Stability and Flexibility from a System Analysis of Gene Regulatory Networks Based on Ordinary Differential Equations

Mika Gustafsson and Michael Hörnquist*

Linköping University, Department of Science and Technology, 601 74 Norrköping, Sweden

Abstract: The inference of large-scale gene regulatory networks from high-throughput data sets has revealed a diverse picture of only partially overlapping descriptions. Nevertheless, several properties in the organization of these networks are recurrent, such as hubs, a modular structure and certain motifs. Several authors have recently claimed cell systems to be stable against perturbations and random errors, but still able to rapidly switch between different states from specific stimuli. Since inferred mathematical models of large-scale systems need to be extremely simple to avoid overfitting, these two features are hard to attain simultaneously for a model. Here we review and discuss possible measures of how system stability and flexibility may be manifested and measured for linearized models based on systems of ordinary differential equations. Furthermore, we review how the network properties mentioned above together with the nature of the interactions contribute to these systems level properties. It turns out that the presence of repressed hubs, together with other phenomena of topological nature such as motifs and modules, contribute to the overall stability and/or flexibility of the model.

Keywords: Systems biology, dynamical modeling, complex networks, gene expression.

INTRODUCTION

Networks have during more than one decade now been extensively used as a unifying language for describing diverse systems, including social contacts, power grids, airport connections, ecological food webs and also diverse inter and intra cellular systems. Among the intra cellular systems there are several different types of networks, ranging from descriptions of physical interactions of proteins, to detailed metabolic and regulatory networks describing biochemical reactions. Regulatory networks, defined from the interaction maps of proteins, RNA molecules metabolites and the genome, determine in combination with detailed kinetics the cellular responses to input signals and govern cellular dynamics [1]. Some parts of these huge systems are well explored, e.g., the yeast cell cycle [2], the lambda switch [3], and the SOS pathway [4], but still many of the parts are unknown. However, the learning of biological networks is essential for both the understanding of how cell systems work and for making predictions of cellular responses. Although measurements of several of these molecular units are available at large-scales, the information is limited due to the size of the system which contains several thousands of units for many organisms. The major sources of information utilized in inference processes have often been the high-throughput data sets measuring gene expression (e.g., microarrays), interaction maps derived from other experimental techniques (e.g., ChIP-chip) and biological knowledge originating from text mining [5].

A popular simplifying approach is to project transcripts and proteins on the associated coding genes, and disregard the metabolites. The corresponding effective gene-to-gene network is called a *Gene Regulatory Network*, GRN (also known as influence network [5] or gene network [6]), and describes regulations directly between genes, although the full regulatory system contains complex interactions among all these entities. This projection leads to loss of details in the model, but is probably a necessary simplification due to sparseness of data from similar conditions. Data sparseness is indeed a major problem for large-scale modeling projects; despite the increasing amount of available high-throughput data sets we still have few experiments from the same condition measuring different entities (e.g., mRNA and protein concentrations). That is, integration of various data sources is crucial, but impeded by different standards between platforms and experimental conditions. Thus, large-scale regulatory network models are still today mainly constructed as projections of the full regulatory system onto subspaces, which contain fewer units and fewer interactions due to the limited number of included conditions. Also, these interactions do not necessarily correspond to biochemical reactions or to physical interactions between proteins and targets; instead they correspond to the effective impact of the regulator gene on its downstream target. Although this reduction sometimes can be considered as a severe limitation, it also has positive effects. For example, it enables a comparison of network properties for models obtained by different inference approaches. Moreover, we may have both pieces of structural information and qualitative information available of the same gene, thus we increase the amount of information known per unit.

Even with detailed knowledge of the kinetic framework in the full system, it would generally be hard to deduce the functional form of the kinetics in the gene space in practise. This is due to the fact that the functional form of the "effective" kinetics in the projected gene space may largely depend

*Address correspondence to this author at the Linköping University, Department of Science and Technology, 601 74 Norrköping, Sweden; Tel: +46 11363381; Fax: +46 11363270; E-mail: michael.hornquist@liu.se

on the exact model parameters of the full system. Since inference approaches do not know the model parameters in advance they should either be based on non-parametric adaptive models or "simple-enough" models as we are short of data. In the present presentation, we focus on the "simple-enough" models in the form of gene regulatory networks. Due to lack of rigidity in the model assumptions the outcome should be taken with some care, thus it should be validated *a posteriori*. Ideally, this validation should contain the validation of individual edges and of responses of the system to perturbations. The validation could be done experimentally or more cost-effective by reusing available databases (not utilized in the inference process, of course). However, several research groups have reported networks depending on different data sources with a great variability [7]. This stresses the complexity of the true regulatory system and also opens the question about other sources of validation. The variability among the models may be the result of either conceptual differences of the approaches or improper modeling (see below).

The gene regulatory networks are effective by nature [6] and context dependent [8], e.g., some metabolites/proteins may be needed to activate a TF (e.g., phosphorylation), or recruiting proteins bringing the TF to the promoter may be needed for a regulatory interaction to work [9]. Thus, these networks represented by the active parts of the regulatory networks are indeed condition specific.

