOD*. The ER method puts p nodes on a canvas and subsequently adds q directed edges, uniformly picked from the set of all possible $p \cdot (p - 1)$ edges. Networks generated by this method have node degree distributions that follow a Poisson distribution, whereas it is commonly observed that biological networks are scale-free, i.e. their

node degree distribution follows a power law [4]. Random networks generated using this method are therefore deemed less suitable for representing biological networks. Another problem is that the probability of generating bidirectional edges is low: bidirectional edges are rare in ER networks. Network motifs with such edges are hardly found in ER random networks, resulting in low (zero) variance and thus in an infinite NZS, which is obviously an undesired effect.

The switching method on the other hand operates as follows [20]. The original network is used as basis and a pair of edges ($A \rightarrow B, C \rightarrow D$) is repeatedly randomly selected and switched to obtain ($A \rightarrow D, C \rightarrow B$). The exchange is only performed if it does not introduce an edge that already exists or a self-edge, i.e. an edge from a metabolite to itself. Furthermore, unidirectional edges are only exchanged with other unidirectional edges and bidirectional edges only with bidirectional ones. Edge “colors” (corresponding to some discrete property) can also be taken into account, i.e. edges are only switched when they have the same color. The process is repeated a sufficient number of times for the random network to show good mixing (for details, see [15]). The switching method preserves the number of incoming, outgoing and bidirectional edges of each node of the real network, and thus the exact degree distribution, making it a more reliable enrichment analysis in biological networks.

For a global structural analysis of metabolic networks, *currency metabolites* (e.g. ATP and ADP) are generally removed. In fact, currency metabolites do not provide (specific) functional information about the metabolites they are connected to [18]. However, for the local topological analysis, we found out that their removal produces results that are qualitatively the same. All results reported here were obtained on a complete metabolic network.

2.4 Structural stability

In [21], the *structural stability score* (SSS) is introduced as a measure for the probability that the dynamical systems that can be associated to a given motif are locally (i.e. around a steady-state) asymptotically stable with no oscillation. This score is determined by first generating a large number of possible Jacobian matrices for a given motif, and subsequently calculating the eigenvalues of each of these Jacobians. The SSS is the fraction of the Jacobians of which all eigenvalues have a negative real part and zero imaginary part. As it is computationally intractable to instantiate every possible Jacobian, we sample from the space of possible Jacobians, which is done by instantiating 10,000 Jacobians in which each nonzero entry is sampled from a given distribution. In [21], Jacobians are constructed by assigning a value sampled from a $U(-1, 1)$ distribution (uniform over range $(-1, 1)$) to all non-diagonal, nonzero entries and a value sampled from a $U(-1, 0)$ distribution to all diagonal entries of the adjacency matrix of a motif. Note that the range of the SSS is $[0, 1]$, with a value of 1 indicating that any dynamic system associated to the the motif is stable (i.e. the interaction signs and strengths do not influence the stability). On the other hand, a low value indicates that only a small fraction of all possible parameters of the Jacobian can guarantee stability.

2.5 Metabolic interactions

In this section, we describe how to determine the type, strength and activity of metabolic interactions based on the

