Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks

A. Siegel[1], O. Radulescu[2], M. Le Borgne[1], P. Veber[1], J. Ouy[1], S. Lagarrigue[3]

Preprint submitted to *BioSystems*

[1] Projet Symbiose. Institut de Recherche en Informatique et Systèmes Aléatoires, IRISA-CNRS 6074-Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France
[2] Institut de Recherche Mathématique de Rennes, UMR-CNRS 6625, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France
[3] UMR Génétique animale, Agrocampus Rennes-INRA, 65 rue de Saint-Brieuc, CS 84215 Rennes, France

Abstract. *We introduce an approach to test the compatibility between differential data and knowledge on genetic and metabolic interactions. A behavioral model is represented by a labeled oriented interaction graph. The predictions of the behavioral model are compared with experimental data.*
We exploit a system of qualitative equations deduced from the interaction graph, which is linear in the sign algebra. We show how to partially solve the qualitative system. We also identified incompatibilities between the model and the data. Independently, we detect competitions in the biological process that is modeled. This approach can be used for the analysis of transcriptomic, metabolic or proteomic data.

Introduction

Experiments and graph behavioral models
DNA microarrays are very useful for the discovery and the understanding of cellular genetic regulations. They allow us to compare differential gene expressions for a very high number of targeted genes. However, the measured expression levels in DNA microarrays are affected by many errors and variability. We interpret DNA microarrays as qualitative data: we only consider the sign of the difference in the expression levels among the two situations.

Gene regulations control cellular functions, including many metabolic functions. Symmetrically, metabolic processes involve interactions between genes and metabolites. In order to study the relations between genes and metabolic regulations, we need information on variations of metabolite fluxes. This type of information is provided by biochemical measurements.

Our goal is to develop testing methods for the compatibility between qualitative measurements and knowledge on the interactions among the measured products (metabolites and genes).

Biological literature is rich in knowledge about interactions between molecules. This information is scattered in many publications and it is not globally compared to the results of large scale experiments. In order to use this knowledge we have developed a two step procedure. In the first step we have used biological literature in order to build a behavioral model of regulation network. In the second step, this model has been confronted to data. In this paper we provide the details about the second step.

We use the information on a system to build a labeled oriented graph (interaction graph). The nodes of this graph represent molecules (such as mRNA, proteins or metabolites). The edges represent interactions, labeled by their sign. Edge signs can be interpreted dynamically: a "+" on an edge between A and B means that an increase of the concentration of A increases the rate of production of B. A "−" on an edge stands for a decreasing effect. The necessary information can be collected either from interaction databases or can be manually extracted from the literature.

Qualitative analysis In this paper, we present a formal analysis of interaction graphs, based on a mathematical model. We suppose that the dynamics of the system can be described by a system of differential equations. An experiment is modeled with a steady state shift of the differential dynamical sys-

tem. Notice that the interaction graph summarizes the qualitative knowledge on the dynamics. We perform transformations on the differential system that allow to connect the variations of products in a linear system (called quantitative linear system).

We suppose that experimental information on products is qualitative, that is, we can find out if the concentration of a product has increased or not. In order to exploit this information, qualitative models are derived from the interaction graph and from the differential dynamics: we transpose quantitative equations as a linear *qualitative* system in the sign algebra [Kui94, TMD03].

The quantitative system leading to the qualitative one represents a generalization of the discrete Laplace equation on graphs. Like for the Laplace equation, we solve the Dirichlet problem; this means relating the interior values to the values on the entrance boundary of the oriented graph [CY00, Soa94, CP03].

We have developed several complementary methods to analyze the system of qualitative equations and test whether experimental data are solutions of the system. The modeling and the algorithms will be illustrated on a simplified model of regulated lipogenesis in liver.

- The *graph valuation algorithm* is an automated method to partially or totally solve the qualitative system of equations. When there is no solution, incompatibilities between the model and the data are detected, as well as the place where these incompatibilities occur. When a solution exists, the algorithm predicts the signs of variations of products that have not been measured.
- The *essential balance computation* allows us to detect the variables in the system that are influenced by competitive pathways and to find which pathway wins the competition. Eventually, this computation detects the presence of errors in the data.

Discussion By the use of qualitative descriptions, our methodology is obviously complementary to quantitative and simulation oriented methods on biological networks.

Our approach is also complementary to the field of metabolic pathway analysis [PSP+04]. Systems we mostly intend to analyze have the following characteristics: either they include both metabolic and genetic regulations, or very low information about interactions is available.

Interaction graphs can be built by network reconstruction techniques [KF02, YVK04, NRF04]). These inference techniques need huge amounts of data which are not always available. The mathematical formalism that we introduce here may lead to new "reverse engineering" algorithms or may serve to assess the validity of the existing ones.

Nevertheless, our practical objective is different. Instead of using data and gene perturbations for building networks from scratch, we intend to refine the analysis of incomplete models. Hence, the networks that we study are inferred from various different sources and contain quite a large number of variables.

This paper is organized as follows. In Section 1 we describe our working example which is the regulation of lipogenesis in liver. Section 2 is devoted to modeling assumptions. We introduce the mathematical concepts and we show how the qualitative equations can be obtained. We detail in Section 3 two methods to solve the system of qualitative equations derived in Section 2. Section 4 is devoted to an independent method for analyzing incompatibilities and competitions.

1 Working example: regulation of the synthesis of fatty acids

Regulation of fatty acid synthesis in liver is our working example. The corresponding interactions are intricate and involve hundreds of nodes. By way of illustration we have kept only the most important nodes and those other nodes that have direct connections to them. The choice of the observed nodes was not optimized; we shall see how our analysis suggests improvements on it.

Two ways of production of fatty acids coexist in liver. Saturated and mono-unsaturated fatty acids are produced from citrates due to the presence of a metabolic pathway composed of four enzymes, namely ACL (ATP citrate liase), ACC (acetyl-Coenzyme A carboxylase), FAS (fatty acid synthase) and SCD1 (Stearoyl-CoA desaturase 1). Polyunsaturated fatty acids (PUFA) such as arachidonic acid and docosahexaenoic acid are synthesized from essential fatty acids provided by nutrition; D5D (Delta-5 Desaturase) and D6D (Delta-6 Desaturase) catalyze the key steps of the synthesis of PUFA.

PUFAs play pivotal roles in many biological functions. They regulate the expression of genes that impact on lipid, carbohydrate, and protein metabolism. The effects of PUFA are mediated directly or indi-

rectly as shown in the following examples. Direct effects are due to bindings leading to changes in the trans-activating activity of nuclear receptors (PPARα – peroxisome proliferator activated receptors, LXRα – Liver-X-Receptor α, HNF-4α). Indirect effects result in changes in the abundance of regulatory transcription factors (SREBP-1c – sterol regulatory element binding-protein–, ChREBP, etc.) [Jum04].

Variables in the model We have considered in our model the transcription factors PPARα, LXRα and SREBP-1c (denoted by PPAR, LXR, SREBP), as they are synthesized from the corresponding genes. We have included the trans-activating active forms of these nuclear receptors: LXR-a (denoting a complex LXRα:RXRα), PPAR-a (complex PPARα:RXRα) and SREBP-a (cleaved form of SREBP-1c). We have also considered SCAP (SREBP cleavage activating protein), a key enzyme involved in the cleavage of SREBP-1c. SCAP interacts with another family of proteins called INSIG, showing the complexity of molecular mechanism. We have included PUFA to symbolize metabolites. Finally, we have considered the enzymes ACL, ACC, FAS, SCD1 (implied in the fatty acid synthesis from citrate) and D5D, D6D (implied in PUFA synthesis).

Interactions in the model Relations between the variables are as following. SREBP-a is an activator of the transcription of ACL, ACC, FAS, SCD1, D5D and D6D [NHT$^+$02, Jum04]. LXR-a is an activator of the transcription of SREBP and FAS. It also indirectly activates ACL, ACC and SCD1 [SG04]. These indirect actions are kept in the model because we don't know whether they are only SREBP-mediated.

PUFA activates the formation of PPAR-a from PPAR. It inhibits the formation of LXR-a from LXR. It also inhibits the formation of SREBP-a (by inducing the degradation of mRNA and inhibiting the cleavage) [Jum04]. SCAP represents the activators of the formation of SREBP-a from SREBP. SCAP is inhibited by PUFA.

PPAR directly activates the transcription of SCD1, D5D, D6D [MN96, TCNC03, MSY$^+$02a].

The activation of SCD1, D5D and D6D by both SREBP and PPAR is paradoxical because fatty acid synthesis (partially governed by SREBP) and oxidation (partially induced by PPAR) are antagonistic in liver. Nevertheless, PUFA have a regulatory role in oxidation. Hence, the induction of D5D and D6D gene by PPAR could be a compensatory response to the increased PUFA demand caused by induction of fatty acid oxidation.

Fasting-refeeding protocols The fasting-refeeding protocols offer favorable conditions for the study of lipogenesis regulation. Our assumption is that during an experimentation, subjects (such as rodents or chicken) were kept in fasted state for several hours. Then, hepatic mRNA of LXR, SREBP, PPAR, ACL, ACC and SCD1 were quantified by DNA microarray analysis. Biochemical measures also provided the variation of PUFA.

A compilation of recent literature on lipogenesis regulation provides hypothetical results of such protocols. During the fasting state, the fatty acid synthesis was inhibited and the fatty acid oxidation was activated. PUFA levels can be considered to be increased in liver following starvation because of the important lipolysis from adipose tissue as shown by Lee et al in mice after 72h fasting [LCL$^+$04]. SREBP, ACL, ACC, FAS and SCD1 was declined [LYH$^+$02]. Tobin et al [TSA$^+$00] have shown that fasting rats for 24h increases the hepatic LXR mRNA.

Matsuzaka et al [MSY$^+$02b] observe no difference in either the hepatic D5D or D6D mRNA level on fasting.

Working data set Our goal is to test whether data can fit properly with the full knowledge on the system. Such data can be an experimental set that have to be analyzed. It could also simply be a set of hypothetical data whose coherence have to be tested.