Further, the research area is still in a premature stage and a lot of development is going on. For example, in 2006 Califano and Stolovitzky initiated the *Dialogue for Reverse Engineering Assessments and Methods* (DREAM) [10], where groups have been assessing the performance of their algorithms in different challenges. The DREAM challenges have been held annually thus far, and in addition to the competition it provides great benchmark sets and evaluation criteria for the development of algorithms. The challenges concerning large-scale network identification have so far been unrealistic by the amount of (*in silico*) data available, but still they constitute the hallmarks known to date and have therefore provided important insights. From the perspective of integrating various data types, denoted as "crucial" above, the DREAM challenge of 2008 was interesting. Here Yip *et al.* [11] presented an interesting winning strategy how steady state and time-series expression data effectively can be integrated to infer edges. Moreover, Gustafsson *et al.* [12] showed how uncertain a priori information about gene interactions from large-scale experiments as well as expression sets from other conditions may be utilized to improve prediction accuracy of expression levels.

The models utilized to describe the time-evolution of the gene levels with gene regulatory networks may be classified with respect to the number of states per node they allow. The Boolean networks [13] and threshold networks [14] are both ideal two state modeling formalisms. They approximate the activity of each gene as being either simply on or off, which is motivated if the upper or lower limit of the gene activity often is attained. However, this approximation discards all intermediate steps and may therefore be of interest for modeling multiple conditions within the same framework, but is probably not suitable for modeling a cell around a stable working point (e.g., the cell cycle). The Bayesian networks

are another modeling framework [15], which is more flexible and can also allow for multiple states per node and can even be extended to model continuous levels [16]. Also models based on Ordinary Differential Equations (ODE-models) assume a continuum of activity levels, thus both are particularly suitable to model systems around a working point. However, for the Bayesian networks it takes either much more data for inference or further a priori assumptions (e.g., restricting the degree distribution) to model continuity, which is problematic since already a Bayesian Network with two levels demands much training data. On the other hand, with proper restrictions (regularizations) that are biologically motivated (e.g., lasso, ridge, or elastic-net [17]) the ODE-models are both computationally tractable and may need fewer experiments for inference than the former categories. Interestingly, the best performing teams of the DREAM2 and DREAM3 network inference challenges have utilized models based on structure and parameter estimation for ODEs [7, 11, 18, 19].

Despite the large network variability over individual edges that has been reported [7], there are several topological findings that the community seems to agree upon. These findings include the presence of hubs regulating peripheral nodes, particular motifs (i.e., recurrent sub-graphs) and a modular network structure. It seems that these features are characteristic for gene regulatory networks (as well as for many other complex networks). In here, we will discuss and review how these findings contribute to the overall system dynamics. We will focus the discussion to some recent work by Gustafsson *et al.* [7], which inferred and analyzed a genome-wide gene regulatory network from time-series profiles of mRNA-levels measured during the yeast cell cycle. The inference was carried out for a linearized version of the ODE-system describing the expression dynamics and by using sparsity as a prior assumption (i.e., keeping the regulations simple) [20]. This inference identified interactions, simultaneously with the strength, direction and whether it had an activating or repressing effect. Together with the system of linearized equations this sets the stage for a dynamical system analysis around the working state of the cell cycle. The purpose of the analysis is to determine what the topological and dynamical pieces bring together to produce a system that is stable against noise yet responsive to stimuli.

In Fig. (**1**) we show a schematic overview of the methodology, which goes from high-throughput experimental data (left), to a network model (middle), which we subsequently can draw conclusions from (right). Conclusions may be drawn both about the topology of the inferred network, i.e*.,* how it is organized, and also about its dynamical properties, e.g*.,* its stability against noise and its ability to respond to stimuli.

The rest of the paper is organized as follows. In the next section, *Dynamics, Stability and Flexibility,* we introduce and review the concepts of dynamics, system stability and flexibility, clarifying the various meanings of the words in the present context. Thereafter, in section *System Analysis,* we review topological and global dynamical findings about hubs, as well as discuss features beyond hubs such as motifs and modules and their impact on system stability and flexibility. We end the paper with a summary and an outlook. Throughout the paper, we have tried to avoid mathematical
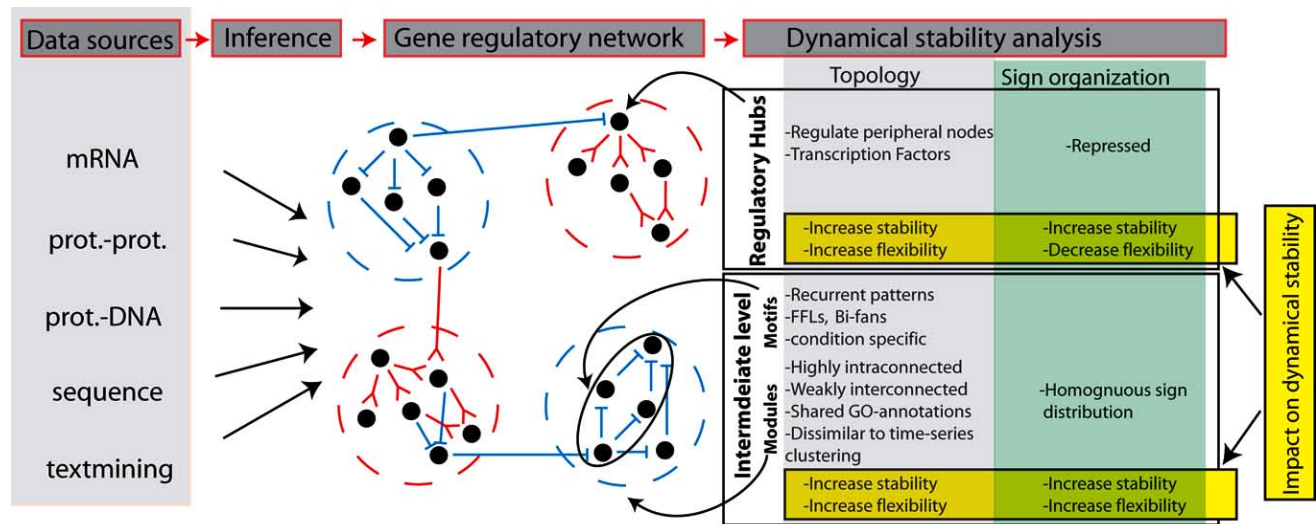
Fig. (1). Map of the steps from experiments to knowledge, and a summary of important features in the organization and their impacts. In principle, all data types mentioned at the left hand side can be utilized for inferring gene regulatory networks; although in the present presentation we discuss networks based solely on mRNA-data.

formulas, and instead we refer to the references for the reader interested in exact formulations.