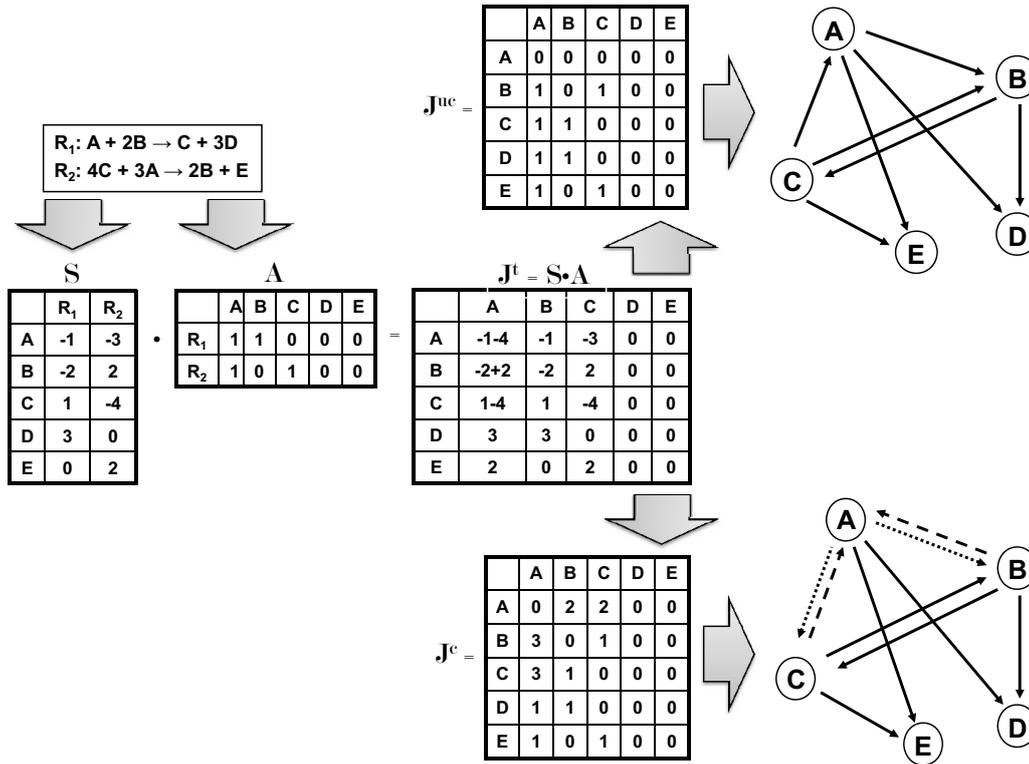


Figure 2: The artificial network consisting of the two reactions shown at the top of this figure is used to illustrate how matrices \mathbf{S} and \mathbf{A} are constructed. Also, an intermediate matrix \mathbf{J}^t (not described in the main text) is shown, which is the result of the multiplication $\mathbf{S} \cdot \mathbf{A}$. Matrix \mathbf{J}^{uc} is constructed by setting all diagonal entries to zero and putting ones at the locations where there are only positive values in \mathbf{J}^t . Matrix \mathbf{J}^c is constructed in a similar manner for the diagonal and positive entries. However, the entries of \mathbf{J}^t consisting of only negative values are set to 2, whereas the entries consisting of both positive and negative values are set to 3. The resulting uncolored and colored networks are shown at the bottom of the figure.

information available in databases and in the literature (see Section 2.1).

2.5.1 Interaction type

The behaviour of a dynamic system (a biochemical network in our case) around a steady state can be described using a linear approximation. The local linear model is defined by the Jacobian, an $n \times n$ matrix (with n the number of nodes in the network). Each element $J_{i,j}$ represents the influence of metabolite j on metabolite i . The structure of the Jacobian (i.e. its nonzero entries) defines the topology of the interactions in the biochemical network. In the case of a metabolic network, the Jacobian \mathbf{J} takes the form

$$\mathbf{J} = \mathbf{S} \cdot \left. \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right|_{\mathbf{x}^0} \quad (3)$$

where \mathbf{S} is a *stoichiometric matrix* of size $m \times r$, with m and r the number of metabolites and reactions respectively, and $\left. \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right|_{\mathbf{x}^0}$ is a matrix of size $r \times m$ of partial derivatives of the vector of reaction rates \mathbf{v} with respect to the vector of metabolite concentrations \mathbf{x} in steady state \mathbf{x}^0 . A stoichiometric matrix contains the reaction coefficients of every reaction of a reaction network. Each reaction is represented by a column in the matrix in which substrates and products of that reaction have some negative or positive integer value

respectively.

