

Algorithmic and Complexity Results for Decompositions of Biological Networks into Monotone Subsystems

Bhaskar DasGupta^{1,5}, German A. Enciso^{2,6}, Eduardo Sontag^{3,7},
and Yi Zhang^{4,8}

¹ Department of Computer Science, University of IL at Chicago, Chicago, IL 60607
dasgupta@cs.uic.edu

² Mathematical Biosciences Institute, 250 Mathematics Building, 231 W 18th Ave,
Columbus, OH 43210
genciso@mbi.osu.edu

³ Department of Mathematics, Rutgers University, New Brunswick, NJ 08903
sontag@math.rutgers.edu

⁴ Department of Computer Science, University of IL at Chicago, Chicago, IL 60607
yzhang3@cs.uic.edu

Abstract. A useful approach to the mathematical analysis of large-scale biological networks is based upon their decompositions into monotone dynamical systems. This paper deals with two computational problems associated to finding decompositions which are optimal in an appropriate sense. In graph-theoretic language, the problems can be recast in terms of maximal sign-consistent subgraphs. The theoretical results include polynomial-time approximation algorithms as well as constant-ratio in-approximability results. One of the algorithms, which has a worst-case guarantee of 87.9% from optimality, is based on the semidefinite programming relaxation approach of Goemans-Williamson [14]. The algorithm was implemented and tested on a *Drosophila* segmentation network [7] and an Epidermal Growth Factor Receptor pathway model [25], and it was found to perform close to optimally.

1 Introduction

In living cells, networks of proteins, RNA, DNA, metabolites, and other species process environmental signals, control internal events such as gene expression, and produce appropriate cellular responses. The field of systems (molecular) biology is largely concerned with the study of such networks, viewed as dynamical systems. One approach to their mathematical analysis relies upon viewing them

⁵ Supported by NSF grants CCR-0206795, CCR-0208749 and IIS-0346973.

⁶ Work done while the author was with the Mathematics Department of Rutgers University and partly supported by NSF grant CCR-0206789.

⁷ Partly supported by NSF grants EIA 0205116 and DMS-0504557.

⁸ Partly supported by NSF grants CCR-0206795 and CCR-0208749.

as made up of subsystems whose behavior is simpler and easier to understand. Coupled with appropriate interconnection rules, the hope is that emergent properties of the complete system can be deduced from the understanding of these subsystems.

A particularly appealing class of candidates for “simpler behaved” subsystems are *monotone systems*, as in [16, 15, 29]. Monotone systems are a class of dynamical systems for which pathological behavior (“chaos”) is ruled out. Even though they may have arbitrarily large dimensionality, monotone systems behave in many ways like one-dimensional systems. For instance, in monotone systems, bounded trajectories generically converge to steady states, and there are no stable oscillatory behaviors. Monotonicity is closely related to positive and feedback loops in systems. The topic of analyzing the behaviors of such feedback loops is a long-standing one in biology in the context of regulation, metabolism, and development; a classical reference in that regard is the work [23] of Monod and Jacob in 1961. See also, for example, [20, 22, 34, 26, 31, 6, 4, 1, 27].

An interconnection of monotone subsystems, that is to say, an entire system made up of monotone components, may or may not be monotone: “positive feedback” (in a sense that can be made precise) preserves monotonicity, while “negative feedback” destroys it. Thus, oscillators such as circadian rhythm generators require negative feedback loops in order for periodic orbits to arise, and hence are not themselves monotone systems, although they can be decomposed into monotone subsystems (cf. [5]). A rich theory is beginning to arise, characterizing the behavior of non-monotone interconnections. For example, [3] shows how to preserve convergence to equilibria; see also the follow-up papers [2, 18, 12, 9, 13]. Even for monotone interconnections, the decomposition approach is very useful, as it permits locating and characterizing the stability of steady states based upon input/output behaviors of components, as described in [4]; see also the follow-up papers [1, 11, 19]. Moreover, a key point brought up in [32] is that new techniques for monotone systems in many situations allow one to characterize the behavior of an entire system, based upon the “qualitative” knowledge represented by general network topology and the inhibitory or activating character of interconnections, combined with only a relatively *small amount of quantitative* data. The latter data may consist of steady-state responses of components (dose-response curves and so forth), and there is no need to know the precise form of dynamics or parameters such as kinetic constants in order to obtain global stability conclusions.