Here, we assume that observations stated that ACL, LXR, PPAR and PUFA increase. SREBP, ACC, FAS and SCD1 are supposed to decrease. We assume that D5D and D6D are unchanged. By doing this, we voluntarily introduce a data that does not fit with results appearing in the literature: ACL should decline instead of inclining. This error is introduced to test whether the analyze is able to notice it and possibly correct it or not.

2 Steady state shift: qualitative description

We justify here the approach and impose some applicability limits. The basic mathematical object is a qualitative graph called the interaction graph. This graph represents the present knowledge on interactions for the biological process under study. Experimental data give variations of the concentrations of some molecules between two external conditions. The corresponding nodes of the interaction graph will be called *observed nodes*.

Differential model The model describes the interactions between a set of molecules indexed by a set $I = \{1, \ldots, n\}$. The concentration of the molecule indexed by i is X_i. The quantity $\frac{dX_i}{dt}$ denotes the speed of variation of X_i.

We assume that the various concentrations involved in the model evolve according to a differential dynamics. The set of control parameters of the dynamics is denoted by $\mathbf{P}$. The state vector $\mathbf{X} = (X_1, \ldots, X_n)$ satisfies a differential equation

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P}).$$

2.1 Interaction graph

Qualitative graph The set of nodes $I \cup \{\mathcal{E}\}$ gathers all the molecules I in the model together with an extra generic node $\mathcal{E}$ to represent the exterior world.

The set of edges $A \subset (I \cup \{\mathcal{E}\}) \times (I \cup \{\mathcal{E}\})$ contains oriented edges joining $j \in I$ to $i \in I$ whenever $\frac{\partial F_i}{\partial X_j} \neq 0$. This description implies that the interaction graph contains no multi-arcs.

Additionally, the set A contains edges from the exterior world $\mathcal{E}$ to I specifying which are the entrance nodes of the system. If there is an edge from $\mathcal{E}$ to $i \in I$, the node i is potentially affected by nodes that are not included in the model. If there is no edge between $\mathcal{E}$ and i, then the node i is supposed to be influenced only by I. Specifying these edges is an important modeling hypothesis.

Every edge in $A \cap (I \times I)$ is associated to $s(j,i) = sign(\frac{\partial F_i}{\partial X_j}) \in \{+, -\}$ and to the real values $a(j,i) = \frac{\partial F_i}{\partial X_j}$ (called interaction coefficients).

The interaction graph $M = (I \cup \{\mathcal{E}\}, A, s)$ is the graph labeled by $s(j,i)$, $(j,i) \in A \cap (I \times I)$. Hence, we are not using the information on the form of the function F or on the absolute values of its partial derivatives here. Notice that the label $s(j,i)$ depends on the variables $\mathbf{X}$ and $\mathbf{P}$. We suppose that $s(j,i)$ does not change within domains of experimental interest.

Exterior and parameters The extra generic node $\mathcal{E}$ (exterior world) represents molecules that are not included in the model and that interact with the molecules in the model. The node $\mathcal{E}$ and the parameters $\mathbf{P}$ both represent variables that act on the equilibrium of the molecules in the model. Nevertheless, there are physical and biological differences among them. The node $\mathcal{E}$ stands for unknown dynamical variables, while $\mathbf{P}$ stands for control parameters. $\mathcal{E}$ may for instance integrate unknown transcription factors that are not considered in the model, but that should affect the dynamics and the equilibrium of the molecules in the model. $\mathbf{P}$ include external conditions such as food, temperature, stress, or internal conditions such as particularities of individual organisms, mutations, etc. It is conceivable to change several parameters in a well defined way, but it is useless to particularize several unknown external molecules. That is why we may keep several control parameters, but it is enough to consider only one node $\mathcal{E}$ for the exterior world.

Extracting an interaction graph from biological facts described in the literature Intuitively, an arc $j \to i$ describes the alteration of the flux of the product i when the concentration of j is modified. We build the interaction graph associated to a given biological question from the biological literature. More precisely, we consider an edge from j to i in the graph when experimentation in a paper shows that the production of i is modified after a change in the concentration of j.

The interaction graph is supplied with negative self-interaction on all nodes of the graph, implying $s(i,i) = -$ for any i (in absence of additional positive self-regulation). These result for instance from degradation processes. Negative self-interaction ensures the existence and the stability of equilibrium.

Such an interaction graph might be incomplete: it is not necessary to gather all the biological facts in a model. Actually, we intend to understand whether experimental data fits with the theoretical model or not.

For the biological problem used as an illustration, we consider a model $M = (I \cup \{\mathcal{E}\}, A, s)$ with vertices $I = \{$PUFA, LXR, LXR-a, PPAR, PPAR-a, S-REBP, SREBP-a, SCAP, ACL, ACC, FAS, SCD1, D5D, D6D$\}$. The interaction graph M is shown in Fig.2.1.

Paths and loops on graph A path $j \rightsquigarrow i$ is a sequence of nodes $\{i_1, i_2, \ldots, i_p\}$ such that $(i_k, i_{k+1}) \in A$, $i_1 = j$ and $i_p = i$. The nodes of a path are visited just once. Let $a_{j \rightsquigarrow i}$ denote the product of interaction coefficients along a path

$$a_{j \rightsquigarrow i} = \prod_{k=1}^{k=p-1} a(i_k, i_{k+1}).$$

The sign of the path is $s(j \rightsquigarrow i) = sign(a_{j \rightsquigarrow i})$.

A loop is a path with identical terminal nodes $i_1 = i_p$. In a loop all nodes are visited just once, excepting the terminal nodes.

Analyzed subgraph, observed nodes The subset of the nodes in the model on which we perform

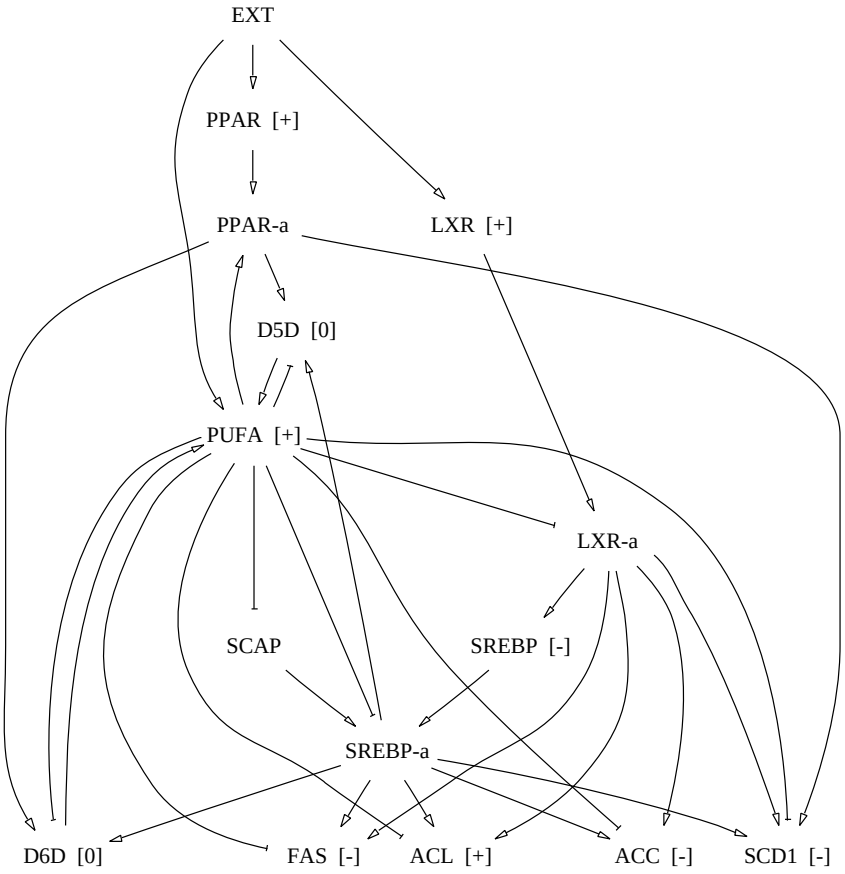

Figure 2.1: Interaction graph A for a model $M = (I, A, s)$ of regulation of the synthesis of fatty acids. The node EXT stands for $\mathcal{E}$ and represents the exterior world. Self-regulation loops on nodes are omitted for sake of clarity. Observed variations are $s(\delta\text{PPAR}) = +$, $s(\delta\text{PUFA}) = +$, $s(\delta\text{LXR}) = +$, $s(\delta\text{SREBP}) = -$, $s(\delta\text{SCD1}) = -$, $s(\delta\text{FAS}) = -$, $s(\delta\text{ACL}) = +$,$s(\delta\text{ACC}) = -$, $s(\delta\text{D5D}) = 0$, $s(\delta\text{D6D}) = 0$.

the analysis is denoted by $G \subset I$. The nodes in G inherit the connections from M.

The set of observed nodes will be denoted as $\mathcal{O}$. One can define the nodes variations $\delta X_i, i \in \mathcal{O}$, given by experimental data on $\mathcal{O}$. The sign of these variations is denoted by $s(\delta X_i)$.

In our example, we consider $G = I$. Only some nodes are observed. The variations of observed nodes are shown in Fig.2.1.

$\mathcal{O}$ = { PPAR, LXR, ACL, ACC, SREBP,
FAS, SCD1, PUFA, D5D, D6D }.

Entrance boundary The analyzed subgraph G is naturally partitioned into two subsets: the entrance boundary of G and the interior nodes.

The entrance boundary of G, denoted $\urcorner^{in}G$ is formed by all the nodes of G that are supposed to have entering connections from the outside:

$$\urcorner^{in}G = \{i \in G \mid \exists j \in (I \cup \{\mathcal{E}\}) \setminus G, \quad (j,i) \in A\}.$$

The set of interior nodes is denoted $\mathring{G} = G \setminus \urcorner^{in}G$.

Notice that when $G = I$, the entrance boundary is entirely made of those nodes that have connections from $\mathcal{E}$.