The metabolic network (Section 2.1.2) results in a stoichiometric matrix \mathbf{S} with $m = 1223$ and $r = 1883$. Note that each of the 477 reversible reactions is represented as two unidirectional reactions.

From the stoichiometric information in \mathbf{S} , a matrix \mathbf{A} can be constructed which has the same dimensions as $\left. \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right|_{\mathbf{x}^0}$ and has an entry of 1 at position (i,j) when the partial derivative of reaction j w.r.t. the concentration of metabolite i is nonzero and 0 otherwise. When we substitute $\left. \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right|_{\mathbf{x}^0}$ in (3) by \mathbf{A} , we obtain a Jacobian which can be used to construct matrices \mathbf{J}^{uc} and \mathbf{J}^c representing an uncolored and a colored model of our network respectively (later, edge color will be used to differentiate different type of interactions). \mathbf{J}^{uc} is constructed such that $J_{i,j}^{uc}$ is 1 when metabolite j has a positive effect on metabolite i , i.e. metabolite j is a substrate in at least one reaction where metabolite i is a product, and zero otherwise. \mathbf{J}^c on the other hand, has three interaction types: positive, negative and combined, which are (arbitrarily) represented by 1, 2 and 3 respectively. A positive interaction is defined similar as for \mathbf{J}^{uc} . A negative interaction at $J_{i,j}^c$ indicates that metabolites i and j are both substrates in the same reaction. A combined interaction is a combination of these two interaction types, which should be thought of as an interaction which can act in both a positive and a neg-

ative way. \mathbf{J}^{uc} contains the structural information without interaction types, or colored edges. This network will from now on be referred to as the *uncolored network*. Conversely, the network represented by \mathbf{J}^{c} will be referred to as the *colored network*. Fig. 2 shows how a small artificial network consisting of only two reactions would be transformed into both networks.

2.5.2 Interaction strength

In a metabolic network, entries in the Jacobian matrix represent interactions between metabolites. In our analysis, we assume that all reaction rates follow the Michaelis-Menten kinetic rate law:

$$v = \frac{x}{x + K_m} \cdot k_{\text{cat}} \cdot [E], \quad (4)$$

where K_m is the Michaelis-Menten constant, $[E]$ is enzyme concentration and k_{cat} represents the maximum number of moles of substrate that the enzyme can convert to product per unit time [1].

Differentiating (4) gives:

$$\frac{dv}{dx} = \frac{k_{\text{cat}} \cdot [E] \cdot K_m}{(x + K_m)^2}. \quad (5)$$

We are interested in the values of $\frac{dv}{dx}$ in order to calculate the matrix $\frac{\partial \mathbf{v}}{\partial \mathbf{x}}|_{\mathbf{x}_0}$. Subsequently, these can be sampled from to generate matrix \mathbf{A} (Section 2.5.1).

Values for K_m and k_{cat} are collected by parsing BRENDA, whereas values for $[E]$ are collected from [10], which contains a list of concentrations for a number of proteins of *S. Cerevisiae*. Values for K_m , k_{cat} and $[E]$ were selected as triplets only when (i) the three values belong to the same protein; (ii) K_m and k_{cat} correspond to the same substrate and (iii) K_m and k_{cat} correspond to the same conditions. Each triplet of values for K_m , k_{cat} and $[E]$ was entered 100 times in (5), each time using a value for x uniformly picked from the range $[1 \times 10^{-5}, 2 \times 10^{-2}]$ moles per liter [12]. The result is a distribution for $\frac{dv}{dx}$, of which each sample represents an entry in the $\frac{\partial \mathbf{v}}{\partial \mathbf{x}}|_{\mathbf{x}_0}$ matrix.

