Qualitative response of interaction networks: application to the validation of biological models
Anne Siegel, Carito Guziolowski, Philippe Veber, Ovidiu Radulescu, Michel Le Borgne

To cite this version:
Anne Siegel, Carito Guziolowski, Philippe Veber, Ovidiu Radulescu, Michel Le Borgne. Qualitative response of interaction networks: application to the validation of biological models. 6th International Congress on Industrial and Applied Mathematics. ICIAM 07, Jul 2007, Zurich, Switzerland. inria-00178854

HAL Id: inria-00178854
https://inria.hal.science/inria-00178854
Submitted on 12 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Qualitative response of interaction networks: application to the validation of biological models

Anne Siegel\(^1\), Carito Guziolowski\(^1\), Philippe Veber\(^1\), Ovidiu Radulescu\(^1,2\), and Michel Le Borgne\(^1\)

\(^1\) CNRS - Université de Rennes 1 - Inria, IRISA, Department of Computer Science, Campus de Beaulieu, 35042 Rennes Cedex - France
\(^2\) Université de Rennes 1, IRMAR, Department of mathematics, Campus de Beaulieu, 35042 Rennes Cedex - France

We advocate the use of qualitative models for the analysis of shift equilibria in large biological systems. We present a mathematical method, allowing qualitative predictions to be made of the behaviour of a biological system. These predictions are not dependent on specific values of the kinetic constants. We show how these methods can be used to improve understanding of a complex regulatory system.

There exists a wide range of technics for the analysis of high-throughput data. Following a review by Slonim [1], we may classify them according to the particular output they compute: 1. list of significantly over/under-expressed genes under a particular condition, 2. dimension reduction of expression profiles for visualisation, 3. clustering of co-expressed genes, 4. classification algorithms for protein function, tissue categorization, disease outcome, 5. inferred regulatory networks. The last category may be extended to all model-based approaches, where experimental measurements are used to build, verify or refine a model of the system under study. Following this line of research, we summary in this abstract how to define and to check consistency between experimental measurements and a graphical regulatory model formalized as an interaction graph.

1 Detecting the sign of a regulation and constraining a model

We suppose that we are given the topology of a network (this topology can be obtained from the manually-curated analysis of the literature, ChIP-chip data or any computational network inference method). In this network, let us consider a node \( A \) with a single predecessor. In other words, the model tells that the protein \( B \) acts on the production of the gene coding for \( A \) and no other protein acts on \( A \). Independently, we suppose that we have several gene expression arrays at our disposal. One of these arrays indicates that \( A \) and \( B \) simultaneously increase during a steady state experiment. Then, the common sense says that \( B \) must have been as activator of \( A \) during the experiment. More precisely, protein \( B \) cannot have inhibited gene \( A \), since they both have increased. We say that the model predicts that the sign of the interaction from \( B \) to \( A \) is positive.

<table>
<thead>
<tr>
<th>Model</th>
<th>( B \rightarrow A )</th>
<th>( B \rightarrow^{[+]} A )</th>
<th>( B \rightarrow^{-[-]} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression profile</td>
<td>( B ) increases ( C ) increases</td>
<td>( B ) increases ( C ) decreases</td>
<td>( B ) increases ( A ) decreases</td>
</tr>
<tr>
<td>Prediction</td>
<td>The action from ( B ) to ( A ) is an activation</td>
<td>Model and data are inconsistent</td>
<td>( C ) increases</td>
</tr>
</tbody>
</table>
use the rule as a constraint for the model. We write constraints for all the nodes of the model and we use several approaches in order to solve the system of constraints. From the study of the set of solutions, we deduce which signs are surely determined by these rules. Then we obtain minimal obligatory conditions on the signs, instead of most probable signs given by probabilistic methods.