In the model M of our working example, the analyzed subgraph is I. Its entrance boundary is $\urcorner^{in}I$ = {PPAR, LXR,PUFA}.

Predecessors We denote by $pred(i)$ the set of predecessors of i: $pred(i) = \{j \in I \cup \{\mathcal{E}\}, j \neq i, (j,i) \in A\}$.

2.2 Qualitative linear equations

Steady state shift In our second modeling assumption, we suppose that for the two experimental situations which are compared the biological system has reached steady state. Steady states are characterized by:

$$F(\mathbf{X}, \mathbf{P}) = 0. \tag{2.1}$$

The steady state shift is the result of a change in the control parameters $\mathbf{P}$. In our example fasting is the only control parameter. Our modeling assumption implies that the parameters take continuous values. This seems to be wrong for fasting, which is either applied or not. Nevertheless, it is not artificial to imagine a continuous set of intermediate steady states corresponding to more or less poor feeding.

Experimental data are interpreted as variations, between two steady states, of the variables associated with observed nodes.

This point of view, which is assumed in the rest of the paper, is not correct in all situations. It cannot be assumed for time series where observations are made at relatively close instants. In such a situation, it is not true that the system has reached steady state at each instant when it is observed. For our example, this hypothesis is justified since data come from metabolism of animals that where kept under normal diet or fasted during a long time before the measurements.

Small variation and self-impedance We differentiate Eq.2.1 in order to understand how the steady state changes for small changes in the parameters. For a node $i \in I$ one gets the equation:

$$\frac{\partial F_i}{\partial X_i}\delta X_i + \sum_{j \neq i} \frac{\partial F_i}{\partial X_j}\delta X_j + \sum_k \frac{\partial F_i}{\partial P_k}\delta P_k = 0.$$

We can notice that if $\frac{\partial F_i}{\partial X_j} \neq 0$ and $j \neq i$ then $j \in pred(i)$. Let us consider a node i such that $\frac{\partial F_i}{\partial P_k} = 0$: the production or the consumption of X_i does not depend directly on the parameters.

Property 2.1
Let us consider a system whose steady state depends on the parameters. Let i be a node that satisfies the following conditions:

- *the node i is not directly influenced by the parameters (all influences come via its predecessors), i.e. $\frac{\partial F_i}{\partial P_k} = 0$,*
- $\left(\frac{\partial F_i}{\partial X_i}\right) \neq 0$,

then the variation in i can be entirely calculated from the variations in the predecessors nodes:

$$\delta X_i = -\left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{k \in pred(i)} a(k,i)\delta X_k. \tag{2.2}$$

Using a mechanical analogy, $\chi_i = -\left(\frac{\partial F_i}{\partial X_i}\right)$ can be called self-impedance: it is the ratio between the force produced by the predecessors of i and the variation in i. As discussed, there are always effects that produce self-impedance (degradation, self-regulation). The variation δX_i can be calculated from the force exerted by the predecessors only if the self-impedance χ_i is non-zero. A zero self-impedance would represent a non-generic case when the sum of all these effects cancel exactly.

Sign algebra If we only know the signs of the variations and of the interaction coefficients, we intend to find the relations between the signs of the variations and the signs of the interaction coefficients. To do so, we transcribe Eq.2.2 into sign algebra. Remind that $s(\delta X_i)$ stands for the sign of the variation of X_i.

By sign algebra, we mean that $+$ and $-$ are embedded in the set of subsets $\{\{+\},\{-\},?\}$, where $? = \{+,-\}$ means undetermined sign. The subsets are then provided with the composition rules: $\{+\}+\{-\} =?$, $\{+\}+\{+\} = \{+\}$, $\{-\}+\{-\} = \{-\}$, $\{+\} \times \{-\} = \{-\}$, $\{+\} \times \{+\} = \{+\}$, $\{-\} \times \{-\} = \{+\}$.

An equality in an equation means that the sign corresponding to the l.h.s. (left hand side) and the sign of the r.h.s. (right hand side) have a non empty intersection [Kui94, TMD03]. In some cases, zero is added to the sign algebra with usual meaning in sums and products.

Generalization to large variations When Eq.2.2 is transcribed in the sign algebra, it is likely that sign relations will be valid not only for small variations, but also for large variations. For this to be true, it is enough to consider that the signs of the interaction coefficients are constant within the experimental range of variation of the parameters and to integrate Eq.2.2 with respect to the parameters:

$$\begin{aligned} \Delta X_i &= \int_P^{P+\Delta P} \frac{dX_i}{dP} dP \\ &= \int_P^{P+\Delta P} \frac{1}{\chi_i} \sum_{k \in pred(i)} a(k,i) \frac{dX_k}{dP} dP. \end{aligned} \tag{2.3}$$

By writing $\frac{dX_i}{dP}$ in Eq.2.3 we supposed that at steady state X_i is a piecewise differentiable function

of the parameters P. This is not incompatible with multistationarity (existence of several steady states for the same values of the parameters). In the case of multistationarity, jumps occur at discrete critical values of the parameters. Between these critical values, the dependence of the steady state on the parameters is smooth. We shall only suppose that the signs of the interaction coefficients $a(k,i)$ do not change at critical values, hence it is the same for all steady states.

Qualitative system of equations Transcribing Eq. 2.3 in the sign algebra provides the following result.

Theorem 2.2
Let M be a biological system involving n variables with values $(X_1, \ldots, X_n)$, and following the dynamics $\frac{d\mathbf{X}}{dt} = F(\mathbf{X}, \mathbf{P})$. Let i be a node that fulfills the following conditions:

- *all influence of the parameter variations on the variable X_i is mediated by the predecessors of i in the interaction graph, i.e. $\frac{\partial F_i}{\partial P} = 0$;*
- *the self-impedances $\chi_i = -\frac{\partial F_i}{\partial X_i}$ are strictly positive, for all the steady states, within the entire range of parameters variations;*
- *the signs of the interaction coefficients $\frac{\partial F_i}{\partial X_j}$ are constant, for all the steady states, within the entire range of parameters variations.*

Then, the signs of finite variations of steady state variables are connected by:

$$s(\Delta X_i) = \sum_{k \in pred(i)} s(k,i) s(\Delta X_k). \qquad (2.4)$$

Example Theorem 2.2 provides the following equations for the working example M (for ease of notations, $s(\Delta A)$ is denoted as A). One can apply Theorem 2.2 because the self-impedances are positive in absence of positive self-regulation. Nodes i that satisfy the hypothesis of Theorem 2.2 are those which are not connected to the exterior world $\mathcal{E}$ (implying that the first condition is satisfied).

System 1

```
(1)  PPAR-a = PPAR + PUFA
(2)  LXR-a = -PUFA + LXR
(3)  SREBP = LXR-a
(4)  SREBP-a= SREBP + SCAP -PUFA
(5)  ACL= LXR-a  + SREBP-a - PUFA
(6)  ACC= LXR-a + SREBP-a - PUFA
(7)  FAS=  LXR-a + SREBP-a - PUFA
(8)  SCD1 = LXR-a  + SREBP-a - PUFA + PPAR-a
(9)  SCAP=  -PUFA
(10) D5D= PPAR-a + SREBP-a - PUFA
(11) D6D = PPAR-a + SREBP-a - PUFA
```

Known values are the following:

Values 1

```
PPAR=+, PUFA=+, LXR=+, SREBP=-, SCD1=-,
FAS=-, ACL=+, ACC=-, D5D=0, D6D=0
```

2.3 Influences and their transmission across a boundary

In order to analyze data, we need to transform the qualitative system derived in Theorem 2.2 to obtain a system where the variables are functions of a fixed set of variables. In general, this kind of system transformation in the sign algebra is a NP-complete problem [Dor88].

In the following paragraphs we show that quantitatively this operation is equivalent to solving the Dirichlet problem for a discrete operator on the interaction graph. A solution to the Dirichlet problem can be found under mild conditions. We use this result for deriving a large class of qualitative systems.

Transmission of influences on graphs One can read Eq.2.2 as the propagation of influence on the interaction graph: the concentration variation of the molecule i is an average of the influences from its predecessors. Let us define as follows the discrete operator $L : \mathbb{R}^I \to \mathbb{R}^I$:

$$(L\delta X)_i = \delta X_i + \left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{k \in pred(i)} a(k,i) \delta X_k. \quad (2.5)$$

Then Eq.2.2 reads $L\delta X = 0$ which is analogous to the Laplace equation on graphs. Let us remind that the Laplace equation fulfills an analogous property: the value of an harmonic function (satisfying the Laplace equation) in a node is the average of the values of the function on the predecessors of this node.

The analogy can be pushed further as one may try to define the influence of the boundary on an interior point (question known as the Dirichlet problem) [CY00].

Laplace equation on graphs was intensively studied in connection to electrical networks and random walks on undirected graphs [Soa94]. Let us notice that our model describes propagation of influence on oriented graphs. This could be related to random

walks on oriented graphs. Nevertheless, the propagation operators resulting from random walks have negative non-diagonal elements [CP03], while in our case there are no local constraints on the signs (a global constraint is imposed by the stability of steady state which asks that all the eigenvalues of the Jacobian lie in the left half of the complex plane). Little is known on the properties of random walks on oriented graphs. For instance it is not known whether the solution to the Dirichlet problem has a representation in terms of averaged stopped random walks or not. The connection between propagation of influences and random walks on oriented graphs is promising. Reasonings and details of this will be presented in a future publication.

Influence of the boundary on the interior From now, we fix a subnetwork $G \subset I$ of the model. We have already defined the entrance boundary $\urcorner^{in}G$, formed by all the nodes of G that are supposed to have entering connections from the outside. The set of interior nodes was denoted $\mathring{G} = G \setminus \urcorner^{in}G$.