DYNAMICS, STABILITY AND FLEXIBILITY

There is a need to clarify our usage of the concepts *dynamics, stability,* and *flexibility* since they are frequently employed for networks within the systems biology community with multiple meanings.

Dynamics is widely used for cellular networks for both the topological evolution of networks [1, 21-23] as well as for describing expression dynamics [7, 24, 25]. The former includes both addition and removal of nodes and edges, as well as rewiring, and can in the context of cellular networks be thought of as representing the long term evolution over several generations based on Darwinian selection. The latter describes how the values of the nodes vary for a network with a fix topology. For ODE-models of gene regulatory networks, this means the trajectories in state-space which the differential equations trace. A biological interpretation is the relatively short term fluctuations of gene levels over parts of an individual life time, e.g., the variation over the cell cycle. We will here use the word dynamics to refer to this latter variation of gene levels within the network model.

TRADITIONAL DEFINITIONS OF STABILITY

There are several forms of stability, such as *structural stability, marginal stability* and *dynamical stability,* and also several reviews in the field discussing these concepts with not totally identical definitions [26-28]. In the present presentation, we will utilize a stability concept suitable for the kind of dynamics we are interested in. Since we are dealing with linearized ODE-systems, a tentative definition of stability is:

A model is (dynamically) stable if any perturbation of a state fades away after some time.

Of course, also other stability concepts could be of interest, such as stability referring to the effect on the solution paths from a perturbation of the model parameters (some-

times called structural stability [26]), or stability of the model to changes of topological quantities such as edge removals (sometimes called topological stability) [21]. Worth noting, the authors of the work [21] concerning topological stability concluded that networks with only a few well connected hubs were stable (or "robust", as they phrased it, a terminology followed by others [29]) against most random removals, but extra sensitive towards removal of some specific edges (fragile). Although these forms of stability probably are of great importance for some networks, where the main source of errors is edge failures (e.g., power grids), we concentrate here on the form of (dynamical) stability defined tentatively above, in order not to grasp too much. On the scale of intracellular processes, the major source of perturbations for gene regulatory networks still comes from changes in the gene expression pattern, based on stochastic variation of expression values as well as external stimuli. However, the study of structural or topological stability will hopefully be a subject for future research also for biological networks, since from an evolutionary perspective it might be of importance, e.g., the impact of gene duplication and the effect of mutations in promoter regions.

STABILITY AND FLEXIBILITY FOR LINEARIZED ODE-MODELS OF GRNs

Dynamical stability in ODE-models is often determined by analyzing the linearized equations around a working point. It is then easy to compute the trajectory of any expression state from the eigenvectors and eigenvalues of the Jacobian matrix, given that it stays close enough to the working point were the linearization is still valid. This matrix can be considered as a generalization of the adjacency matrix and each element contains the regulatory strength and nature (activation/repression) of the directed interaction. Also, our tentative definition of (dynamical) stability can from a mathematical point of view be phrased such that the real part of all eigenvalues of the matrix should be negative [7]. Furthermore, the eigensolutions to the Jacobian matrix describe either exponentially growing (or decaying) states or oscilla-

tory states. Conventional analysis for linear systems then concludes that the presence of any growing state eventually will lead to system instabilities, due to noise. This may be problematic in a large complex system (with sufficiently number of edges) if it is not properly tuned, since increased complexity leads to an increased probability for instabilities in random systems [30] (which, however, may be compensated by a strong self-degradation in the system). In a series of work Sinha *et al.* [31-33] studied the impact of some topological features on stability. In particular, they concluded that the presence of hubs (with random sign distribution) increases the probability of some growing eigenstates, which may be compensated for by a modular organization. Notably, their studies reveal that dynamical stability in the classic strict sense is less probable with the heavy tailed out-degree distribution found in gene networks by other groups [20, 34]. These studies reveal interesting couplings between topology and dynamical stability, but completely lack the influence of the model parameters on the dynamical stability. This influence is probably crucial for gene regulatory networks, since some studies have observed an intricate balance between activation and repression [7, 24].

The flip-side of dynamical stability is the responsiveness of the system to stimuli to switch into new working states; evidently this flexibility comes on the expense of stability in the system. In a Boolean setting a system with a compromise between these two is called *critical* and a recent study by Balleza *et al.* [35] revealed several different organisms to have topologies and model parameters to facilitate a near critical system. Even though our primary interest in this article is networks with an ODE dynamics, the studies of Boolean networks [13, 35] are of importance for the discussion of system stability in gene regulatory networks since not much have been studied for ODEs.