Generally, a graph, whose edges are labeled by “+” or “−” signs (sometimes one writes $+1, -1$ instead of $+, -$, or uses respectively activating “ \rightarrow ” or inhibiting “ \dashrightarrow ” arrows), is said to be *sign-consistent* if all paths between any two nodes have the same net sign, or equivalently, all closed loops have positive parity, i.e. an even number, possibly zero, of negative edges. (For technical reasons, one ignores the direction of arrows, looking only at undirected graphs; see more details in Section 2.)

When applying decomposition theorems such as those described in [3, 4, 1, 32, 2, 18, 11, 19, 12, 9, 13], it tends to be the case that *the fewer the number*

of interconnections among components, the easier it is to obtain useful conclusions. One may view a decomposition into interconnections of monotone subsystems as the “pulling out” of “inconsistent” connections among monotone components, the original system being a “negative feedback” loop around an otherwise consistent system. In this interpretation, the number of interconnections among monotone components corresponds to the number of variables being fed-back. In addition, and independently from the theory developed in the above references, one might speculate that nature tends to favor systems that are decomposable into small monotone interconnections, since “negative” feedback loops, although required for homeostasis and for periodic behavior, have potentially destabilizing effects, especially if there are signal propagation delays. Some evidence is provided by work in progress such as [21], where the authors compare certain biological networks and appropriately randomized versions of them and show that the original networks are closer to being consistent, and by [28], where the authors show that, in a Boolean setting, and using a mean-field calculation of sensitivity, networks of Boolean functions behave in a more and more “orderly” fashion the closer that the components are to being monotone.

Thus, we are led to the subject of this paper, namely computing the smallest number of edges that have to be removed so that there remains a consistent graph. In this paper, we study the computational complexity of the question of how many edges must be removed in order to obtain consistency, and we provide a polynomial-time approximation algorithm guaranteed to solve the problem to about 87.9% of the optimum solution, which is based on the semidefinite programming relaxation approach of Goemans-Williamson [14] (A variant of the problem is discussed as well). We also observe that it is not possible to have a polynomial-time algorithm with performance too close to the optimal. While our emphasis is on theory, one of the algorithms was implemented, and we show results of its application to a *Drosophila* segmentation network and to an Epidermal Growth Factor Receptor pathway model. It turns out that, when applying the algorithm, often the solution is much closer to optimal than the worst-case guarantee of 87.9%, and indeed often gives an optimal solution.

2 Monotone Systems and Consistency

We will illustrate the motivation for the problem studied here using systems of ordinary differential equations

$$\dot{x} = F(x) \tag{1}$$

(the dot indicates time derivative, and $x = x(t)$ is a vector), although the discussion applies as well to more general types of dynamical systems such as delay-differential systems or certain systems of reaction-diffusion partial differential equations. In applications to biological networks, the component $x_i(t)$ of the vector $x = x(t)$ indicates the concentration of the i th species in the model at time t . We will restrict attention to models in which the direct effect that one given variable in the model has over another is either consistently inhibitory or consistently promoting. Thus, if protein A binds to the promoter region of

gene B, we assume that it does so either to consistently prevent the transcription of the gene or to consistently facilitate it. (Of course, this condition does not prevent protein A from having an indirect influence, through other molecules, perhaps dimmers of A itself, that can ultimately lead to the opposite effect on gene B.) Mathematically, we require that for every $i, j = 1 \dots n, i \neq j$, the partial derivative $\partial F_i / \partial x_j$ be either ≥ 0 at all states or ≤ 0 at all states.

Given any partial order \leq defined on \mathbb{R}^n , a system (1) is said to be *monotone with respect to* \leq if $x_0 \leq y_0$ implies $x(t) \leq y(t)$ for every $t \geq 0$. Here $x(t), y(t)$ are the solutions of (1) with initial conditions x_0, y_0 , respectively. Of course, whether a system is monotone or not depends on the partial order being considered, but we one says simply that a system is *monotone* if the order is clear from the context. Monotonicity with respect to nontrivial orders rules out chaotic attractors and even stable periodic orbits; see [16, 15, 29], and is, as discussed in the introduction, a useful property for components when analyzing larger systems in terms of subsystems.