By differentiating the components of Eq.2.1 corresponding only to the interior nodes of G one gets:

$$\forall i \in \mathring{G}, \quad \sum_{j\in\mathring{G}} \frac{\partial F_i}{\partial X_j}\delta X_j + \sum_{k\in\urcorner^{in}G} \frac{\partial F_i}{\partial X_k}\delta X_k + \sum_{l\in I\setminus G} \frac{\partial F_i}{\partial X_l}\delta X_l + \frac{\partial F_i}{\partial P}\delta P = 0.$$

Noticing that $\frac{\partial F_i}{\partial X_l} = 0$ for all $i \in \mathring{G}$, $l \in I \setminus G$ and assuming that $\frac{\partial F_i}{\partial P} = 0$ for all $i \in \mathring{G}$ it follows:

$$\forall i \in \mathring{G}, \sum_{j\in\mathring{G}} \frac{\partial F_i}{\partial X_j}\delta X_j = -\sum_{k\in\urcorner^{in}G} \frac{\partial F_i}{\partial X_k}\delta X_k. \quad (2.6)$$

Eq.2.6 is the starting point for the proof of the following theorem (details can be found in the Appendix):

Theorem 2.3
Let G be a subgraph of I. Let $\mathring{A}$ be the restriction of the jacobian of F to the interior nodes: $\mathring{A}_{ij} = \frac{\partial F_i}{\partial X_j}, i, j \in \mathring{G}$. Let us suppose that:

- *$det(\mathring{A}) \neq 0$,*
- *there is no direct influence of the parameters on the interior nodes of G, i.e. $\frac{\partial X_i}{\partial P_k} = 0$, $\forall i \in \mathring{G}$.*

Then, the response of $i \in \mathring{G}$ to small changes on the boundary of G is given by:

$$\delta X_i = \sum_{j\in\urcorner^{in}G} \sum_{j\rightsquigarrow i\in\mathcal{P}_{\mathring{G}}} \frac{a_{j\rightsquigarrow i}}{\mathring{C}_{j\rightsquigarrow i}}\delta X_j, \quad (2.7)$$

with the following notations:

- *$\mathcal{P}_{\mathring{G}}$ denotes the set of paths included in G, starting on the boundary and that do not return to the boundary: $\mathcal{P}_{\mathring{G}} = \{j \rightsquigarrow i = (i_1 = j, \ldots, i_k, \ldots, i_p = i), i_k \notin \urcorner^{in}G, \forall k > 1\}$;*
- *$a_{j\rightsquigarrow i} = \prod_{k=1}^{p-1} a(i_k, i_{k+1})$;*
- *$\mathring{C}_{j\rightsquigarrow i} = (-1)^{l_{k(j)\rightsquigarrow i}+1} \frac{det\mathring{A}}{det\mathring{A}_{\{k(j)\rightsquigarrow i\}^c}}$ denotes the "path impedance" of the internal path $k(j) \rightsquigarrow i$ where*
  - *$k(j) \in \mathring{G}$ is the second node after j of the path $j \rightsquigarrow i$;*
  - *$l_{k(j)\rightsquigarrow i}$ is the length of the path $k(j) \rightsquigarrow i$, i.e. the number of arcs;*
  - *$det\mathring{A}_{\{k(j)\rightsquigarrow i\}^c}$ is the principal minor obtained by eliminating all the lines and all the columns whose indices are in the path $k(j) \rightsquigarrow i$ from the jacobian $\mathring{A}$. Conventionally, if the resulting minor is empty we choose it equal to one: $det\mathring{A}_{\mathring{G}^c} = 1$.*

  If $k(j) = i$ the path impedance becomes the node impedance $\mathring{C}_i = -\frac{det\mathring{A}}{det\mathring{A}_{\{i\}^c}}$.

If we consider $G = \{i\} \cup pred(i)$ to be the set of a predecessors of a node i we obtain the Property 2.1 as a corollary.

2.4 Signs of impedances

Path impedance Eq.2.7 describes the influence on an interior node of variations on the boundary. j acts with a *force* $a_{j\rightsquigarrow i}\delta X_j$ on the node i, along the path $j \rightsquigarrow i$. This force is bigger when the product of interaction coefficients $\frac{\partial F_{i_{k+1}}}{\partial X_{i_k}}$ along the path is bigger.

$\mathring{C}_{k(j)\rightsquigarrow i}$ is the ratio force/response and therefore can be called *path impedance* or *path modulus*. According to Theorem 2.3, the contribution to the impedance comes only from the internal part of the path: $k(j) \rightsquigarrow i$. If the internal part reduces to a point ($k(j) = i$, meaning that the path from j to i is a direct arc), then the contribution to the impedance comes from the node i only: the path impedance becomes node impedance. A large path (or node) impedance implies a small response at the end of the path, even if the force is big. Therefore, the impedance can be related to sensitivity.

Signs of impedance One way to test the compatibility between micro-array data and interaction models is to compare predicted and observed signs of the variations. In order to do so, we need to know the

signs of the path impedances. The following property is very useful:

Property 2.4
If the interior of the subgraph G contains no positive loops, then all internal path impedances $\mathring{C}_{k(j)\rightsquigarrow i}$ and node impedances $\mathring{C}_i$ in Theorem 2.3 are strictly positive.

If G has no positive self-loops, than all the self-impedances χ_i are positive.

Using Property 2.4, we obtain from Theorem 2.3 the following corollary. This result will be used in order to write down linear qualitative equations.

Corollary 2.5
Suppose that

- *all influences of the parameter variations on the interior nodes of G are mediated by the boundary, i.e. $\frac{\partial X_i}{\partial P_k} = 0$, $\forall i \in \mathring{G}$;*
- *the restriction of the Jacobian of F to $\mathring{G}$ is non-singular;*
- *the interaction graph has no positive loops in the interior of G, for all the steady states, within the entire range of parameter variations.*

Then the signs of the finite variations of the internal variables X_i satisfy the following relation:

$$\forall i \in \mathring{G},\ s(\Delta X_i) = \sum_{j \in \urcorner^{in} G} \sum_{j \rightsquigarrow i \in \mathcal{P}_{\mathring{G}}} s(j \rightsquigarrow i) s(\Delta X_j). \tag{2.8}$$

3 Model and data compatibility

In this section we introduce two methods of qualitative analysis of biological systems and experimental data. These methods use the systems of qualitative equations generated by Theorem 2.2.

Nevertheless, Corollary 2.5 can be used to generate other systems of qualitative equations. Indeed, it is enough to consider any covering of the network I by subgraphs $\{G_j\}_{j=1,p}$. Then, the Eq. 2.8 written for all nodes $i \in G_j$ and for all subgraphs G_j provide another system of qualitative equations.

Hence, even if algorithms for incompatibility are presented for system 1 in this section, they can be easily adapted for another system of qualitative equations as well.

3.1 Solving qualitative systems

Let us recall the qualitative equations provided by Theorem 2.2

System 1

```
(1)  PPAR-a = PPAR + PUFA
(2)  LXR-a = -PUFA + LXR
(3)  SREBP = LXR-a
(4)  SREBP-a= SREBP + SCAP -PUFA
(5)  ACL= LXR-a  + SREBP-a - PUFA
(6)  ACC= LXR-a + SREBP-a - PUFA
(7)  FAS=  LXR-a + SREBP-a - PUFA
(8)  SCD1 = LXR-a  + SREBP-a - PUFA + PPAR-a
(9)  SCAP=  -PUFA

(10) D5D= PPAR-a + SREBP-a - PUFA
(11) D6D = PPAR-a + SREBP-a - PUFA
PPAR=+, PUFA=+, LXR=+, SREBP=-, SCD1=-,
FAS=-, ACL=+, ACC=-, D5D=0, D6D=0
```

Compatible system A valuation of the qualitative system is called compatible if it satisfies:

1. all nodes are determined (have values different from ?);
2. all the observed nodes are given the observed values;
3. the qualitative Eqs. 2.4 are fulfilled.

Solving the qualitative system means finding all compatible valuations. If there is no compatible valuation, the system is said to be incompatible.

The interest of finding compatible valuations lies in the following property: any solution of the quantitative system provides a compatible valuation. Consequently, the incompatibility of the qualitative system implies that the quantitative system is incompatible as well. This immediately indicates a contradiction between the model and the observations. Furthermore, the signs corresponding to a solution of the quantitative system should be given by one of the solutions of the qualitative system.

Let us notice that when the system is compatible, the number of compatible valuations can be big and this could bring very limited information on the biological system. In general we would like to know which kind of information one can extract from compatible valuations. In order to do this an useful concept is the following notion of competition.

Competitions We say that there is a competition or a balance on a node i if the right side of the Eq.2.4 is equal to "?". The biological interpretation of this indetermination is that the node i is submitted to competitive actions of different signs. An example is provided by the node LXR-a in the working example, that satisfies

$$\begin{array}{lcl} \text{PUFA} & = & + \\ \text{LXR} & = & + \\ \text{LXR-a} & = & -\,\text{PUFA} + \text{LXR} \quad = \quad ?. \end{array}$$

Determining the sign of the balanced node says which paths win the competition. If the variation of the balanced node is zero we say that the competition is neutral. If this variation is negative (positive) we say that the negative (positive) paths win the competition, or that the competition leans toward the negative (positive) direction. The information on competitions has biological interest, because it replaces undetermined actions by determined ones and facilitates the understanding of the interaction graph.

In the next subsections we use simple rules of qualitative algebra to solve System 1 and extract information on competitions. Then, we propose an algorithmic approach to solve the system obtained from Theorem 2.2 by using these rules systematically. Algorithmic methods related to competitions are detailed in Section 4.

3.2 Hand solving of qualitative equations

Let us solve System 1 by using the rules of the sign algebra and the following computation rule:

Property 3.1
If $A = X + B$ and if the sign of A is opposite to the sign of B, then the sign of X is determined and it is equal to the sign of A.

The result of the application of these rules to System 1 is summarized as following:

Property 3.2
System 1 is not compatible with the observed variations.

Simple change of the sign of ACL (ACL=−) renders it compatible and determined. The values of non-observed variables are PPAR-a=+, LXR-a=−, SCAP=−, SREBP-a=−.

