# IN4176 example proposal

*Based on the thesis proposal of Peter van Nes, 2007*

## Motivation

Knowing the dynamics of metabolic networks in living cells is a prerequisite for simulation or prediction of the effects of a certain perturbation, like a heat shock, a change in $pH$-level, or the injection of a certain chemical. Such simulations and predictions could greatly facilitate several tasks, such as metabolic engineering and drug design. However, the dynamics of large networks can not be accurately measured or even modeled, as too little is known about the kinetics of the enzymes that regulate the reactions in a metabolic network.

## Problem

Although modeling the dynamics of a metabolic network fully is currently not possible, it is possible to obtain its structure. The challenge is to devise methods to infer dynamic properties of large networks based on their structure alone.

## Background of problem

In (1), an attempt was made to construct a link between the structure and the dynamic behaviour of a metabolic network. This was done by introducing the notion of a *motif*, a small subnetwork inside a larger one. Roughly, the work in (1) consists of two steps:

I. determining the structural stability of each motif;

II. determining the over-abundance of each motif in real networks as compared to random networks.

Step I is performed by generating a lot of instances of each motif, by randomly assigning values in the range [-1,1] to the reactions. Step II generates a large number of random networks, counts occurrences of each motif and

finally uses the resulting distribution to come up with a $p$-value for the rarity of the occurrence of each motif.

It was observed that structurally stable motifs are over-abundant in real networks as compared to random networks, whereas less structurally stable motifs are not. The general conclusion was that apparently, evolution favours structurally stable motifs over less structurally stable ones. There are, however, a number of problems with the approach in (1):

- The biological plausibility of the conclusion. According to (2), structural stability is not an intrinsic property of biology itself. Furthermore, in general, a network made up of a lot of structurally stable motifs is not necessarily stable itself. It is not obvious why evolution should prefer structurally stable motifs, if these do not necessarily result in a stable network.

- In (3) it is shown that even a small motif can already exhibit a large range of biological functionality. This is a major obstacle in predicting dynamic behaviour of a network from its structure.

- Generating instances of motifs by sampling edge values (reaction rates) from the range [-1,1]. Although it can be argued that this approach takes all possible reaction rates into account, also those that are influenced by the kinetics of the system, this is biologically not very plausible. Some reactions may have much higher rates than others.

- In step II, the kinetics of the metabolic network are not taken into account. The metabolic networks that are evaluated contain only information from the stoichiometric matrix. This matrix contains, for all reactions, entries for the metabolites that play a role in that reaction, a negative entry indicating that a metabolite acts as a substrate in a particular reaction and a positive entry that it is the product. In (1), a reaction occurring in the stoichiometric matrix is assumed to always take place. In reality, however, chemical reactions are subject to regulation by enzymes. This means that when a certain enzyme, necessary for a certain reaction, is absent, the edge does not exist. Taking enzyme regulation into account might therefore result in different, probably more reliable motif counts.

## Proposal

We propose to extend the work in (1) as follows:

1. finding a better distribution for instantiating values for the edges of the motifs (in step I);
2. taking the kinetics of metabolic networks into account (in step II);
3. looking at another structural property for motifs, monotonicity (replacing step I); and
4. defining a continuous measure for the monotonicity of an entire network, based on its consistency deficit (CD) (replacing step II).

Below, the approach taken in each of these steps is discussed in more detail.

## Approach

### 1. Better edge value distributions

We will use a more biologically plausible distribution to draw values from, e.g. one estimated based on experimental data such as those found in the BRENDA database (4). If we have information on regulatory interactions (by other metabolites or enzymes) we could use this to couple the distributions used for different edges in the network.

### 2. Taking kinetics into account

The experiments of (1) will be repeated, taking kinetics into account by incorporating the approach in (5) for simulating the kinetics of a metabolic network. All possible reaction rates, resulting from the unknown kinetic rate laws and parameters are considered. These unknown parameters constitute a parameter space. For a given point in this space, a Jacobian matrix is constructed by parameterising the original Jacobian matrix. This parameterised Jacobian matrix is roughly equal to a product of the stoichiometric matrix and a saturation matrix. The latter matrix indicates for each reaction the degree of saturation in terms of a particular metabolite. These saturation matrices are randomly generated, so as to simulate the kinetics of the reactions in the metabolic network that is studied. When the saturation for a certain reaction is below some threshold, the reaction could be regarded as switched off and the edge is therefore excluded. This will result in finding a different set of motifs than when the edge is included.

This procedure is repeated $n$ times. On the basis of the $n$ motif counts, we can construct a probability distribution for each motif. Using the mean or median of this distribution might be more accurate than the count used in (1).

### 3. Using monotonicity rather than stability

We will study monotonicity (6) as a possible alternative to stability. Monotone systems behave in a predictable way to perturbations and they have robust dynamical characteristics, which makes them reliable components of larger, more complex networks. A network is monotone if every closed-loop inside this network has an even number of inhibitory reactions. Just as (1) defined structural stability as the stability of a motif, we will define *structural monotonicity* as the monotonicity of a motif. This can be seen as the probability that a certain instance of a motif is monotone.