A useful way to define partial orders in \mathbb{R}^n , and the only one to be further considered in this paper, is as follows. Given a tuple $s = (s_1, \dots, s_n)$, where $s_i \in \{1, -1\}$ for every i , we say that $x \leq_s y$ if $s_i x_i \leq s_i y_i$ for every i . For instance, the “cooperative order” is the orthant order \leq_s generated by $s = (1, \dots, 1)$. This is the order \leq defined by $x \leq y$ if and only if $x_i \leq y_i$ for all $i = 1, \dots, n$. It is not difficult to verify if a system is cooperative with respect to an orthant order; the following lemma, known as “Kamke’s condition,” is not hard to prove, see [29] for details (also [3] in the more general context of monotone systems with input and output channels).

Lemma 1. *Consider an orthant order \leq_s generated by $s = (s_1, \dots, s_n)$. A system (1) is monotone with respect to \leq_s if and only if*

$$s_i s_j \frac{\partial F_j}{\partial x_i} \geq 0, \quad i, j = 1 \dots n, \quad i \neq j. \quad (2)$$

An equivalent way to phrase this condition is to ask that $\partial F_i / \partial x_j \geq 0$ at all states for every $i, j, i \neq j$, which is the Kamke condition for the special case of the cooperative order. The name of the order arises because in a monotone system with respect to that order each species promotes or “cooperates” with each other.

A rephrasing of this characterization of monotonicity with respect to orthant orders can be given by looking at the *signed digraph* associated to (1) and defined as follows. Let $V(G) = \{1, \dots, n\}$. Given vertices i, j , let $(i, j) \in E(G)$ and $f_E(i, j) = 1$ if both $\partial F_j / \partial x_i \geq 0$ and the strict inequality holds at least at one state. Similarly let $(i, j) \in E(G)$ and $f_E(i, j) = -1$ if both $\partial F_j / \partial x_i \leq 0$ and the strict inequality holds at least at one state. Finally, let $(i, j) \notin E(G)$ if $\partial F_j / \partial x_i \equiv 0$. Recall that we are assuming that one of the three cases must hold. Now we can define an orthant cone using any function $f_V : V(G) \rightarrow \{-1, 1\}$, by letting $x \leq_{f_V} y$ if and only if $f_V(i)x_i \leq f_V(i)y_i$ for all i . Given f_V , we define the consistency function $g : E(G) \rightarrow \{\text{true}, \text{false}\}$ by $g(i, j) = f_V(i)f_V(j)f_E(i, j)$. Then, the following analog of Lemma 1 holds.

Lemma 2. Consider a system (1) and an orthant cone \leq_{f_V} . Then (1) is monotone with respect to \leq_{f_V} if and only if $g(i, j) \equiv 1$ on $E(G)$.

For the next lemma, let the *parity* of a chain in G be the product of the signs $(+1, -1)$ of its individual edges. We will consider in the next result closed *undirected chains*, that is, sequences $x_{i_1} \dots x_{i_r}$ such that $x_{i_1} = x_{i_r}$, and such that for every $\lambda = 1, \dots, r - 1$ either $(x_{i_\lambda}, x_{i_{\lambda+1}}) \in E(G)$ or $(x_{i_{\lambda+1}}, x_{i_\lambda}) \in E(G)$.

Lemma 3. Consider a dynamical system (1) with associated directed graph G . Then (1) is monotone with respect to some orthant order if and only if all closed undirected chains of G have parity 1.

2.1 Systems with Inputs and Outputs

As we discussed in the introduction, a useful approach to the analysis of biological networks consists of decomposing a given system into an interconnection of monotone subsystems. The formulation of the notion of interconnection requires subsystems to be endowed with “input and output channels” through which information is to be exchanged. In order to address this we consider *controlled* dynamical systems ([33], which are systems with an additional parameter $u \in \mathbb{R}^m$, and which have the form

$$\dot{x} = g(x, u). \tag{3}$$

The values of u over time are specified by means of a function $t \rightarrow u(t) \in \mathbb{R}^m$, $t \geq 0$, called an *input* or *control*. Thus each input defines a time-dependent dynamical system in the usual sense. To system (3) there is associated a *feedback function* $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, which is usually used to create the closed loop system $\dot{x} = g(x, h(x))$. Finally, if $\mathbb{R}^n, \mathbb{R}^m$ are ordered by orthant orders \leq_{f_V}, \leq_q respectively, we say that the system is monotone if it satisfies (2) for every u , and also

$$q_k f_V(j) \frac{\partial g_j}{\partial u_k} \geq 0, \text{ for every } k, j \tag{4}$$

(see also [3].) As an example, let us consider the following biological model of testosterone dynamics [10, 24]:

$$\begin{aligned} \dot{x}_1 &= \frac{A}{K + x_3} - b_1 x_1 \\ \dot{x}_2 &= c_1 x_1 - b_2 x_2 \\ \dot{x}_3 &= c_2 x_2 - b_3 x_3. \end{aligned} \tag{5}$$