In this case there are competitions on LXR-a, SCD1 that are observed to lean towards the negative directions. There are also competitions on D5D, D6D that are observed to be neutral. There is no competition on FAS, ACC and ACL.

Proof From Eq.1 PPAR-a=+. From Eq.3 LXR-a=SREBP=−. From Eq.9 SCAP=−. From Eq.4 SREBP-a=−.

We have already some biologically interesting results: by Eq.2, LXR-a is supposed to be undetermined, that is, the result of a competition between the positive action of LXR and the negative action of PUFA. Using Eq.3 we have shown that the competition on LXR-a should lean toward the negative direction. From Eq.8 it follows that there is a competition on SCD1. Data on SCD1 implies that this competition leans toward the negative direction.

From Eq.5 and Property 3.1, SREBP-a=+. This is inconsistent with the previously found value, which proves that the system is incompatible. There is necessarily an error, either in the interactions or in the data. Let us assume that the interactions are correct. Then, considering that data on ACL is wrong is the simplest way to make the system compatible (because Eqs.6 and 7 do not force signs of FAS and ACC).

With the correction ACL=−, the unknown values are determined in the same way as before. From Eq.1 PPAR-a=+. From Eq.3 LXR-a=SREBP=−. From Eq.9 SCAP=−. From Eq.4 SREBP-a=−. Now Eq.5 is satisfied with no competition. The same is true for Eqs.6,7: there are no competitions on ACC and FAS. From Eqs.2,8 we find that there are competitions on LXR-a,SCD1 which leans towards the negative direction. From Eqs.10,11 we find that there are competitions on D5D, D6D. The observed variations of D5D, D6D are vanishing, therefore these competitions are neutral.

■

Manual resolution of a qualitative system is only possible for small examples. In the next section we describe an algorithm that allows the systematic application of these rules to large systems.

3.3 Graph valuation algorithm

Basic ideas underlying the algorithm A natural way to reduce the number of unknowns in a system of equations is to proceed by elimination. This means to compute the value of one unknown from an equation and then substitute this value in all the other equations. Unfortunately, in the sign algebra, the usual elimination rules do not apply (qualitative equality of undetermined quantities is not transitive).

In the proposed algorithm, we will rely on the rules used in Section 3.2. Let us formulate them more precisely.

- Eq. 2.4 implies that the sign of ΔX_i is determined if $s(k,i)s(\Delta X_k)$ are determined and have the same value for all the predecessors k of the node i. This first elementary rule is used in the forward propagation step of the algorithm.

- The second rule follows from Property 3.3. Suppose that $\Delta X_i = s(k,i)s(\Delta X_k) + \sum_{l \neq k} s(l,i)s(\Delta X_l)$. If $s(\Delta X_i)$ and the sign of $\sum_{l \neq k} s(l,i)s(\Delta X_l)$ are determined and are opposite then $s(k,i)s(\Delta X_k) = s(\Delta X_i)$. The second rule is used in the backward propagation step of the algorithm.

In the case of a sign algebra with 0 value, the first rule is unchanged while the second one needs a straightforward adaptation.

The graph valuation algorithm

1. *Initialization.* All the nodes are given a value in $\{+,-,?\}$. Observed nodes are initialized with the value corresponding to the observed variations of the concentration of the molecules. The other nodes are initialized with the value ?. The value associated with a node i is denoted $val(i)$.

   Let us consider two sets of nodes U and T. The nodes with values in $\{+,-\}$ are put in T. The remaining nodes (with value "?") are put in U.

2. *Forward propagation*:

   Create a set F that contains every node of U that is also a successor of a node in T.

   Repeat: Find a node i in F which has at least a predecessor and such that for all predecessors j of i, $s(j,i)val(j) = +$, or for all predecessors j of i, $s(j,i)val(j) = -$. Then

   - set $val(i) = s(j,i)val(j)$,
   - remove i from F,
   - remove the successors of i which are in U and put them in F
   - put i in T.

   until no node found

3. *Backward propagation.*

   Create a set B that contains every node of U that is also a predecessor of a node in T.

   Repeat: Find a node j in B which has a successor i in T, such that such that $val(i) \times \sum_{k \in pred(i), k \neq j} s(k,i)val(k) = -$ or $\sum_{k \in pred(i), k \neq j} s(k,i)val(k) = 0$. Then

   - $val(j) = s(j,i)val(i)$,
   - remove j from B,
   - remove the predecessors of ^{j} which are in U and put them in B,
   - put j in T.

   until no node found

4. *Alternate propagations.* Alternate forward and backward propagations until no new value is added to T.

Notice that each time that a node is valuated, it is moved to T and it never returns to F or B. T is an increasing subset of a finite set, so that the algorithm stops.

Application to the detection of incompatibilities All the valuations introduced by the algorithm are unique sign choices compatible with the equations provided by Theorem 2.2. By doing so, no compatible valuation is lost. More precisely:

Property 3.3
The graph valuation algorithm provides a new graph such that any compatible valuation of the nodes in the initial graph is a compatible valuation of the new graph.

Nevertheless, nothing guarantees that a compatible valuation exists, i.e. that all Eqs. 2.4 are satisfied. The resulting graph may contain incompatibilities. These incompatibilities are detected simply by checking Eq. 2.4 on each node.

Criterion to compare model and data 1
If the graph output by the algorithm described below contains a node where Eq. 2.4 is not satisfied, then experimental data and the model are incompatible.

Working example In the example M the initialization is PUFA=+, LXR=+, LXR-a=?, PPAR=+, PPAR-a=?, SREBP=-, SREBP-a=?, ACL=+, ACC=+, FAS=-, SCD1=-, SCAP=?, D5D=?, D6D=?. We assume that the variations of D5D and D6D are not known.

The forward propagation begins with F={ PPAR-a, LXR-a, SREBP-a, SCAP, D6D, D5D } and ends with $F = \emptyset$, PPAR-a=+ (from PUFA=+ and PPAR=+), SREBP-a=- (from SREBP=-) and SCAP=- (from PUFA=+). D5D and D6D are removed from F and keep their value ?.

The backward propagation begins with B={ LXR-a } and ends with $B = \emptyset$, LXR-a=- (from SREBP=-),

or LXR-a=+ (from ACL=+ , SREBP-a=- and PUFA=+). The second forward propagation begins with F={D6D, D5D} and stops with $F = \emptyset$ but no new value for D5D or D6D.

After running the algorithm, if LXR-a=- equation (5) is not satisfied at node ACL and if LXR-a=+ equation (3) is not satisfied at node SREBP. This shows a contradiction between the model and the values assigned to some nodes by experimental data or hypothesis.

In this example, Eq.2.4 is not satisfied in the nodes SREBP or ACL. The algorithm could be easily improved in order to keep track of the nodes involved in application of the rules used to determinate the values of LXR-a and SREBP. In this case, the nodes are ACL, SREBP-a, SREBP and LXR-a. The nodes ACL and SREBP are observed and that incompatibility involves LXR-a and SREBP-a. The observations or the model may be wrong. In the later case, a modeling error should be tracked first in the subgraph containing the nodes involved in the incompatible valuation.

A similar computation can be done assuming D6D = D5D=0. It leads to the same contradictions.

Improvements The graph valuation algorithm is not complete for the resolution of linear qualitative systems. As a consequence, the algorithm concludes only when it detects an incompatibility or when all the values are determined. The following example shows a system for which an incompatibility exists, but the graph valuation algorithm does not detect it. The elimination of the variable x is possible between the two first equations ([TMD03]) and leads to an incompatibility for the sign of $a + b$.

$$\begin{array}{rcl} a & = & -x \ + \ y \\ b & = & x \ + \ y \\ w & = & a \ + \ b \\ x = +, \quad y = +, & & w = - \end{array}$$

The design of an improved algorithm based on value propagation and variable elimination is under consideration.

4 Assessment of competitions

The graph valuation algorithm aims to detect incompatibilities or to produce compatible valuations. Nevertheless, even when these tasks are successful the result may be difficult to use. If for instance data is wrong for a node that happens to be very connected to the rest of the graph, then many incompatibilities appear and it is difficult to localize and to correct the error. Also, if there are not enough observations then compatible valuations are produced in huge number. It is difficult to extract useful biological information from this situation.

In this section, we develop another approach, that exploits the global qualitative information on the propagation of influences on the graph. This approach is complementary to the graph valuation algorithm since it identifies competition processes and the places where these occur.

Competitions As already explained, we say that there is a competition or a balance on a node i if the right side of the Eq.2.4 is equal to "?". The biological interpretation of this indetermination is that the node i is submitted to competitive actions of different signs.

This definition can be extended to any equation connecting i to other nodes in the graph instead of Eq.2.4 (related to the predecessors of a node i). Such equations are provided by Corollary 2.5 that can be applied to every subgraph G containing i. Then, we say that there is a competition on a node i when there exists an equation whose left side is $s(\delta X_i)$ and whose right side is equal to "?".

Set of equations describing the sign of variation of a variable X_{i_0} Let $M = (I \cup \{\mathcal{E}\}, A, s)$ be a network with no positive loop in the interior of I and such that the sign function s is constant within the entire range of parameter variations. Let us focus on a variable X_{i_0} with index i_0 in the model M. From Corollary 2.5, as soon as X_{i_0} is an internal variable of a subgraph G of M, the sign of variation of X_{i_0} can be expressed as a combination of the signs of paths starting on the boundary of G and never coming back to it:

$$\forall G \subset I, X_{i_0} \in \mathring{G}, \ s(\Delta X_{i_0}) = \sum_{\substack{j \in \neg^{in} G \\ j \rightsquigarrow i_0 \in \mathcal{P}_{\mathring{G}}}} s(j \rightsquigarrow i_0) s(\Delta X_j). \tag{4.9}$$

The *set of equations describing the variable* X_{i_0} is the set of equations (4.9) for all subgraphs G such that X_{i_0} is an internal variable of G.