As a first step, we will have to obtain a signed graph from a metabolic network (7; 8). Consider the following example:

$$R_1 : A + B \rightarrow C + D$$

$$R_2 : D + E \rightarrow F + G$$

$$R_3 : E + H \rightarrow I + J$$

Since metabolite $D$ is a product of reaction $R_1$ but a substrate of $R_2$, one could say that reaction $R_1$ and reaction $R_2$ *co-operate*. Therefore, the substrates of reaction $R_1$ (metabolites $A$ and $B$) have an *activating* effect on the products of reaction $R_2$ ($F$ and $G$), leading to positive edges $A \rightarrow F$, $A \rightarrow G$, $B \rightarrow F$ and finally $B \rightarrow G$. On the other hand, reactions $R_2$ and $R_3$ both have metabolite $E$ as a substrate. Therefore, one could say these two reactions *compete* with each other and that therefore the other substrates of reaction $R_2$ have a *negative* or *inhibitory* effect on the products of reaction $R_3$ and vice versa. So there would be negative edges $D \dashv I$, $D \dashv J$, $H \dashv F$ and finally $H \dashv G$.

### 4. A measure for the monotonicity of a network

It can be shown (6) that networks consisting of monotone subnetworks are themselves monotone or near-monotone. This near-monotonicity can be assessed by using the consistency deficit (CD), the minimal number of edges that should be removed so that the remaining graph is consistent (monotone). If structurally monotone motifs appear to be over-abundant in a large system that is monotone or near-monotone, it might be argued that

these motifs have driven the evolution of real networks. In other words, we propose to examine the correlation between the over-abundance of structurally monotone motifs in a certain network and the CD of this network.

The CD concept can be further extended as follows. A system that is near-monotone could become monotone by switching of some reactions, i.e. edges in the network. The CD counts how many reactions should be switched of for the system to become fully monotone. Calculation of the CD does not take into account, however, that some reactions are more important than other. Some edges in the network are traversed much more frequently than other.

This can be taken into account using the *edge betweenness* of an edge. Edges that are traversed in many shortest paths between all vertices have a higher edge betweenness than those that do not. The edge betweenness for edge $e$ is defined as the number of shortest paths from node $s$ to node $t$ traversing edge $e$, divided by the total number of shortest paths between $s$ and $t$. The same thing can be done for nodes, resulting in the *node betweenness*. A metabolic network is scale-free, meaning that some nodes have a very high degree, the so called hubs, whereas the majority have very low degree. Hubs will have a high node betweenness as a lot of shortest paths will pass through these hubs. It is improbable that a system will switch off an edge that has either a high edge betweenness, or that is connected to a node that has a high node betweenness. Therefore, we could extend the CD taking betweenness into account. One way of doing this is by assigning costs to removing edges, such that edges with high betweenness will have high removal cost.

## Planning and budget

| Month | Activities |
|---|---|
| 1-2 | Gathering and investigating data |
| 2-3 | 1. Finding a better reaction rate distribution |
| 4-5 | 2. Implementing (5) in (1) |
| 6-7 | 3. Implementing structural monotonicity and basic CD |
| 7-8 | 4. Incorporating edge and node betweenness in CD |
| 9 | Writing paper |

The project will need one researcher for a period of 9 months (EUR 20k). Besides a reasonably equipped workstation (EUR 5k), no other resources are required. This brings the total budget to EUR 25k.

### Expected results

We expect to obtain more reliable abundance estimates for the motifs as we take kinetics into account. Also, we expect to see a high correlation between the structural monotonicity of a motif and its abundance in a real network. Furthermore we expect to see a high correlation between the number of structurally monotone motifs and the consistency deficit in a real metabolic network. These two expectations are based on the fact that a metabolic network consisting of many monotone motifs is itself monotone or near-monotone.

## References


[1] R.J. Prill, P.A. Iglesias, and A. Levchenko, "Dynamic properties of network motifs contribute to biological network organization", *PLoS Biology*, vol. 3, no. 11, pp. e343, 2005.

[2] J. Doyle and M. Csete, "Motifs, control, and stability", *PLoS Biology*, vol. 3, no. 11, pp. e392, 2005.

[3] P.J. Ingram, M.P.H. Stumpf, and J. Stark, "Network motifs: structure does not determine function", *BMC Genomics*, vol. 7, no. 1, pp. 108, 2006.

[4] Cologne University Bioinformatics Center, "Brenda, the comprehensive enzyme information system", `http://www.brenda.uni-koeln.de/`, release 2007.1.

[5] R. Steuer, T. Gross, J. Selbig, and B. Blasius, "Structural kinetic modeling of metabolic networks", *Proc. National Academy of Sciences U S A*, vol. 103, no. 32, pp. 11868, 2006.

[6] E Sontag, "Monotone and near-monotone network structure", *arXiv preprints q-bio.MN/0612032 and q-bio.MN/0612033*, 2007.

[7] B. DasGupta, G.A. Enciso, E. Sontag, and Y. Zhang, "Algorithmic and complexity results for decompositions of biological networks into monotone subsystems", *arXiv preprint q-bio.MN/0509040*, 2005.

[8] D. Angeli, P. De Leenheer, and E.D. Sontag, "On the structural monotonicity of chemical reaction networks", *Proc. IEEE Conf. Decision and Control, San Diego, CA*, 2006.