When a stability analysis is performed for a linear system, a single growing eigenstate will induce instability of the whole system. On the other hand a growing eigenstate in the linearization of a non-linear model may also reflect a switch to a new working point. For example, linear instabilities are deliberately used in some air-fighters control system to increase its maneuverability [36]. Hence, in a gene regulatory network, a growing eigenstate may reflect a fast switch from the current working state to some other which beneficially should be rapid, e.g., a switch from the cell cycle to stress conditions. However, as the presence of growing eigenstates leads to a drift from the current working point, which must be compensated by non-linear effects, growing eigenstates must be utilized only cautiously to be advantageous for the regulation of the cell. Therefore to reflect this system flexibility, but also because of errors when inferring a large complex network on noisy incomplete data, a linearized model of a gene regulatory network will most probably contain some (but not too many) fast growing eigenstates. The definition of stability based on its eigenvalues needs as a consequence to be adapted.

REDEFINING STABILITY AND QUANTIFYING FLEXIBILITY

To quantify the degree of responsiveness to selective stimuli and the stability against noise for linearized/incomplete gene regulatory networks, Gustafsson *et al.* [7] made an adaptation of the two quantities *flexibility* and *stability* based on the distribution of the growth rates (i.e., on the real parts of the corresponding positive eigenval-
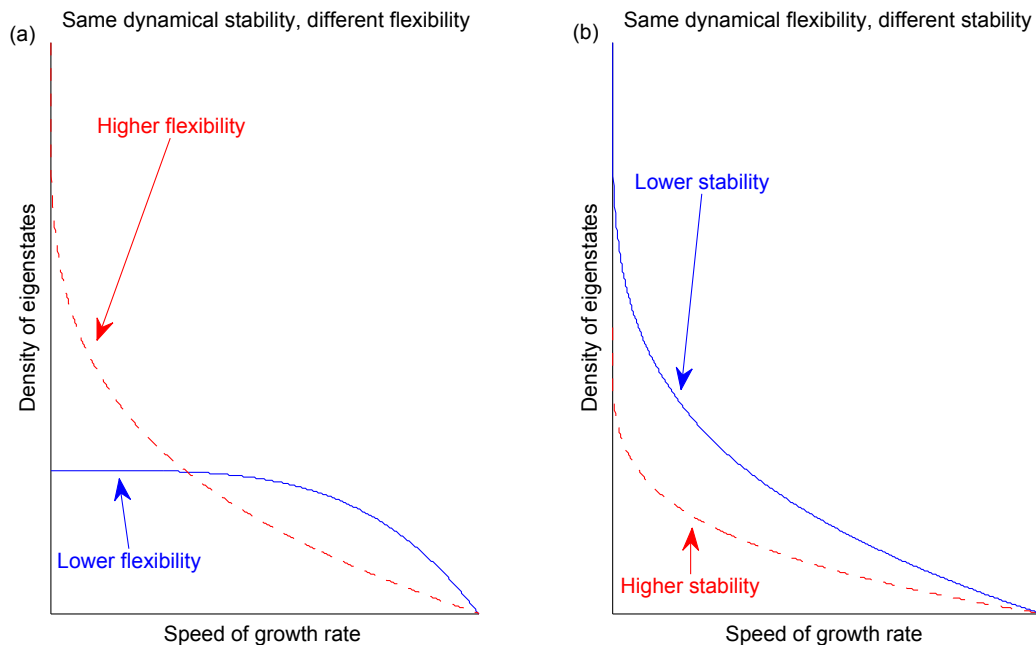
Fig. (2). Illustration of the concepts of flexibility and stability based on the distribution of real parts of the eigenvalues to the Jacobian matrix. Each curve corresponds to one network, and the integral of the curve (that is, the area under the curve) corresponds to the number of growing eigenstates. The vertical axis depicts the density of states corresponding to a certain growth speed given by the horizontal axis. In (**a**) we can see two curves corresponding to two networks with the same stability but with different flexibilities. The network with a relatively uniform density of growing eigenstate growth rates (blue solid curve) has a lower flexibility than the network with a skewed distribution (red dashed curve). The right figure (**b**) depicts two networks with same flexibility but different stabilities, which is manifested from the two distributions sharing the shape.

ues) of the growing eigenstates, suitable for GRNs. Now, *instability* is first defined as the normalized sum of amplitude growth rates of all growing eigenstates. Thus, this sum is positively correlated to the rate with which the system drifts due to random fluctuations from the linearized state. Thereafter, *stability* is defined as one minus the instability.

Next, given an arbitrary stability, it is informative to study the source of it, i.e., how the growth rates of the instabilities are distributed. A distribution of growth rates like the blue solid curve in Fig. (**2a**) reflects a non-specific divergence, i.e., the drift will go in a random direction spanned by the growing eigenstates. Contrary, a skewed spectrum of growth rates like the red dashed curve in Fig. (**2a**) reflects that the divergence will relatively quickly be confined to a small subspace (spanned by the most rapidly growing eigenstates). Therefore, in Gustafsson *et al.* [7] *flexibility* was defined from the skewness of the distribution of the growth rates of the growing eigenstates (see [7] for exact definition), which then reflects the specificity of the drift. In Fig. (**2b**) we have the opposite situation of Fig. (**2a**), with two networks with growth rate distributions corresponding to the same flexibility but different stabilities.

SYSTEM ANALYSIS

In order to understand the effect of the global topological design principles in gene regulatory networks, it is important to analyze their corresponding impact on the dynamics of the system. Below we will review some important topological findings in gene regulatory networks, and how they contribute to the system stability and flexibility. We call an analysis based on these concepts for a (dynamical) system analysis, since it is based on the possible dynamical states of the system from its organization.