Drawing the digraph of this system, it is easy to see that it is not monotone with respect to any orthant order, as follows by application of Lemma 3. On the other hand, replacing x_3 in the first equation by u , we obtain a system that is monotone with respect to the orders $\leq_{(1,1,1)}, \leq_{(-1)}$ for state and input respectively. Defining $h(x) = x_3$, the closed loop system of this controlled system is none other than (5). The paper [10] shows how, using this decomposition

together with the “small gain theorem” from monotone input/output theory ([3]) leads one to a proof that the system does not have oscillatory behavior, even under arbitrary delays in the feedback loop, contrary to the assertion made in [24]. We can carry out this procedure on an arbitrary system (1) with a directed graph G as follows: given a set E of edges in G , enumerate the edges in E^C as $(i_1, j_1), \dots, (i_m, j_m)$. For every $1 \leq k \leq m$, replace all appearances of x_{i_k} in the function F_{j_k} by the variable u_k to form the function $g(x, u)$. Define $h(x) = (x_{i_1}, \dots, x_{i_m})$. It is easy to see that this controlled system (3) has closed loop (1).

Let the set E be called *consistent* if the undirected subgraph of G generated by E has no closed chains with parity -1 . Note that this is equivalent to the existence of f_V such that $g \equiv 1$ on E , by Lemma 4 applied to the open loop system (3). If E is consistent, then the associated system (3) itself can also be shown to be monotone: to verify condition (4), simply define each q_k so that (4) is satisfied for k, j_k . Since $\partial g_{j_k} / \partial u_k = \partial F_{j_k} / \partial x_{i_k} \neq 0$, this choice is in fact unambiguous. Conversely, if (3) is monotone with respect to the orthant orders \leq_{f_V}, \leq_q , then in particular it is monotone for every fixed constant u , so that E is consistent by Lemma 3. We thus have the following result.

Lemma 4. *The set of edges E of the digraph G is consistent iff the corresponding controlled system (3) is monotone with respect to some orthant orders.*

3 Statement of Problem

A natural problem is therefore the following. Given a dynamical system (1) that admits a digraph G , use the procedure above to decompose it as the closed loop of a monotone controlled system (3), while minimizing the number $\|E^C\|$ of inputs. Equivalently, *find f_V such that $P(E_+) = \|E_+\|$ is maximized and $P(E_-) = \|E_-\| = \|E_+^C\|$ minimized.* This produces the following problem formulation.

Problem 1 (Undirected Labeling Problem(ULP))

An instance of this problem is (G, h) , where $G = (V, E)$ is an undirected graph and $h: E \mapsto \{0, 1\}$. A valid solution is a vertex labeling function $f: V \rightarrow \{0, 1\}$. Define an edge $\{u, v\} \in E$ to be consistent iff $h(u, v) \equiv (f(u) + f(v)) \pmod{2}$. The objective is then to find a valid solution maximizing $|F|$ where F is the set of consistent edges.

There is a second, slightly more sophisticated way of writing a system (1) as the feedback loop of a system (3) using an arbitrary set of edges E . Given any such E , define $S(E^c) = \{i \mid \text{there is some } j \text{ such that } (i, j) \in E^c\}$. Now enumerate $S(E^c)$ as $\{i_1, \dots, i_m\}$, and for each k label the set $\{j \mid (i_k, j) \in E^c\}$ as j_{k1}, j_{k2}, \dots . Then for each k, l , one can replace each appearance of x_{i_k} in $F_{j_{kl}}$ by u_k , to form the function $g(x, u)$. Then one lets $h(x) = (x_{i_1}, \dots, x_{i_m})$ as above. The closed loop of this system (3) is also (1) as before but with the advantage that there are $|S(E^c)|$ inputs, and of course $|S(E^c)| \leq |E^c|$.

If E is a consistent and *maximal* set, then one can make (3) into a monotone system as follows. By letting f_V be such that $g \equiv 1$ on E , we define the order \leq_{f_V} on \mathbb{R}^n . For every i_k, j_{kl} such that $(i_k, j_{kl}) \in E^C$, it must hold that $f_V(i_k)f_V(j_{kl})f_E(i_k, j_{kl}) = -1$. Otherwise $E \cup \{(i_k, j_{kl})\}$ would be consistent, thus violating maximality. By choosing $q_k = -f_V(i_k)$, equation (4) is therefore satisfied. Conversely, if the system generated by E using this second algorithm is monotone with respect to orthant orders, and if h is a negative function, then it is easy to verify that E must be both consistent and maximal.