Next, 1000 matrices \mathbf{A}^l , $l \in [1 \dots 1000]$ are generated, by assigning all nonzero entries in \mathbf{A} a value randomly sampled from the $\frac{dv}{dx}$ distribution. Using these \mathbf{A}^l s, 1000 matrices \mathbf{J}^l are generated:

$$\mathbf{J}^l = \mathbf{S} \cdot \mathbf{A}^l, \quad l \in [1 \dots 1000]. \quad (6)$$

Finally, all nonzero entries in the 1000 Jacobians thus generated are distributed over vectors \mathbf{j}_{pos} , \mathbf{j}_{neg} , \mathbf{j}_{com} and \mathbf{j}_{diag} , representing positive, negative, combined or diagonal entries of a Jacobian matrix respectively, such that $\mathbf{j}_{\text{pos}} = \{J_{i,j}^l | J_{i,j}^c = 1\}$, $\mathbf{j}_{\text{neg}} = \{J_{i,j}^l | J_{i,j}^c = 2\}$, $\mathbf{j}_{\text{com}} = \{J_{i,j}^l | J_{i,j}^c = 3\}$ and $\mathbf{j}_{\text{diag}} = \{J_{i,j}^l | i = j\}$. These four vectors together will from now on be referred to as the *BRENDA distribution*.

In Section 2.4, it was discussed how the method described in [21] determines the SSS of a motif. It instantiates Jacobians by sampling values from uniform distributions and entering them in the adjacency matrix. We instead sample from the BRENDA distribution, which seems biologically more relevant. This is done by assigning diagonal values, non-diagonal values of 1, non-diagonal values of 2, and non-diagonal values of 3 in the adjacency matrix of a motif a value sampled from \mathbf{j}_{diag} , \mathbf{j}_{pos} , \mathbf{j}_{neg} , or \mathbf{j}_{com} respectively. The remaining part of the procedure is the same as described

in Section 2.4, yielding a new structural stability score based on the BRENDA distribution. This new SSS and the SSS of [21] will from now on be distinguished as the *B-SSS* (for BRENDA) and the *P-SSS* (for Prill), respectively.

2.5.3 Interaction activity

The reactions described in the stoichiometric matrix represent all the chemical reactions taking place in the cell. However, not all reactions are simultaneously active. In fact, for a reaction to be active the corresponding enzyme (catalyzing the reaction) has to be present, and the presence of the enzyme is in turn regulated by genes.

The compendium in [17] provides the expression levels of all the annotated genes in *S. Cerevisiae*. We selected only the 902 genes known to regulate reactions in the metabolic network in [8]. For each experiment, we used expression values obtained by robust multi-chip analysis (RMA) values [14] and scaled to a range of $[0, 1]$, and removed reactions for which the corresponding gene had a value less than a threshold τ of 0.5. Other values for τ (0.05 – 0.6) produced comparable results in terms of motif counts (although the number of eliminated reactions varies by an order of magnitude).