In Fig. (**3**) we see a system analysis more exactly, a depiction of stability and flexibility) of the Yeast network (mentioned in the Introduction, inferred from genome-wide microarray data in time-series from the cell cycle by imposing a sparse linear ODE-model, originally derived in [20]). Moreover, we see several ensembles of networks, with increased similarity to the Yeast network. From the difference of the ensemble averages and the corresponding standard deviations with various topological features of interest kept constant, it is possible to determine the impact (and the statistical significance) of these topological features. Each point in the coordinate system of Fig. (**3**) corresponds to a network, or an ensemble of networks, relating to a curve similar to those presented in Fig. (**2**). For example, the blue solid and red dashed curve of Fig. (**2a**) should give rise to two separate points in Fig. (**3**), where the point corresponding to the red dashed curve is positioned higher up than the point corresponding to the blue solid curve, reflecting the non-skew distribution for the blue curve, but at the same horizontal position since they have the same stability. Thus, the networks in the upper part of Fig. (**3**) have skewed distributions of growth rates, i.e., they are dominated by a few growth eigenstates. Moreover, the curves of Fig. (**2b**) correspond to two points with different horizontal position, but same vertical, since they have the same flexibility but different stabilities.

HUBS

Hubs, that is, nodes with a large number of incoming or outgoing connections, have been observed in various biological and other networks since the late 1990's [1, 21, 38-40], when the global effect of hubs in a network was suggested by Barabasi and Albert to produce a system stable against random failures, yet fragile towards targeted attacks (i.e., topological stability). In a gene regulatory network a significant portion of the transcription factors (TFs) was shown to act as regulatory hubs [34, 41] which then may act as master switches between different states of the cell. Moreover, Maslov and Sneppen [41] demonstrated that regulatory hubs in the yeast transcriptional network tend to regulate genes with low in-degree (peripheral genes) on average (assortative mixing). This suggests that many regulatory hubs are not master switches; instead their main goal is to mediate a signal in a cost efficient manner. Furthermore, two recent studies [7, 24] reported these hubs to be kept quiet by a negative in-regulation. All these aspects associated to hubs may contribute both to the dynamical stability and flexibility of the system. In Fig. (**3**) we isolate the effect of the presence of hubs, and the regulation of hubs, in two of the steps between ensembles.

In the first step for hubs, the first arrow of Fig. (**3**), we explore the topological effect of having out-degree hubs as well as a high portions of genes with out-degree zero (non-regulators). This is done by comparing two ensembles of networks with the same number of nodes and edges, but different topologies. The Erdös-Renyi (ER) network ensemble, obtained from a random distribution of edges among the nodes, and the ensemble with the same degree distribution as the inferred Yeast network (REWIRED), obtained from the randomization process developed by Maslov and Sneppen [37], have the same numbers of nodes and edges, and also have the same weights, but have different degree distributions. The ensemble of ER-networks has a degree distribution following a Poissonian curve, and hence contains no hubs, while the REWIRED ensemble, having the same degree distribution as Yeast Topology, has a broad distribution of out-degrees and hence contains several hubs [20]. Strikingly, the REWIRED ensemble has both significantly higher stability and flexibility. This may intuitively be understood from the introduction of both a small portion of genes with high regulatory influence on others (hubs) and from a large portion of genes with no influence (non-regulators). The large fraction of non-regulators induce an equal amount of non-growing eigenstates, since if we perturb one of these genes it could not propagate downstream the network, hence a perturbation of a random gene in the REWIRED ensemble has lower probability to be of growing nature than a perturbation of a random gene in one of the networks in the ER ensemble, thus making REWIRED more stable than the ER. On the other hand a perturbation of any of the regulatory hubs is likely to propagate downstream in a few steps and due to the complex network structure grow quickly. For stability we see that the presence of non-growing eigenstates (from the non-regulators) is of most importance, while it is the presence of only some fast growing eigenstates (a small portion of hubs) for flexibility. These intuitive arguments are analogous to the ones proposed by Barabasi and Albert [39], mentioned above, to explain the degree distribution observed
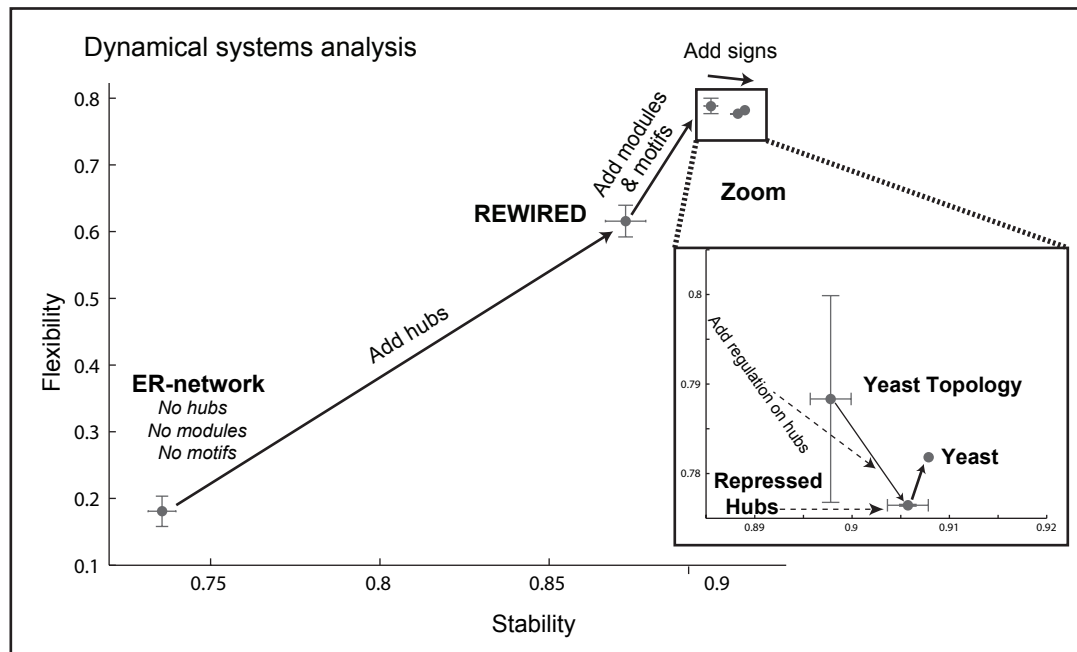
Fig. (3). Stability and flexibility for the inferred network (Yeast) and several randomized versions thereof. The error bars cover two standard deviations for each ensemble of networks (see text). Starting in the lower left corner of the figure, we have the ensemble of ER-like networks, thereafter successively more and more topological and dynamical features are added to the network, thus obtaining new ensembles of networks more and more similar to the Yeast network. All ensembles contain the same model parameters (see text), but are randomized in various ways.