Thus the problem of finding E consistent and such that $P(E_-) = \|S(E_-)\| = \|S(E^C)\|$ is smallest, when restricted to those sets that are maximal and consistent (this does not change the minimum $\|S(E^C)\|$), is equivalent to the following problem: decompose (1) into the negative feedback loop of an orthant monotone control system, using the second algorithm above, and using as few inputs as possible. This produces the following problem formulation.

Problem 2 (Directed Labeling Problem(DLP))

An instance of this problem is (G, h) where $G = (V, E)$ is a directed graph and $h: E \rightarrow \{0, 1\}$. A valid solution is a vertex labeling function $f: V \rightarrow \{0, 1\}$. Define an edge $(u, v) \in E$ to be consistent iff $h(u, v) \equiv (f(u) + f(v)) \pmod{2}$. The objective is then to find a valid solution minimizing $|g(E - F)|$ where $g(C) = \{u \in V \mid \exists y \in V, (u, y) \in C\}$ for any $C \subseteq E$ and F is the set of consistent edges.

4 Theoretical Results

Theorem 5

- (a) For some constant $\varepsilon > 0$, it is not possible to approximate in polynomial time the ULP and the DLP problems to within an approximation ratio of $1 - \varepsilon$ and $1 + \varepsilon$, respectively, unless $P=NP$.
- (b) For ULP, we provide a polynomial time α -approximation algorithm where $\alpha \approx 0.87856$ is the approximation factor for the MAX-CUT problem obtained in [14] via semidefinite programming.
- (c) For DLP, if d_{in}^{max} denotes the maximum in-degree of any vertex in the graph, then we give a polynomial-time approximation algorithm with an approximation ratio of at most $d_{in}^{max} \cdot O(\log |V|)$.

5 Two Examples of Applications of the ULP Algorithm

5.1 Drosophila Segment Polarity

An important part of the development of the early Drosophila (fruit fly) embryo is the differentiation of cells into several stripes (or *segments*), each of which eventually gives rise to an identifiable part of the body such as the head, the wings, the abdomen, etc. Each segment then differentiates into a posterior and an anterior part, in which case the segment is said to be *polarized*. (This differentiation process continues up to the point when all identifiable tissues of the fruit

fly have developed.) Differentiation at this level starts with differing concentrations of certain key proteins in the cells; these proteins form striped patterns by reacting with each other and by diffusion through the cell membranes.

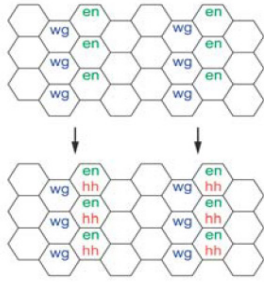


Fig. 1. A digram of the Drosophila embryo during early development. A part of the segment polarization process is displayed. Courtesy of N. Ingolia and PLoS [17].

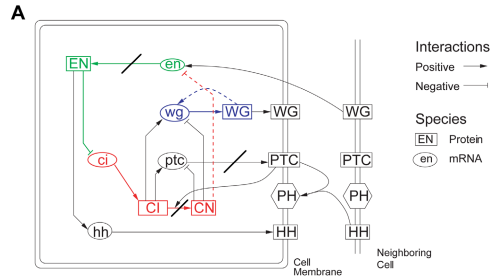


Fig. 2. The network associated to the Drosophila segment polarity, as proposed in [7], Courtesy of N. Ingolia and PLoS. The three edges that have been crossed have been chosen in order to let the remaining edges form an orthant monotone system.

A model for the network that is responsible for segment polarity [7] is illustrated in Figure 2. As explained above, this model is best studied when multiple cells are present interacting with each other. But it is interesting at the one-cell level in its own right — and difficult enough to study that analytic tools seem mostly unavailable. The arrows with a blunt end are interpreted as having a negative sign in our notation. Furthermore, the concentrations of the membrane-bound and inter-cell traveling compounds PTC, PH, HH and WG(membrane) on all cells have been identified in the one-cell model (so that, say, $HH \rightarrow PH$ is now in the digraph). Finally, PTC acts on the reaction $CI \rightarrow CN$ itself by promoting it without being itself affected, which in our notation means $PTC \overset{+}{\rightarrow} CN$ and $PTC \overset{-}{\rightarrow} CI$.