Formally, we consider a system of \( n \) chemical species \( \{1, \ldots, n\} \). These species interact with each other and we model these interactions using an interaction graph. The set of nodes is denoted by \( V = \{1, \ldots, n\} \). There is an edge \( j \to i \in E \) if the level of species \( j \) influences the production rate of species \( i \). Edges are labeled by a sign \( \{+,-\} \) which indicates whether \( j \) activates or represses the production of \( i \).

In a typical stress perturbation experiment a system leaves an initial steady state following a change in control parameters. After waiting long enough, the system may reach a new steady state. Expression arrays provide an order of magnitude for the ratio between initial and final levels. We denote by \( \text{sign}(X_i) \in \{+,-,0\} \) the sign of variation of species \( i \) during a given perturbation experiment, and by \( \text{sign}(j \to i) \in \{+,-\} \) the sign of the edge \( j \to i \) in the interaction graph. Let us fix species \( i \) such that there is no positive self-regulating action on \( i \). For every predecessor \( j \) of \( i \), \( \text{sign}(j \to i) \ast \text{sign}(X_j) \) provides the sign of the influence of \( j \) on the species \( i \). Then, we can write a constraint on the variation to interpret the rule previously stated.

\[
\text{sign}(X_i) \approx \sum_{j \to i} \text{sign}(j \to i) \text{sign}(X_j).
\]

Each variable belongs to the sign algebra \( \{+,-,?,0\} \), provided with natural arithmetic operations + and \( \times \) on signs. Since we are using signs instead of numerical values, the equality relation is replaced by sign compatibility relation \( \approx \), ensuring that \( + \approx -\) [4]. This equation has to be slightly modified when the experiment is a genetic perturbation. For a given interaction graph, we refer to the qualitative system associated to \( G \) as the set made up of constraint (1) for each node in \( G \).

### 2 Applications

When the user has a signed interaction graph at its disposal, the constraint equation (1) is used to validate and predict new signs variations in the network. In this case, the signs \( \text{sign}(j \to i) \) in Eq. (1) are all known. This yields a system of qualitative equations with a variable \( s_i = \text{sign}(X_i) \) for each specie.

We first check the internal consistency of the network, that is, whether there exists at least a value for each variation \( s_i \) such that all constraints are satisfied simultaneously.

When the system is not self-consistent, we apply a diagnosis process allowing to identify subsets of the network that cause inconsistency. Then the user should propose a correction of the model. This procedure was applied to E. Coli transcriptional regulatory network as provided in RegulonDB [5]. We deduced that transcriptional regulations are not sufficient to explain any stress perturbation of the system: complex formation implying IHF has to be taken into account.

When the system is self-consistent, we add perturbation data to the system of equations, fixing the value of some variables \( s_i \). We then check the consistency of the new qualitative system of equation. When inconsistencies appear, we apply again our diagnosis algorithm to isolate pieces of the network that carry inconsistencies.

When the system and data are consistent, the qualitative system of equations has not only one solution but a huge number of solutions. However, some variables \( s_i \) may be assigned the same value in all solutions of the system. We refer to these inferred variation signs as predictions of the qualitative system. Applied to E. Coli transcriptional regulatory network including IHF complex, and corrected observations during stationary phase on 40 species, we were able to predict the variation of 381 (25%) components of our network [6].

When the user has a unsigned interaction graph at its disposal, we use expression profiles to test the consistency of network and infer signs of regulations. The signs of interactions \( i \to j \) correspond to variables \( s_{ij} \) and each different expression profiles generate a set of variables \( x_i^k \) for species. These variables are constrained by Eq. (1). The same process as before (self-consistency, diagnosis and correction, consistency with data, new diagnosis and prediction) allows to isolate pieces of the network that are not modelled correctly and to infer signs of interaction in the remaining part of the network. Applied to several S. Cerevisiae networks, we obtain information on 15% to 30% of S. Cerevisiae regulation signs.

On a computational background, we use Decision Diagrams and constraints programming to study the system of qualitative equations [4]. The tools are available on the web page [http://www.irisa.fr/symbiose/bioquali](http://www.irisa.fr/symbiose/bioquali/)

### References