in the protein-protein network, in which case the non-regulators induce robustness to random deletion and hubs induce fragility to targeted attacks.

For the second step concerning hubs, the third arrow of Fig. (**3**) (seen at the inset), we explore also the dynamical effect from the organization of repression and activation. This is the step "Add regulation on hubs", which separates the ensemble Yeast Topology corresponding to networks with identical structure to the inferred network, but randomized signs and weights and the ensemble Repressed Hubs with the same topology and also identical signs and weights as the Yeast network on all ingoing regulations to regulatory hubs (now defined as nodes with out-degree of two or more), but with all other signs and weights randomized. Here we can see that the particular direct regulation of the hubs increases the stability of the system, with the expense of its flexibility. Since there is an excess of negative regulations for these hubs [7], this probably come from a stabilization of the fastest growing states, which both has an increasing effect of the stability and a decreasing effect of the flexibility. It is likely that the fastest growing states are to be the targets of the in-regulation of hubs, since they form the computational core [41] and therefore are crucial as master switches between different working points.

BEYOND HUBS – MOTIFS AND MODULES

Evidently, the degree distribution is an important design principle, which both has high descriptive power of the architecture of gene networks and great impact on the network dynamics. However, to describe and gain understanding of the system it is necessary also to study features which involve the interaction of several entities.

Several authors have found recurrent graph structures called *motifs* to be significant for networks [23, 42]. They are frequently associated with specific structures of the dynamical parameters [7, 22, 23], e.g., the signs in the Feed Forward Loop (FFL) are often organized to yield coherent signals to the downstream target. The dynamical effect of some motifs in isolation has been studied in experimental detail [22] and also their frequencies were shown to change considerably between exogenous and endogenous processes [8].

Another popular concept which is frequently reported in the context of networks is *modules*. The core idea is that a module is a functional unit which works relatively isolated and performs a specific task; hence it consists of genes with a high degree of process similarity [43]. For a gene regulatory network, this can be a group of genes where several of the genes in the module share a common GO process annotation [44]. In engineered systems, modules may be stable (in both our meaning and in the topological meaning [45]) subunits performing different tasks. In biological systems, modularity brings an evolutionary flexibility to the system, and recent studies suggest it origins from time-varying evolutionary goals [46, 47], with each module serving its own special task. Despite the general consensus about modularity in networks describing cells, the concept of modules has been used in various ways (but still motivated from a high process similarity within the modules [43, 48, 49]). The first attempts used time-series clustering to detect gene clusters with a high degree of process similarity [48]. More recently, network approaches originally based on the graph theoretic concept of tightly connected sub-graphs, called *communities* [50], have been applied to the same problem. Furthermore, integrated approaches [49, 51, 52] taking into account both

structural and expression data have also been developed for this problem, see Wang *et al.* [52] and references therein for a review on the subject. However, it was observed [7] that methods based on communities and methods based on co-expression approaches may lead to completely different modules/clusters. Evidently, both co-expression based clustering and network clustering detect gene sets with high process similarity, but otherwise little in common. Moreover, some studies have revealed that modules can be organized into meta-modules [53] in a hierarchical structure [1].

From a dynamical system analysis point of view the individual effects from all topological aspects beyond hubs are hard to deduce. However, in Fig. (**3**) we can detect in the second step the combined effect on stability and flexibility for all topological phenomena consistent with the present degree distribution. This is illustrated by the arrow "Add modules & motifs", which separates the REWIRED ensemble of networks sharing only the node degree distribution with the Yeast network, from the Yeast Topology ensemble of networks with the exact same topology as the Yeast network, but randomized signs and weights. In [7] it is shown that the Yeast network has significantly higher density of motifs and modularity index than its rewired versions corresponding to the REWIRED ensemble. The significant increase in both stability and flexibility for this step indicates that these features are important for the system functionality. This is further repeated by the fourth and last arrow of Fig. (**3**) (seen at the inset), which separates the ensemble Repressed Hubs from the Yeast network. That is, this arrow corresponds to settle the inferred weights and signs on all remaining edges, the edges going to non-regulators. Since the Yeast network has the highest stability and also flexibility of the two, i.e., the arrow points upwards to the right, this reflects that the weights and signs on the edges to the non-regulators are organized to make the system stable yet flexible.