The Implementation. The Matlab implementation of the algorithm on this digraph with 13 nodes and 20 edges produced several partitions with as many as 17 consistent edges. One of these possible partitions simply consists of placing the three nodes ci , CI and CN in one set and all other nodes in the other set, whereby the only inconsistent edges are $CL \overset{+}{\rightarrow} wg$, $CL \overset{+}{\rightarrow} ptc$, and $PTC \overset{+}{\rightarrow} CN$. But note that it is desirable for the resulting open loop system to have as simple remaining loops as possible after eliminating all inconsistent edges. In this case, the remaining directed loops $EN \overset{-}{\rightarrow} ci \overset{+}{\rightarrow} CI \overset{+}{\rightarrow} CN \overset{-}{\rightarrow} en \overset{+}{\rightarrow} EN$ and $EN \overset{-}{\rightarrow} ci \overset{+}{\rightarrow} CI \overset{+}{\rightarrow} CN \overset{-}{\rightarrow} wg \overset{+}{\rightarrow} WG \overset{+}{\rightarrow} WG(membrane) \overset{+}{\rightarrow} en \overset{+}{\rightarrow} EN$ can still cause difficulties.

A second partition which generated 17 consistent edges is that in which EN, hh, CN, and the membrane compounds PTC, PH, HH are on one set, and the remaining compounds on the other. The edges cut are $\text{ptc} \xrightarrow{+} \text{PTC}$, $\text{CI} \xrightarrow{+} \text{CN}$ and $\text{en} \xrightarrow{+} \text{EN}$, each of which eliminates one or several positive loops. By writing the remaining consistent digraph in the form of a cascade, it is easy to see that the only loop whatsoever remaining is $\text{wg} \leftrightarrow \text{WG}$; this makes the analysis proposed in [12] easier. In this relatively low dimensional case we can prove that in fact $\text{OPT} = 17$ as stated below.

Lemma 6. *Any partition of the nodes in the digraph in Figure 2 generates at most 17 consistent edges.*

It is surprising that a realistic biological system with as many as 13 variables and 20 edges can be transformed into a monotone system after the deletion of only three nodes. It is conceivable that this restricts the possible dynamics of the system. This is especially the case given that the open loop digraph has almost no closed oriented paths (except for $\text{WG} \leftrightarrow \text{wg}$), which is evidence that the dynamics of the control system under constant inputs may be especially simple, e.g. such that all solutions converge towards a unique equilibrium.

Multiple Copies. It was mentioned above that the purpose of this network is to create striped patterns of protein concentrations along multiple cells. In this sense, it is most meaningful to consider a *coupled* collection of networks as it is given originally in Figures 1 and 2. Consider a row of k cells, each of which has independent concentration variables for each of the compounds, and let the cell-to-cell interactions be as in Figure 2 with cyclic boundary conditions (that is, the k -th cell is coupled with the first in the natural way). We show that the results can be extended in a very similar manner as before.

Lemma 7. *For the k -cell linearly coupled network described in Figure 2, $\text{OPT} = 17k$.*

5.2 EGFR Signaling

In their May 2005 paper [25], Oda et al. integrate the information that has become available about the *epidermal growth factor receptor* (EGFR) signalling process from multiple sources, and they define a network with 330 known molecules under 211 chemical reactions. The network itself is available from the supplementary material in SBML format (*Systems Biology Markup Language*, www.sbml.org), and will most likely be subject to continuous updates.

The Implementation. Each reaction in the network classifies the molecules as reactants, products, and/or modifiers (enzymes). This information was imported into Matlab using the Systems Biology Toolbox. The digraph G that is used for this analysis has many more edges than the digraph considered in the digraph displayed in [25]. The reason for this is as follows: if molecules A and B are both reactants in the same reaction, then the presence of A will have an indirect inhibiting effect on the concentration of B , since it will accelerate the reaction

which consumes B (assuming B is not also a product). Therefore a negative edge must also appear from A to B , and vice versa. Similarly, modifiers have an inhibiting effect on reactants. We thus define G by letting $\text{sign}(i, j) = 1$ if there exists a reaction in which j is a product and i is either a reactant or a modifier. We let $\text{sign}(i, j) = -1$ if there exists a reaction in which j is a reactant, and i is also either a reactant or a modifier. Similarly $\text{sign}(i, j) = 0$ if the nodes i, j are not simultaneously involved in any given reaction, and $\text{sign}(i, j)$ is undefined (NaN) if the first two conditions above are both satisfied. An undefined edge can be thought of as an edge that is *both* positive and negative, and it can be dealt with, given an arbitrary partition, by deleting exactly one of the two signed edges so that the remaining edge is consistent. Thus, in practice, one can consider undefined edges as edges with sign 0, and simply add the number of undefined edges to the number of inconsistent edges in the end of each procedure, in order to form the total number of inputs. This is the approach followed here; there are exactly 7 such entries in the digraph G .