Example of SREBP In the working example of lipogenesis, let us consider the variable X = SREBP. We want to determine all the subgraphs G that contain X as an internal variable. A necessary and sufficient condition is that G contains all the predecessors

$$s(\Delta\text{SREBP}) \underset{(-)}{=} s(P_1)s(\Delta\text{LXR-a}) \tag{4.1}$$

$$s(\Delta\text{SREBP}) \underset{(-)}{=} s(P_2)s(\Delta\text{PUFA})_{(-)} + s(P_3)s(\Delta\text{LXR})_{(+)}. \tag{4.2}$$

$$s(\Delta\text{SREBP-a}) = s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_5)s(\Delta\text{SREBP})_{(-)} + s(P_7)s(\Delta\text{PUFA})_{(-)} \tag{4.3}$$

$$s(\Delta\text{SREBP-a}) = s(P_8)s(\Delta\text{LXR-a}) + s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_7)s(\Delta\text{PUFA})_{(-)} \tag{4.4}$$

$$s(\Delta\text{SREBP-a}) = s(P_9)s(\Delta\text{LXR})_{(+)} + s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_6)s(\Delta\text{PUFA})_{(-)} + s(P_7)s(\Delta\text{PUFA})_{(-)} \tag{4.5}$$

$$s(\Delta\text{SREBP-a}) = s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_5)s(\Delta\text{SREBP})_{(-)} + s(P_{10})s(\Delta\text{SCAP}) \tag{4.6}$$

$$s(\Delta\text{SREBP-a}) = s(P_8)s(\Delta\text{LXR-a}) + s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_{10})s(\Delta\text{SCAP}) \tag{4.7}$$

$$s(\Delta\text{SREBP-a}) = s(P_9)s(\Delta\text{LXR})_{(+)} + s(P_4)s(\Delta\text{PUFA})_{(-)} + s(P_6)s(\Delta\text{PUFA})_{(-)} + s(P_{10})s(\Delta\text{SCAP}) \tag{4.8}$$

Table 4.1: System of qualitative equations describing the variations of SREBP and SREBP-a; sign of influences associated to each path.

of X, that is, LXR-a. In each subgraph G, SREBP is determined by all the paths that start from the boundary of G and never go back to it. Notice that any path arriving in SREBP contains LXR-a. If it is longer than the path LXR-a $\rightarrow$ SREBP, it must contain PUFA or LXR, that both belong to the boundary of $\mathring{G}$ (since they are connected to the exterior). Hence, paths to SREBP that start from the boundary of such subgraphs G and never come back to it are

- (P_1) LXR-a$\rightarrow$SREBP,
- (P_2) PUFA $\rightarrow$ LXR-a $\rightarrow$ SREBP,
- (P_3) LXR$\rightarrow$ LXR-a$\rightarrow$SREBP.

The set of equations associated with SREBP contains two equations shown in table 4.1.

Paths appearing in an equation (4.9)
Let $\mathcal{P}(M)$ contain all paths of M that appear in an equation (4.9). Hence, a path $j \rightsquigarrow i$ belongs to $\mathcal{P}(M)$ if there exists a subgraph $G \subset I$ such that j belongs to the boundary of G and $j \rightsquigarrow i$ is an internal path of G. The reasoning detailed in the example allows to characterize this set. A full proof is given in the Appendix.

Property 4.1
A path $j \rightsquigarrow i \subset I$ belongs to $\mathcal{P}(M)$ iff $pred(j) \neq \emptyset$ and no node of $j \rightsquigarrow i$ except possibly j is connected to the exterior $\mathcal{E}$.

If so, there exists a subgraph G such that the variations of X_i satisfy:

$$s(\Delta X_i) = s(\Delta X_j)s(j \rightsquigarrow i) + \sum_{\substack{k \in \neg^{in}G \\ k \rightsquigarrow i \in \mathcal{P}_{\mathring{G}} \\ k \rightsquigarrow i \neq j \rightsquigarrow i}} s(k \rightsquigarrow i)s(\Delta X_k).$$

Example of SREBP-a Let us fix X =SREBP-a. As an application of Property 4.1 paths that start from the boundary of a graph G such that $X \in \mathring{G}$ and the path never come back to the boundary are

- (P_4) PUFA$\rightarrow$SREBP-a,
- (P_5) SREBP$\rightarrow$SREBP-a,
- (P_6) PUFA$\rightarrow$ LXR-a$\rightarrow$SREBP$\rightarrow$SREBP-a,
- (P_7) PUFA$\rightarrow$SCAP$\rightarrow$SREBP-a,
- (P_8) LXR-a$\rightarrow$SREBP$\rightarrow$SREBP-a,
- (P_9) LXR$\rightarrow$ LXR-a$\rightarrow$SREBP$\rightarrow$SREBP-a.
- (P_{10}) SCAP$\rightarrow$SREBP-a.

The equations describing SREBP-a are given in Table 4.1. Notice that Eq. 4.6, 4.7 and 4.8 correspond to eq. 4.3, 4.4 and 4.5 in which $s(P_7)s(\Delta\text{PUFA})$ has been replaced by $s(P_{10})s(\Delta\text{SCAP})$. This corresponds to the particular case of SCAP, whose only predecessor is PUFA, so that SCAP can be considered either on the boundary or in the interior of a subgraph G, providing two different equations.

Influences Property 4.1 suggests the following definition. For a path $j \rightsquigarrow i \in \mathcal{P}(M)$ we say that

$s(\Delta X_j)s(j \rightsquigarrow i)$ is the *influence of $j \rightsquigarrow i$ on the variable X_i*.

Let us also define the sets of positive and negative influences acting on X_i, denoted by $\mathcal{I}^+(i)$ and $\mathcal{I}^-(i)$. The set of all influences is denoted by $\mathcal{I}(i)$.

$\mathcal{I}^+(i) = \{j \rightsquigarrow i \in \mathcal{P}(M), j \in \mathcal{O},\ s(\Delta X_j)s(j \rightsquigarrow i) = +\}$
$\mathcal{I}^-(i) = \{j \rightsquigarrow i \in \mathcal{P}(M), j \in \mathcal{O},\ s(\Delta X_j)s(j \rightsquigarrow i) = -\}$
$\mathcal{I}(i) = \mathcal{I}^+(i) \cup \mathcal{I}^-(i)$.

Example In the examples of SREBP and SREBP-a, the observed variations are ΔPUFA= +, ΔSREBP = + and ΔLXR= +, so that the positive and the negative influences are

$$\begin{array}{lcl} \mathcal{I}^+(\text{SREBP}) & = & \{P_3\} \\ \mathcal{I}^-(\text{SREBP}) & = & \{P_2\} \\ \mathcal{I}^+(\text{SREBP-a}) & = & \{P_9\} \\ \mathcal{I}^-(\text{SREBP-a}) & = & \{P_4, P_5, P_6, P_7\}. \end{array}$$

The influence of paths starting from LXR-a and SCAP is not taken into account since the variations of these molecules are not observed.

Counterbalanced influences In Table 4.1, the right sides of Eqs. 4.2, 4.3 and 4.5 are undetermined since they are the sums of at least two terms of different signs. This means that there are competitions on SREBP and SREBP-a. The observation of SREBP shows that the competition between P_2 and P_3 leans toward the negative direction: we say that the influence of P_3 is counterbalanced by the influence of P_2. Since SREBP-a is not observed, we can say that the influence of P_9 is counterbalanced by the influence of P_4, P_5, P_6, P_7, but we do not know the inclination of the competition.

This example justifies the following:

Definition 4.2
The set of counterbalanced influences *on i is denoted by $\mathcal{C}(i)$. It is defined as follows:*

- *if i is observed, the influence of $j \rightsquigarrow i$ is counterbalanced if it is different from the variation of X_i, i.e. if $s(\Delta X_j)s(j \rightsquigarrow i) \neq s(\Delta X_i)$. Then*
  if $s(\Delta X_i) = +$ then $\mathcal{C}(i) = \mathcal{I}^-(i)$
  if $s(\Delta X_i) = -$ then $\mathcal{C}(i) = \mathcal{I}^+(i)$
  if $s(\Delta X_i) = 0$ then $\mathcal{C}(i) = \mathcal{I}(i)$.
- *If i is not observed, the influence of $j \rightsquigarrow i$ is counterbalanced if there exists another path $k \rightsquigarrow i$ whose influence on X_i has a different sign, i.e. if $\mathcal{I}^+(i)$ and $\mathcal{I}^-(i)$ are both nonempty. Then*

if $i \notin \mathcal{O}$, $\mathcal{I}^+(i) \neq \emptyset$, $\mathcal{I}^+(i) \neq \emptyset$ then $\mathcal{C}(i) = \mathcal{I}(i)$
if $i \notin \mathcal{O}$, $\mathcal{I}^+(i) = \emptyset$ or $\mathcal{I}^-(i) = \emptyset$ then $\mathcal{C}(i) = \emptyset$.

A direct consequence of the definition is that an influence which is not counterbalanced absorbs all the influences in the equations it appears in, so that it does not introduce indetermination's:

Property 4.3
If a path $j \rightsquigarrow i$ is not counterbalanced then the sign of its influence on the node i is the same as the sign of ΔX_i.

To summarize, influences are counterbalanced when they compete with other influences. When we can observe which is the result of this competition, the counterbalanced influence is the dominated one. Influences are not counterbalanced either if they are non-competing, or if they dominate their competitors.

Example of SREBP-a and SREBP In the example of SREBP-a, positive influences $\mathcal{I}^+$(SREBP-a) $= \{P_9\}$ compete with negative ones $\mathcal{I}^-$(SREBP-a) $= \{P_4, P_5, P_6, P_7\}$ for determining the unknown sign of the variation of SREBP-a. All these influences are considered to be counterbalanced:

$$\begin{array}{rcl} \mathcal{C}(\text{SREBP-a}) & = & \mathcal{I}(\text{SREBP-a}) \\ & = & \{P_9, P_4, P_5, P_6, P_7\}. \end{array}$$

The observed negative variation of SREBP implies that positive influences are dominated, hence counterbalanced:

$$\mathcal{C}(\text{SREBP}) = \mathcal{I}^+(\text{SREBP}) = \{P_3\}.$$

Redundant and essential influences Let us consider the influences of P_1 and P_2 on SREBP. The observed sign of SREBP implies that P_1 is dominated (thus counterbalanced) by P_2. With no loss, we can skip P_1 in our analysis of influences. We can not skip P_2. We say that P_2 is *essential*, while P_1 is *redundant*. Let us now consider the influences of P_9 and P_6 on SREBP-a. These two influences are competing, but we do not know which one wins. We decide to consider that both are redundant.