3. RESULTS

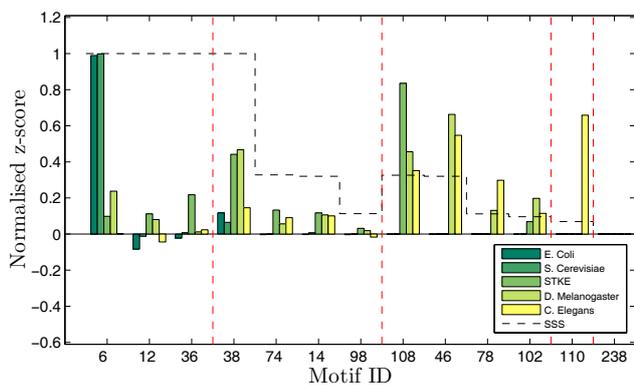
3.1 Baseline method

Fig. 3(a) shows the results for all the networks considered in [21]. The motifs are divided into *density classes*, groups of motifs having an equal number of edges (delimited by vertical dashed lines). The dashed function represents the SSS, the bars the NZS. In the case of 3-node motifs, the number of edges ranges from 2 to 6, yielding 5 density classes. In Fig. 3(a), a descending stairs-like behaviour can be observed (at least for some of the networks): within each density class, the highest scoring motifs appear on the left, the lower scoring motifs on the right (positive correlation between the SSS and the NZS of a network motif). However, it is important to point out that in Fig. 3(a) infinite NZS are set to zero (consistent with the results in [21]). An example is provided by motif 78 of the transcriptional regulatory network of the *S. Cerevisiae*, consisting of two bidirectional edges (Fig. 1). Since the Erdős-Renyi (ER) randomization method produces very few bidirectional edges (Section 2.3), it is likely that the motif is not generated at all, resulting in an infinite NZS (as in [21]). We can conclude that if we take into account infinite NZSes, the correlation between the SSS and the NZS is not much evident anymore.

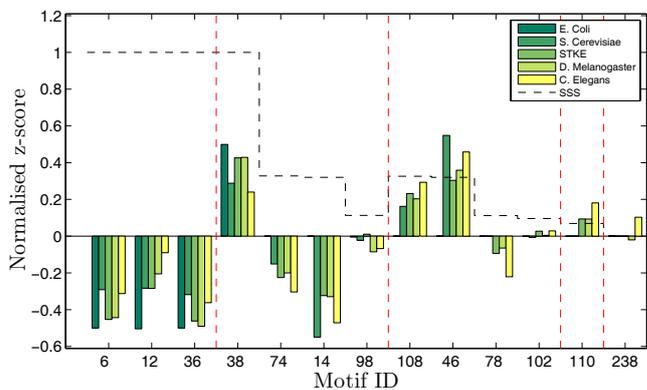
3.2 Null model

The calculation of motif enrichment requires the definition of a proper null model (see Section 2.3). Different choices are possible for the generation of the random networks, and it has been recognized that random networks generated by the *switching method* should be preferred over ER random models, since they better preserve the topology of the original network (see Section 1). Therefore, we have re-calculated motif enrichment for the transcriptional regulatory network in [21], using the switching method (see Fig. 3(b)).

It can be clearly seen that the choice of randomization method has considerable influence on the results in [21]. In

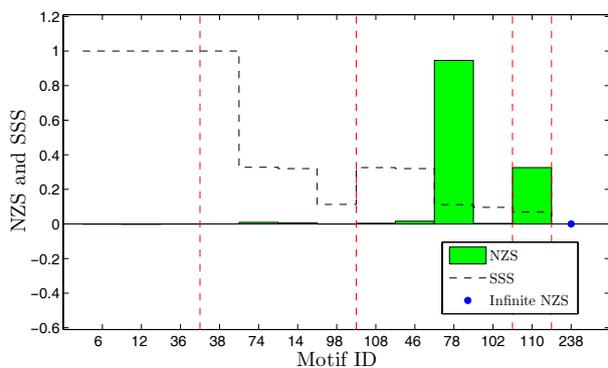


(a) Baseline method.

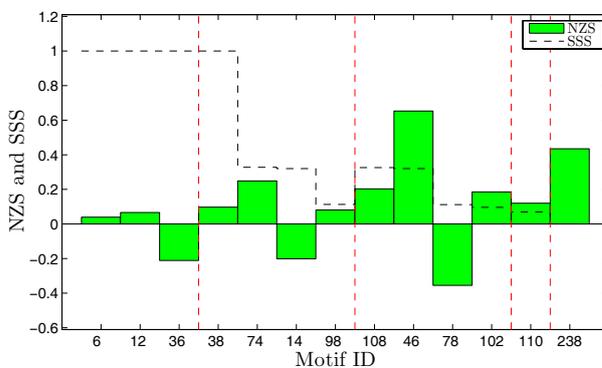


(b) Switching method.

Figure 3: Network in Prill: structural stability score (SSS) and normalized z -scores (NZS) as a function of the motifs. Motif enrichment calculation is based on random networks generated with the baseline method and the switching method.



(a) Baseline method.