The Results. After running the algorithm 100 times for this problem, and choosing that partition which produced the highest number of consistent edges, the induced consistent set contained 633 out of 852 edges (ignoring the edges on the diagonal and the 7 undefined edges). See the supplementary material for the relevant Matlab functions that carry out this algorithm. A procedure analogous to that carried out for system (5) allows to decompose the system as the feedback loop of a controlled monotone system using $852 - 633 = 219$ inputs. Since the induced consistent set is maximal by definition, we are guaranteed that the function h is a negative feedback. Contrary to the previous application, many of the reactions involve several reactants and products in a single reaction. This induces a denser amount of negative and positive edges: even though there are 211 reactions, there are 852 (directed) edges in the 330×330 graph G . It is very likely that this substantially decreases OPT for this system. The approximation ratio of the SDP algorithm is guaranteed to be at least 0.87 for some r , which gives the estimate $\text{OPT} \leq \approx 633/0.87 \approx 728$ (valid to the extent that r has sampled the right areas of the 330-dimensional sphere, but reasonably accurate in practice).

One procedure that can be carried out to lower the number of inputs is a hybrid algorithm involving *out-hubs*, that is, nodes with an abnormally high out-degree. Recall from the description of the *DLP* algorithm that all the out-edges of a node x_i can be potentially cut at the expense of only one input u , by replacing all the appearances of x_i in $f_j(x)$, $j \neq i$, by u . We considered the k nodes with the highest out-degrees, and eliminated all the out-edges associated to these hubs from the reaction digraph to form the graph G_1 . Then we run the *ULP* algorithm on G_1 to find a partition f_V of the nodes and a set of m edges that can be cut to eliminate all remaining negative closed chains. Finally, we put back on the digraph those edges that were taken in the first step, and which are consistent with respect to the partition f_V . The result is a decomposition of the system as the negative feedback loop of a controlled monotone system, using at most $k + m$ edges.

An implementation of this algorithm with $k = 60$ yielded a total maximum number of inputs $k + m = 137$. This is a significant improvement over the 226 inputs in the original algorithm. Clearly, it would be worthwhile to investigate further the problem of designing efficient algorithms for the *DLP* problem to generate improved hybrid algorithmic approaches. The approximation ratios in Theorem 5(c) are not very satisfactory since d_{in}^{max} and $\log |V|$ could be large factors; hence future research work may be carried out in designing better approximation algorithms.

5.3 Supplementary Material: MATLAB Implementation Files

A set of MATLAB programs have been written to implement the algorithms described in this paper. They can be accessed from the URL http://www.math.rutgers.edu/~sontag/desz_README.html.

References

1. D. Angeli, J. E. Ferrell Jr. and E. D. Sontag. *Detection of multi-stability, bifurcations, and hysteresis in a large class of biological positive-feedback systems*, Proc. Nat. Acad. Sci. USA, 101: 1822-1827, 2004.
2. D. Angeli, P. De Leenheer and E. D. Sontag. *A small-gain theorem for almost global convergence of monotone systems*, Systems and Control Letters, 51: 185-202, 2004.
3. D. Angeli and E.D. Sontag. *Monotone control systems*, IEEE Trans. Autom. Control, 48: 1684-1698, 2003.
4. D. Angeli and E. D. Sontag. *Multistability in monotone I/O systems*, Systems and Control Letters, 51: 185-202, 2004.
5. D. Angeli and E. D. Sontag. *An analysis of a circadian model using the small-gain approach to monotone systems*, proceedings of the IEEE Conf. Decision and Control, Paradise Island, Bahamas, Dec. 2004, IEEE Publications, 575-578, 2004.
6. O. Cinquin and J. Demongeot. *Positive and negative feedback: striking a balance between necessary antagonists*, J. Theor. Biol., 216: 229-241, 2002,
7. G. von Dassaw, E. Meir, E.M. Munro, and G.M. Odell. *The segment polarity network is a robust developmental module*, Nature 406: 188-192, 2000.
8. D. L. DeAngelis, W. M. Post and C. C. Travis. *Positive Feedback in Natural Systems*, Springer-Verlag, New York, 1986.
9. G.A. Enciso, H.L. Smith, and E. D. Sontag. *Non-monotone systems decomposable into monotone systems with negative feedback*, J. of Differential Equations, 2005, to appear.
10. G. Enciso and E. Sontag. *On the stability of a model of testosterone dynamics*, Journal of Mathematical Biology 49: 627-634, 2004.
11. G. Enciso and E. D. Sontag. *Monotone systems under positive feedback: multistability and a reduction theorem*, Systems and Control Letters, 54: 159-168, 2005.
12. G. Enciso and E. Sontag. *Global attractivity, I/O monotone small-gain theorems, and biological delay systems*, Discrete and Continuous Dynamical Systems, to appear.
13. T. Gedeon and E. D. Sontag. *Oscillation in multi-stable monotone system with slowly varying positive feedback*, submitted for publication (abstract in Sixth SIAM Conf. on Control and its Applications, New Orleans, July 2005).