Generally, we call essential those influences that can not be skipped without loss. The discussion of the examples justifies the following:

Definition 4.4
A path is said to have an essential influence *on a node i if it belongs to the set $\mathcal{E}(i)$ defined as follows:*

$$\mathcal{E}(i) = \{j \rightsquigarrow i = (j_0 = j, j_1, \ldots, j_l = i) \in \mathcal{I}(i), \forall 0 < k < l, j \rightsquigarrow j_k \notin \mathcal{C}(k)\}.$$

Essentially balanced nodes Let us consider LXR-a. It collects two essential influences of different signs (from PUFA and LXR). We say that LXR-a is *essentially balanced.*

Let us also consider SREBP. This node collects two competing influences: P_1 and P_2. Nevertheless, we know that P_1 is dominated, hence redundant. We say that SREBP is *not essentially balanced.*

More generally, we have the following:

Definition 4.5
An observed variable X_i is said to be essentially balanced *if it collects essential influences that are also counterbalanced, i.e. if $\mathcal{E}(i) \cap \mathcal{C}(i) \neq \emptyset$.*

A non observed variable X_i is said to be essentially balanced *if it collects essential influences with opposite signs, that is $\mathcal{E}(i) \cap \mathcal{I}^+(i) \neq \emptyset$ and $\mathcal{E}(i) \cap \mathcal{I}^-(i) \neq \emptyset$.*

An essential balance on a variable X_i gathers competing influences such that no information allows to say that one of the competing influences has been previously absorbed in a previous competition. This implies that the essential balance is localized precisely at the node i. The variation of X_i gives the inclination of the competition.

From the above discussion it follows:

Criterion to compare model and data 2
Let us consider a model of molecular influences. Let us consider that the interaction graph contains no positive loop made of nodes that are not connected to the exterior. Let X_i be a variable that collects essential influences, i.e. $\mathcal{E}(i) \neq \emptyset$. Then several situations are possible:

1. *X_i is essentially balanced. Then there is a non-redundant competition process localized at i. If X_i is observed, we can say which one of the competing influences is dominating.*
2. *X_i is not essentially balanced and all the collected essential influences have the same sign.*

   If X_i is observed then:

   (a) *if the sign of ΔX_i coincides with the one of the essential influences then we can say nothing.*

   (b) *if the sign of ΔX_i is different from the one of the essential influences then there is probably an error in data.*

Complementarity with the graph valuation algorithm As already stated, the assessment of competitions is not devoted to proving incompatibilities. For instance the case 2b) of the Criterion 2 suggests, but does not prove the presence of an error. Its utility is not obvious in our simple working example, but should appear from its application to more complex networks. Case 2b) of Criterion 2 is a hint for the location of the error.

Case 1) of Criterion 2 is obviously useful as it allows to localize non-redundant competitions. These non-redundant competitions are usually important regulation checkpoints.

Working example and biological discussion
Computation on the graph related to the working examples M of regulation of lipid synthesis showed that there are competitions on 9 nodes among 11: LXR-a, SREBP, SREBP-a, ACL, ACC, FAS, SCD1, D5D, D6D. However, some of these competitions are redundant. Among the 32 counterbalanced paths, 20 are essential. Essentially balanced nodes are LXR-a, ACL and SCD1, D5D and D6D.

- LXR-a collects non-redundant competitive paths.
- ACL collects four essential negative influences and no essential positive influence. This suggests that the data on ACL is false.

  The system has to be studied more precisely, in order to check whether there is a real incompatibility between the model and the data. A way to do this is by using the methods of the previous section. An alternative way is a refined assessment of competitions (see the paragraph *improvements*).
- SCD1, D5D and D6D are observed nodes that collect essential negative and positive influences. This suggests that a non-redundant competition occurs on each of these nodes.

  The observed negative variation on SCD1 suggests that the negative path through SREBP dominates the positive path through PPAR.

Consequently, two points can be extracted from such an analysis: at the SCD1 level, the negative pathway containing SREBP dominates the positive pathway containing PPAR. At the ACL level, the data on ACL should be checked carefully or paths in the model leading to ACL should be studied in detail.

Essential balance algorithm Whether i is an essentially balanced node or not, it depends on competitions collected by the predecessors of i. An algorithmic computation of their number can be done as soon as the interaction graph is acyclic.

Let $I^+(i)$, $I^-(i)$ denote the number of paths in $\mathcal{I}^+(i)$ and $\mathcal{I}^-(i)$; $C^+(i)$, $C^-(i)$ denote the number of counterbalanced influences in $\mathcal{C}^+(i)$ and $\mathcal{C}^-(i)$; and $E^+(i)$, $E^-(i)$ denote the number of essential influences in $\mathcal{E}(i) \cap \mathcal{I}^+(i)$ and $\mathcal{E}(i) \cap \mathcal{I}^-(i)$. Let also p^+ and $p-$ denote the experimental information.

Initialization For every observed node on the boundary, let $I^+(i) = I^-(i) = C^+(i) = C^-(i) = E^+(i) = E^-(i) = 0$. For every observed node in G, let $p^+(i) = 1$ iff i is observed and $s(\delta X_i) = +$, $p^+(i) = 0$ otherwise. Let $p^-(i) = 1$ iff i is observed and $s(\delta X_i) = -$, $p^-(i) = 0$ otherwise.

Propagation Since we suppose that the interaction graph contains no loop, we can suppose that all the functions I^+, I^-, C^+, C^-, E^+, E^- are computed for every predecessor of a node i. Then the functions are computed as following for the node i.

$I^+(i) = \sum_{k \in pred(i)} I^{s(k,i)}(k) + p^{s(k,i)}(k)$
$I^-(i) = \sum_{k \in pred(i)} I^{-s(k,i)}(k) + p^{-s(k,i)}(k)$
$C^+(i) = \sum_{k \in pred(i)} C^{s(k,i)}(k) + E^{s(k,i)}(k)$
$C^-(i) = \sum_{k \in pred(i)} C^{-s(k,i)}(k) + E^{-s(k,i)}(k)$
X_i observed, $s(\delta X_i) = \epsilon$

$$\begin{cases} E^{\epsilon}(i) = 0 \\ E^{-\epsilon}(i) = (I^{-\epsilon} - C^{-\epsilon})(i) \end{cases}$$

X_i nonobserved, $(I^+ - C^+)(I^- - C^-)(i) = 0$

$$\begin{cases} E^+(i) = 0 \\ E^-(i) = 0 \end{cases}$$

X_i nonobserved, $(I^+ - C^+)(I^- - C^-)(i) \neq 0$

$$\begin{cases} E^+(i) = (I^+ - C^+)(i) \\ E^-(i) = (I^- - C^-)(i). \end{cases}$$

Essentially balanced nodes can be identified by using the following:

Property 4.6
In a directed graph with no internal loop, a node i collects a essential balance if and only if $E^+(i) \neq 0$ or $E^-(i) \neq 0$, where $E^{+,-}$ are defined in the algorithm described below. The case 1) of Criterion 2 occurs when $(I^+ - C^+)(I^- - C^-)(i) \neq 0$, the case 2b) occurs when $(I^+ - C^+)(I^- - C^-)(i) = 0$.

Improvements Our discussion of competitions aims at identifing and localizing essential competitions. Let us describe modifications of the definitions that allow us to refine the analysis.

Definitions 4.2, 4.4 declare redundant influences that compete with others, even if the result of the competition is unknown. We have did so in order to localize essential competitions: once they occur somewhere, competitions are transported along the graph as non-essential.

With altered definitions we can identify and localize incompatibilities. Let us change the Definition 4.2 of counterbalanced influences by skipping its second part. This means deciding that influences of different signs on an unobserved node are not counterbalanced. Then, according to Definition 4.4, the redundant influences are only those that are dominated somewhere. Hence the case 2b) of the Criterion 2 identifies an incompatibility: it is impossible to act with non-dominated influences of even sign on a node and to obtain a different sign on the node. This is actually the case of the node ACL in our working example. This node collects non-dominated negative influences and one dominated positive influence.

Thus, by altering Definitions 4.2, 4.4 we can detect incompatilities, but we can no longer localize competitions. In order to satisfy both purposes we can introduce two types of redundancy, one suitable for incompatibility detection, the other for localizing competitions.

5 Remarks and conclusion

In this paper we have discussed how a qualitative theoretical model of a mixed gene and metabolic network can be confronted to DNA-microarray and metabolic data.

Using a linear approximation, we have obtained quantitative equations connecting the theoretical variations of genes and metabolic products to the changes of the external conditions. The linear approximation is valid for small variations.

These equations are analogous to the discrete Laplace equations on graphs that have been studied in the context of Markov chains [CY00, Soa94, CP03]. In order to calculate the variations we have to compute the matrix inverse of the Jacobian of a differential dynamical system. This matrix inverse is in fact the discrete Green function of the problem. Contrary to [CY00] we do not want explicit formulas for the Green function (difficult to obtain for irregular graphs), but we want to emphasize how one node collects influences from the other nodes. A similar path representation of the Green function is used in statistical mechanics [CP03].

Quantitative graph equations have lead us to systems of qualitative equations. The validity of the latter extend beyond the validity of the linear approximation, implying uniform signs of the interac-

tion coefficients (elements of the Jacobian) within the domain of variation of the experimental conditions. We introduce the graph valuation algorithm that can reduce or solve the qualitative system. The computation rules are not those of usual linear spaces, but belong to the sign algebra.