(b) Switching method.

Figure 4: Metabolic network: structural stability score (SSS) and normalized z -scores (NZS) as a function of the motifs. Motif enrichment calculation is based on random networks generated with the baseline method and the switching method.

Fig. 3(b), motif 38 still has a high NZS, but for the other motifs, the trend seen in the original results is lost. Note in particular motif 6, whose NZS has gone from around 1 to -0.3 , indicating that it has gone from occurring much more frequently in the real networks than in the random ones, to less.

In summary, by replacing the ER randomization method by the switching method, we obtained results which are in contrast with the ones reported in [21]: over-represented motifs are now under-represented.

3.3 Metabolic interactions

In this section, we first apply the same analysis as in Section 3.1 and Section 3.2 to the metabolic network of *S. Cerevisiae* [8]. Then, we exploit the additional information, about the interactions between metabolites, available for the considered metabolic network: (i) interaction type, (ii) strength and (iii) activity.

3.3.1 Baseline method and null model

The baseline method applied to the metabolic network produces results in contrast to those in [21] (see Fig. 4(a)): structurally stable motifs are clearly not enriched. Only two motifs have a high NZS (78 and 110, which both have an infinite NZS in Fig. 3(a)). The motif with highest NZS in Fig. 3(a), 6, is not over-represented at all in our network. For the metabolic network, motif 78 was generated only a few times in the random networks, resulting in a low standard deviation in (1), and thus a high NZS. Motif 238 (indicated by a dot), consisting of even more bidirectional edges, is never generated and thus has infinite NZS (the same holds in Fig. 3(a)). Finally, note that motif 78 does not have the highest SSS in its density class.

If we now apply the switching method, the results change significantly (see Fig. 4(b)). The top scoring motif in Fig. 4(a), motif 78, now has the lowest NZS, caused by a high frequency of motif 78 in the random networks. Motif 238 has a quite high NZS, but no longer infinite as in Fig. 4(b). These observations indicate that the switching method has

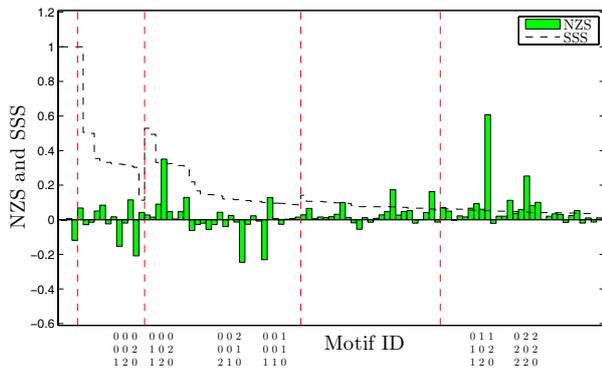


Figure 5: Structural stability score and normalized z -scores. The motifs are defined by structure and interaction type (colored motifs).

generated more bidirectional edges than the ER method, as expected. The top scoring motif is now motif 46. This is the motif with the highest SSS of its density class. However, this happens only in this density class and we cannot infer from this finding that SSS and NZS are positively correlated.

3.3.2 Interaction type (colored edges)

The stoichiometric matrix entirely defines the mass-action kinetic part of the reaction rates, and from that we can associate a sign to the interactions between metabolites (neglecting the enzymatic regulation part). In the colored model \mathbf{J}^c , to each edge a label 1, 2 or 3 is assigned, indicating a positive, negative or combined interaction (Section 2.5.1). Results in Sections 3.1-3.2 are based on the 13 possible 3-node motifs of the uncolored model \mathbf{J}^{uc} (see Section 2.5.1). With the colored model, we have NZSes for 97 motifs, instead of only 13. As a consequence, the results in [21] can no longer be compared directly. However, a similar NZS vs. SSS plot can still be created (Fig. 5). A number of density classes show some correlation between NZS and SSS. For instance, inside classes 3 and 5 the over-represented motifs generally appear on the left, whereas the under-represented ones appear on the right. However, the other classes do not show this trend. Therefore, we can conclude that taking into account the interaction type, does not produce any definite correlation between SSS and NZS.