14. M. Goemans and D. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM, 42 (6): 1115-1145, 1995.
15. M. Hirsch. *Systems of differential equations that are competitive or cooperative II: Convergence almost everywhere*, SIAM J. Mathematical Analysis, 16: 423-439, 1985.
16. M. Hirsch. *Differential equations and convergence almost everywhere in strongly monotone flows*, Contemporary Mathematics, 17: 267-285, 1983.
17. N. Ingolia. *Topology and robustness in the drosophila segment polarity network*, Public Library of Science, 2 (6): 0805-0815, 2004.
18. P. De Leenheer, D. Angeli and E.D. Sontag. *On predator-prey systems and small-gain theorems*, J. Mathematical Biosciences and Engineering, 2: 25-42, 2005.
19. P. De Leenheer and M. Malisoff. *Remarks on monotone control systems with multi-valued input-state characteristics*, proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference ECC 05 (Seville, Spain, December 2005), to appear.
20. J. Lewis, J.M. Slack and L. Wolpert. *Thresholds in development*, J. Theor. Biol., 65: 579-590, 1977.
21. A. Maayan, R. Iyengar and E. D. Sontag. *A study of almost-sign consistency on biological networks*, in preparation.
22. H. Meinhardt. *Space-dependent cell determination under the control of morphogen gradient*, J. Theor. Biol., 74: 307-321, 1978.
23. J. Monod and F. Jacob. *General conclusions: telenomic mechanisms in cellular metabolism, growth, and differentiation*, Cold Spring Harbor Symp. Quant. Biol., 26: 389-401, 1961.
24. J.D. Murray. *Mathematical Biology, I: An introduction*, New York, Springer, 2002.
25. K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano. *A comprehensive pathway map of epidermal growth factor receptor signaling*, Molecular Systems Biology doi:10.1038/msb4100014, 2005.
26. E. Plathe, T. Mestl and S.W. Omholt. *Feedback loops, stability and multistationarity in dynamical systems*, J. Biol. Syst., 3: 409-413, 1995.
27. E. Remy, B. Mosse, C. Chaouiya and D. Thieffry. *A description of dynamical graphs associated to elementary regulatory circuits*, Bioinformatics, 19 (Suppl. 2): ii172-ii178, 2003.
28. I. Shmulevich and E.D. Sontag. *Order and Distance to Monotonicity in Boolean Networks*, in preparation.
29. H. L. Smith. *Monotone Dynamical Systems*, Providence, R.I., AMS 1995.
30. H. L. Smith. *Systems of ordinary differential equations wich generate an order-preserving flow: A survey of results*, SIAM Reviews, 30: 87-111, 1988.
31. E. H. Snoussi. *Necessary conditions for multistationarity and stable periodicity*, J. Biol. Syst., 6: 39, 1998.
32. E. D. Sontag. *Some new directions in control theory inspired by systems biology*, Systems Biology, 1: 9-18, 2004.
33. E. D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, Springer, New York, 1990 (Second Edition, 1998).
34. R. Thomas. *Logical analysis of systems comprising feedback loops*, J. Theor. Biol., 73: 631-656, 1978.
35. A. I. Volpert, V. A. Volpert, and V. A. Volpert. *Traveling Wave Solutions of Parabolic Systems*, volume 140 of Translations of Mathematical Monographs, AMS, 2000.