SUMMARY AND OUTLOOK

The understanding of gene regulatory networks inferred from experimental data is of great importance for understanding complex diseases and crucial for modeling cells *in silico*. Although the details of these systems are under debate, several statistical characteristics regarding the topology and dynamics are fairly consistent across different studies. Indeed, some parts of the regulatory systems are known, and some of them in terms of non-linear differential equations. However, to learn a large-scale model, possibly genome-wide, the experimental data are limited, and therefore simple models are still important. This includes rough dynamical descriptions by models based on linear first order ordinary differential equations. These models are computationally tractable and possible to regularize to fit into the problem of the limited amount of data. Further, the dynamics of linearized ODE network models is analytically tractable, but the analysis reveals that most of the models are unstable in a strict sense. However, by focusing on the fact that linear models of biological systems are only crude linearizations of the true systems, one can modify the conventional stability concept of linear systems into a system analysis of biological systems. In here, we discussed these new concepts and their relation to topological features and dynamical design principles recently found in gene regulatory networks. The intro-

duction of the investigated topological features (motifs and modules, and hubs, respectively) increased both stability and flexibility. Furthermore, we presented a study on the effects of repressed hubs, and of signs and weights on edges going to non-hubs. The effect of a high fraction of repressed hubs was found to produce a system more stable against noise but less flexible. The effect of having the "true" distribution of signs and weights on edges going to non-hubs, i.e., the distribution inferred from real data, was to increase both system stability and flexibility. The clear trend of the studied topological features to high stability and flexibility and the high separation of ensembles stress the meaning of the concepts and the importance of a system analysis. In the future, we believe these concepts will play important roles both for an increased understanding of the gene regulatory network models and for discriminating among such models. However, they first need to be further refined to better fit systems of various sizes before one can start to make comparisons between networks for other processes or organisms. It would also be very interesting to include topological changes into the framework in order to shed further light on the evolution of networks.

ACKNOWLEDGEMENTS

REFERENCES

[1]   A. L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat. Rev. Genet.,* vol. 5, pp. 101-113, Feb. 2004.

[2]   M. Ptashne, *A Genetic Switch: Phage Lambda Revisited, 3rd ed.* Cold Springs Harbor, New York: Cold Spring Harbor Laboratory Press, 2004.

[3]   K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak and J. J. Tyson, "Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle," *Mol. Biol. Cell,* vol. 11, pp. 369-391, Jan. 2000.

[4]   T. S. Gardner, D. di Bernardo, D. Lorenz and J. J. Collins, "Inferring genetic networks and identifying compound mode of action *via* expression profiling," *Science,* vol. 301, pp. 102-105, Jul. 2003.

[5]   M. Hecker, S. Lambeck, S. Toepfer, E. van Someren and R. Guthke, "Gene regulatory network inference: data integration in dynamic models-a review," *BioSystems,* vol. 96, pp. 86-103, Apr. 2009.

[6]   P. Brazhnik, A. de la Fuente and P. Mendes, "Gene networks: how to put the function in genomics," *Trends Biotechnol.,* vol. 20, pp. 467-472, Nov. 2002.

[7]   M. Gustafsson, M. Hörnquist, J. Björkegren and J. Tegnér, "Genome-Wide System Analysis Reveals Stable yet Flexible Network Dynamics in Yeast," *IET Syst. Biol.,* vol. 3, pp. 219-228, 2009.

[8]   N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann and M. Gerstein, "Genomic analysis of regulatory network dynamics reveals large topological changes," *Nature,* vol. 431, pp. 308-312, Sep. 2004.

[9]   M. Ptashne and A. Gann, *Genes & Signals.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 2002.

[10]   G. Stolovitzky, D. Monroe and A. Califano, "Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference," *Ann. N. Y. Acad. Sci.,* vol. 1115, pp. 1-22, Dec. 2007.

[11]   K.Y. Yip, R.P. Alexander, K.K. Yan and M. Gerstein, "Improved Reconstruction of *In Silico* Gene Regulatory Networks by Integrating Knockout and Perturbation Data", *PLoS One,* vol. 5(1): pp. e8121, 2010. [Online] Available: http://www.plos.org. [Accessed Aug. 5 2010].

[12]   M. Gustafsson and M. Hörnquist. "Gene Expression Prediction by Soft Integration and the Elastic Net - Best Performance of the DREAM3 Gene Expression Challenge", *PLoS One,* vol. 5(2):

p. e9134, 2010. [Online] Available: http://www.plos.org. [Accessed Aug. 5 2010].

[13] M. Helikar, N. Kochi, J. Konvalina and J.A. Rogers, "Boolean Modeling of Biochemical Network", *Open Bioinformatics*, vol. 5. pp. 16-25, 2011.

[14] M. Hörnquist, "Scale-free networks are not robust under neutral evolution," *Europhys. Lett.,* vol. 56, pp. 461-467, Nov. 2001.

[15] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artif. Intell.,* vol. 29, pp. 241-288, Sep. 1986.

[16] S. Kim, S. Imoto and S. Miyano, "Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data," *BioSystems,* vol. 75, pp. 57-65, July 2004.

[17] H. Zou and T. Hastie, "Regularization and variable selection *via* the elastic net," *J. R. Stat. Soc. Ser. B,* vol. 67, pp. 301-320, 2005.

[18] A. Scheinine, W. I. Mentzen, G. Fotia, E. Pieroni, F. Maggio, G. Mancosu and A. de la Fuente, "Inferring gene networks: dream or nightmare?" *Ann. N. Y. Acad. Sci.,* vol. 1158, pp. 287-301, Mar. 2009.

[19] M. Lauria, F. Iorio and D. di Bernardo, "NIRest: a tool for gene network and mode of action inference," *Ann. N. Y. Acad. Sci.,* vol. 1158, pp. 257-264, Mar. 2009.

[20] M. Gustafsson, M. Hornquist and A. Lombardi, "Constructing and analyzing a large-scale gene-to-gene regulatory network--lasso-constrained inference and biological validation," *IEEE/ACM Trans. Comput. Biol. Bioinform.,* vol. 2, pp. 254-261, Jul-Sep. 2005.