3.3.3 Interaction strength

In [21], the calculation of the SSS was based on parameters drawn from uniform distribution (P-SSS), since no further biological knowledge was available. For metabolic networks, a more realistic parameter distribution can be built based on the parameter values in BRENDA (the *BRENDA distribution*), which can be in turn used for the calculation of the SSS (B-SSS).

Comparison of the P-SSS and the B-SSS (results not shown) shows that (i) B-SSS never becomes 1 (there are no motifs that are always stable) in contrast to the P-SSS (motifs 6, 12, 36 and 38 are always stable); (ii) the difference between the B-SSS for the colored and the uncolored model is smaller than for the P-SSS (meaning that for the B-SSS the sign of an edge is less important for the stability); (iii) the P-SSS has a wider range than the B-SSS.

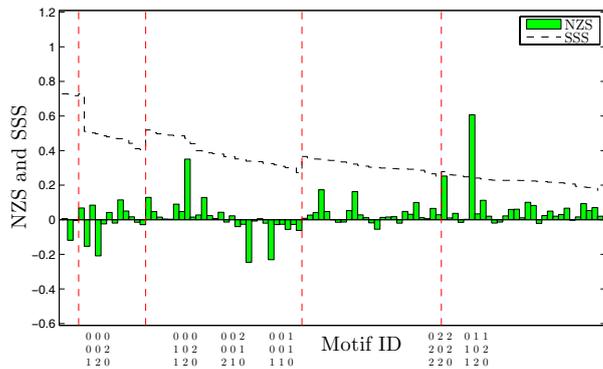


Figure 6: Brenda based structural stability score and normalized z -scores.

Despite these differences between P-SSS and B-SSS, we found that the SSS vs. NZS plot does not change significantly (see Figs. 5 and 6). We can conclude that the use of the more biological Brenda distribution for the calculation of the SSS does not produce the expected positive correlation between SSS and NZS (with the last density class in Fig. 6 the sole exception).

3.3.4 Interaction activity

If we take enzymatic regulation into account (Section 2.5.3), we obtain the results in Fig. 7. The main effect of removing reactions is that many of the peaks in Fig. 5 and Fig. 6 disappear or get flattened. Moreover, we can see that in the last density class the location of the highest peak is shifted to the right. However, it can be clearly seen that the new motif distribution still does not have any correlation with the structural stability score.

4. CONCLUSIONS

Many studies have attempted to relate local topological properties (i.e. network motifs) to (global) dynamic behaviour, showing that certain motifs are enriched in biochemical networks, and that over-represented motifs may confer some “desirable” dynamic feature on the network (e.g. robustness, stability, etc.). However, some of these findings are controversial and demand more investigation.

In this paper, we have investigated the results presented in [21], where it is claimed that structurally stable motifs are over-represented in 5 different biochemical networks. Our study clearly shows that the conclusions in [21] critically depend on the choice of null model in the motif enrichment analysis (and on the choice to set to zero infinite z -scores). Since the null model in [21] (Erdős-Rényi random networks) is not appropriate to describe the (scale-free) topology of the biochemical networks, the claim that structurally stable motifs are over-represented cannot be confirmed.