When the qualitative system is incompatible, it is generally difficult to propose corrections. When it is compatible, we may obtain many solutions. For this reason we have also introduced a complementary approach which is the study of competitions. This approach treats all nodes on equal footing. It does not prove incompatibilities, but identifies biologically important competition processes and the places where they occur.

The analysis of the examples taken from the regulation of lipogenesis suggests that this type of methods could be used to optimize experimental protocols. The optimization criteria are not formalized here. Nevertheless, as a general rule, if a node is observed we would like to have observations as complete as possible on its predecessors. In the working example, SCAP is not observed. Instead, we have predicted the sign of its variation. The experimental knowledge of SCAP variation would further facilitate the localization of the discovered incompatibility between interaction and data.

Acknowledgments This research was supported by the ACI IMPBio, French Ministry for Research program on interdisciplinarity. We would like to thank M. Crouzeix, I. Gruais, Y. Guivarch, D. Petritis, and C. Soulé for many inspiring discussions.

Appendix: Proofs

Proof of Theorem 2.3 From Eq. 2.6 it follows:

$$\delta X_i = -\sum_{j\in\mathring{G}} \mathring{B}_{ij} \sum_{k\in\neg^{in}G} \frac{\partial F_j}{\partial X_k}\delta X_k, \qquad (5.1)$$

where $\mathring{B} = \mathring{A}^{-1}$.

Let us consider the following well-known property of the inverse matrix [Blo79]:

$$\mathring{A}^{-1}_{ij} = (-1)^{i+j}\frac{det(M_{ji})}{det\mathring{A}} \qquad (5.2)$$

where M_{ij} is the minor obtained by deleting line i and column j in matrix $\mathring{A}$.

Furthermore, it is a simple exercise [Blo79] to show that:

$$M_{ji} = \begin{cases} (-1)^{i+j}\sum_{j\rightsquigarrow i}(-1)^{l_{j\rightsquigarrow i}}a_{j\rightsquigarrow i}det(\mathring{A}_{(j\rightsquigarrow i)^c}) \\ \qquad\qquad \text{if } i\neq j \\ det\mathring{A}_{\{j\}^c} \qquad \text{if } i=j. \end{cases} \qquad (5.3)$$

where $j \rightsquigarrow i$ is any path leading from j to i with no loop, $l_{j\rightsquigarrow i}$ is the length of the path, $a_{j\rightsquigarrow i}$ is the product of elements of $\mathring{A}$ along this path, $\mathring{A}_{(j\rightsquigarrow i)^c}$ is the principal minor defined by the set of indices complementary to those in the path. Conventionally, $det\mathring{A}_{\mathring{G}^c} = 1$ (determinant of the empty matrix).

Using Eqs.5.2,5.3 it follows

$$\begin{aligned}\delta X_i \;=\; & -\sum_{k\in\neg G}\frac{det\mathring{A}_{\{i\}^c}}{det\mathring{A}}\frac{\partial F_i}{\partial X_k}\delta X_k \\ & -\sum_{j\in\mathring{G},j\neq i}\sum_{j\rightsquigarrow i}\sum_{k\in\neg G} \\ & \qquad (-1)^{l_{j\rightsquigarrow i}}a_{j\rightsquigarrow i}\frac{det\mathring{A}_{\{j\rightsquigarrow i\}^c}}{det\mathring{A}}\frac{\partial F_j}{\partial X_k}\delta X_k.\end{aligned}$$

We can identify in the sum the contributions of direct (one arc) and indirect (at least two arcs) paths from the boundary to the interior. Thus, $\delta X_i = \sum_{k\in\neg^{in}G}\left[\sum_{k\rightsquigarrow i}^{\text{indirect}}\frac{a_{k\rightsquigarrow i}}{C_{k(j)\rightsquigarrow i}} + \sum_{k\rightsquigarrow i}^{\text{direct}}\frac{a(k,i)}{C_i}\right]\delta X_k$ from which it follows Eq.2.7. ■

Proof of Property 2.4 If there are no positive loops, then it can be shown (see Soulé [Sou03]) that the signs of the principal minors of the Jacobian $det\mathring{A}$ are alternating. Thus,

$$\begin{aligned} sign(det\mathring{A}_{\{j\}^c}) &= (-1)^{\#\mathring{G}-1} \\ sign(det\mathring{A}_{(j\rightsquigarrow i)^c}) &= (-1)^{\#\mathring{G}-l_{j\rightsquigarrow i}-1} \\ sign(\mathring{A}_{jj}) = sign(a_{jj}) &= -1 \end{aligned}$$

from which it follows that all the impedances are positive. ■

Proof of Property 4.1 If $P = j \rightsquigarrow i = (j, j_1, \ldots, i)$ contains an intermediate node j_k that is connected to the exterior, then for every subgraph G, j_k belongs to the boundary of G, hence $P \notin \mathcal{P}_{\mathring{G}}$.

Conversely, if P contains no intermediate node connected to the exterior $\mathcal{E}$ let $G = \{j_0 = j, j_1, \ldots, j_n = i\} \cup_{0<k\leq n} pred(j_k)\}$. Then $P \in \mathcal{P}_{\mathring{G}}$, $j \in \neg^{in}G$, hence $P \in \mathcal{P}(M)$.

In order to prove the second part of the Property, Corollary2.5 can be applied to G. ■

Proof of Property 4.6 The formulas given in the pure competition computation simply compute the number of influences and counterbalances by using the information provided by the predecessor of a node. If i is a node, then positive influences $I^+(i)$ are

the union of positive influences $I^+(k)$ on any predecessor of i with a positive edge, negative influences of any predecessor of i with a negative edge, and influences directly starting in k if it is an observed node.

Essential influences on i are given by counterbalances on each predecessor of i with the appropriate sign and competitions identified on these predecessors. Then one simply has to compare carried influences on i with the total number of influences to get essential influences. ■

References

[Blo79] DM. Bloom. *Linear algebra and geometry*. Cambridge University Press, Cambridge, 1979.

[CP03] M. Campanino and D. Petritis. Random walks in randomly oriented lattices. *Markov Processes and Related Fields*, 9:391–412, 2003.

[CY00] F. Chung and ST. Yau. Discrete green's functions. *J. Combin. Theory Ser.A*, 91:191–214, 2000.

[Dor88] J.L. Dormoy. Controlling qualitative resolution. In *Proceedings of the seventh National Conference on Artificial Intelligence, AAAI88', Saint-Paul, Minn.*, 1988.

[Jum04] DB Jump. Fatty acid regulation of gene transcription. *Crit. Rev. Clin. Lab. Sci.*, 41(1):41–78, 2004.

[KF02] N. Kaminski and N. Friedman. Practical approaches to analyzing results of microarray experiments. *American Journal of Respiratory and Cell Molecular Biology*, 27:125–13, 2002.

[Kui94] B.J. Kuipers. *Qualitative reasoning. Mdeling and simulation with incomplete knowledge*. MIT Press, 1994.

[LCL+04] SS Lee, WY Chan, CK Lo, DC Wan, DS Tsang, and WT Cheung. Requirement of pparalpha in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *J Lipid Res.*, 45(11):2025–37, 2004.

[LYH+02] G Liang, J Yang, JD Horton, RE Hammer, JL Goldstein, and MS. Brown. Diminished hepatic response to fasting/refeeding and liver x receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *J Biol Chem*, 277(15):9520–8, Jan 2002.

[MN96] CW Miller and JM Ntambi. Peroxisome proliferators induce mouse liver stearoyl-coa desaturase 1 gene expression. *Proc Natl Acad Sci U S A.*, 93(18):9443–8, 1996.

[MSY+02a] T Matsuzaka, H Shimano, N Yahagi, M Amemiya-Kudo, and alt. Dual regulation of mouse delta(5)- and delta(6)-desaturase gene expression by srebp-1 and pparalpha. *J Lipid Res.*, 43(1):107–14, 2002.

[MSY+02b] T Matsuzaka, H Shimano, N Yahagi, M Amemiya-Kudo, T Yoshikawa, AH Hasty, Y Tamura, J Osuga, H Okazaki, Y Iizuka, A Takahashi, H Sone, T Gotoda, S Ishibashi, and N Yamada. Dual regulation of mouse delta(5)- and delta(6)-desaturase gene expression by srebp-1 and pparalpha. *J Lipid Res*, 43(1):107–14, Jan 2002.

[NHT+02] TY Nara, WS He, C Tang, SD Clarke, and MT Nakamura. The e-box like sterol regulatory element mediates the suppression of human delta-6 desaturase gene by highly unsaturated fatty acids. *Biochem. Biophys. Res. Commun.*, 296(1):111–7, 2002.

[NRF04] I. Nachman, A. Regev, and N. Friedman. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, 20:i248 – i256, 2004.

[PSP+04] JA Papin, J Stelling, ND Price, S Klamt, S Schuster, and Palsson BO. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22:400–405, 2004.

[SG04] KR Steffensen and JA. Gustafsson. Putative metabolic effects of the liver x receptor (lxr). *Diabetes*, 53(Supp 1):36–52, Feb 2004.

[Soa94] P.M. Soardi. *Potential theory on infinite networks*. Springer, Berlin, Heidelberg, New York, 1994.

[Sou03] C. Soulé. Graphic requirements for multistationarity. *Complexus*, 1(123-133), 2003.

[TCNC03] C Tang, HP Cho, MT Nakamura, and S-D Clarke. Regulation of human delta-6 desaturase gene transcription: identification of a functional direct repeat-1 element. *J Lipid Res*, 44(4):686–95, 2003.

[TMD03] L. Travé-Massuyès and P. Dague, editors. *Modèles et raisonnements qualitatifs*. Hermes sciences, 2003.

[TSA+00] KA Tobin, HH Steineger, S Alberti, O Spydevold, J Auwerx, JA Gustafsson, and HI. Nebb. Cross-talk between fatty acid and cholesterol metabolism mediated by liver x receptor-alpha. *Mol Endocrinol*, 14(5):741–52, May 2000.

[YVK04] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*, 20:i363 – i370, 2004.