Moreover, we have tested the method in [21] against a biochemical network [8] for which information was available about (i) the type, (ii) strength and (iii) activity of the interactions. The main conclusion of this paper stays the same: structurally stable motifs are not over-represented, and hence stability cannot be derived from local structure (we reached similar conclusions with a 4-node motif analysis, and with the metabolic networks of *E. Coli* and *H.*

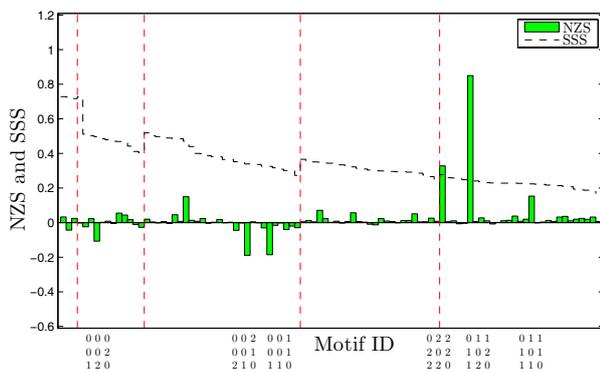


Figure 7: Structural stability score and normalized z -scores (micro-array data).

Sapiens).

Our findings do not rule out the possibility that motif enrichment in biochemical networks can be explained by dynamic properties other than structural stability. Recent results seem to suggest that monotonicity could be such a property [19, 22].

5. ACKNOWLEDGMENTS

The authors would like to thank Robert Prill for providing the data in [21]; Theo Knijnenburg for supplying the yeast micro-array compendium data (see [17]), and Rogier van Berlo for supplying part of the code used for the analysis.

6. REFERENCES

- [1] <http://www.brenda.uni-koeln.de>, 1-6-2007.
- [2] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.
- [3] U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8:450–461, 2007.
- [4] A. L. Barabási and Z. N. Oltvai. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.
- [5] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks. Part I: the injectivity property. *SIAM Journal of Applied Mathematics*, 65(5):1526–1546, 2005.
- [6] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks. Part II: the species-reaction graph. *SIAM Journal of Applied Mathematics*, 66(4):1321–1338, 2006.
- [7] J. Doyle and M. Csete. Motifs, control, and stability. *PLoS Biology*, 3(11):e392, 2005.
- [8] N. C. Duarte, M. J. Herrgard, and B. O. Palsson. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, 14:1298–1309, 2004.
- [9] J. S. Edwards, M. Covert, and B. O. Palsson. Metabolic modelling of microbes: The flux-balance approach. *Environmental Microbiology*, 4:133–140, 2002.
- [10] S. Ghaemmaghami, W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O’Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, 2003.
- [11] M. Heinemann and S. Panke. Synthetic biology – putting engineering into biology. *Bioinformatics*, 22(22):2790, 2006.
- [12] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, 92(5):1792, 2007.
- [13] P. J. Ingram, M. P. H. Stumpf, and J. Stark. Network motifs: structure does not determine function. *BMC Genomics*, 7(1):108, 2006.
- [14] R. A. Irizarry, B. Hobbs, F. Collin, Y. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264, 2003.
- [15] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11):1746–58, 2004.
- [16] S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends in Biotechnology*, 21(2):64–69, 2003.
- [17] T. A. Knijnenburg, L. F. A. Wessels, and M. J. T. Reinders. Combinatorial influence of environmental parameters on transcription factor activity. *Bioinformatics*, 24(13):i172–181, 2008.
- [18] H. Ma and A. P. Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2):270–277, 2003.
- [19] A. Maayan, R. Iyengar, and E. D. Sontag. Intracellular regulatory networks are close to monotone systems. *NET Systems Biology*, 2:103–112, 2008.
- [20] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. Uniform generation of random graphs with arbitrary degree sequences. *Arxiv preprint cond-mat/0312028*, 2003.
- [21] R. J. Prill, P. A. Iglesias, and A. Levchenko. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biology*, 3(11):1881–1892, 2005.
- [22] E. D. Sontag, A. Veliz-Cuba, R. Laubenbacher, and A. S. Jarrah. The effect of negative feedback loops on the dynamics of Boolean networks. *To appear in Biophysical Journal*, 2008.
- [23] R. Steuer, T. Gross, J. Selbig, and B. Blasius. Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences USA*, 103(32):11868–73, 2006.
- [24] S. Wernicke and F. Rasche. FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.