[21] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science,* vol. 286, pp. 509-512, Oct. 1999.

[22] U. Alon, "Network motifs: theory and experimental approaches," *Nat. Rev. Genet.,* vol. 8, pp. 450-461, Jun. 2007.

[23] U. Alon, *An Introduction to Systems Biology : Design Principles of Biological Circuits*, Boca Raton, Fla.: Chapman and Hall/CRC, vol. 10, 2007.

[24] A. Ma'ayan, G. A. Cecchi, J. Wagner, A. R. Rao, R. Iyengar and G. Stolovitzky, "Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 105, pp. 19235-19240, Dec. 2008.

[25] R. Steuer, "Computational approaches to the topology, stability and dynamics of metabolic networks," *Phytochemistry,* vol. 68, pp. 2139-2151, Aug-Sep. 2007.

[26] S. Flach and C. R. Willis, "Discrete breathers," *Physics Reports,* vol. 295, pp. 181-264, March 1998.

[27] A. Lesne, "Robustness: confronting lessons from physics and biology," *Biol. Rev. Camb. Philos. Soc.,* vol. 83, pp. 509-532, Nov. 2008.

[28] S. Nikolov, E. Yankulova, O. Wolkenhauer and V. Petrov, "Principal difference between stability and structural stability (robustness) as used in systems biology," *Nonlinear Dynamics Psychol. Life Sci.,* vol. 11, pp. 413-433, Oct. 2007.

[29] H. Kitano, "Computational systems biology," *Nature,* vol. 420, pp. 206-210, Nov. 2002.

[30] R. M. May, "Will a large complex system be stable?" *Nature,* vol. 238, pp. 413-414, Aug. 1972.

[31] S. Sinha and S. Sinha, "Evidence of universality for the May-Wigner stability theorem for random networks with local dynamics," *Phys. Rev. E. Stat. Nonline Soft Matter Phys.,* vol. 71, article number 020902, Feb. 2005.

[32] S. Sinha, "Complexity *vs.* stability in small-world networks," *Physica A,* vol. 346, pp. 147-153, Feb. 2005.

[33] R. K. Pan and S. Sinha, "Modular networks emerge from multiconstraint optimization," *Phys. Rev. E,* vol. 76, article number 045103, Oct. 2007.

[34] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K.

Gifford and R. A. Young, "Transcriptional regulatory networks in Saccharomyces cerevisiae," *Science,* vol. 298, pp. 799-804, Oct. 2002.

[35] E. Balleza, E. R. Alvarez-Buylla, A. Chaos, S. Kauffman, I. Shmulevich and M. Aldana, "Critical dynamics in genetic regulatory networks: examples from four kingdoms," *PLoS One,* vol. 3, pp. e2456, Jun. 2008.

[36] R. Singh. The contenders: Gripen JAS-39. Feb. 2007. [Online] Available: http://www.domain-b.com/aero/gripen_jas-39.htm [Accessed Aug. 5, 2010].

[37] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science,* vol. 296, pp. 910-913, May 2002.

[38] R. Albert, H. Jeong and A.L. Barabasi, "Error and attack tolerance of complex networks," *Nature,* vol. 406, pp. 378-382, July 2000.

[39] H. Jeong, S. P. Mason, A.L. Barabási and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature,* vol. 411, pp. 41-42, May 2001.

[40] H. Jeong, B. Tombor, R. Albert, Z. N. Oltavai, and A.L. Barabási, "The large-scale organization of metabolic networks," *Nature,* vol. 407, pp. 651-654, Oct. 2001.

[41] S. Maslov and K. Sneppen, "Computational architecture of the yeast regulatory network," *Phys. Biol.,* vol. 2, pp. S94-S100, Dec. 2005.

[42] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, "Network motifs: simple building blocks of complex networks," *Science,* vol. 298, pp. 824-827, Oct. 2002.

[43] M. Gustafsson, M. Hörnquist and A. Lombardi, "Comparison and validation of community structures in complex networks," *Physica A,* vol. 367, pp. 559-576, July 2006.

[44] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.,* vol. 25, pp. 25-29, May. 2000.

[45] U. Alon, "Biological networks: the tinkerer as an engineer," *Science,* vol. 301, pp. 1866-1867, Sep. 2003.

[46] N. Kashtan and U. Alon, "Spontaneous evolution of modularity and network motifs," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 102, pp. 13773-13778, Sep. 2005.

[47] M. Parter, N. Kashtan and U. Alon, "Facilitated variation: how evolution learns from past environments to generalize to new environments," *PLoS Comput. Biol.,* vol. 4, pp. e1000206, Nov. 2008. [Online] Available: http://www.plos.org. [Accessed Aug. 5 2010].

[48] M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 95, pp. 14863-14868, Dec. 1998.

[49] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller and N. Friedman, "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data," *Nat. Genet.,* vol. 34, pp. 166-176, Jun. 2003.

[50] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 99, pp. 7821-7826, June 2002.

[51] J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv and N. Barkai, "Revealing modular organization in the yeast transcriptional network," *Nat. Genet.,* vol. 31, pp. 370-377, Aug. 2002.

[52] X. Wang, E. Dalkic, M. Wu and C. Chan, "Gene module level analysis: identification to networks and dynamics," *Curr. Opin. Biotechnol.,* vol. 19, pp. 482-491, Oct. 2008.

[53] P. Langfelder and S. Horvath, "Eigengene networks for studying the relationships between co-expression modules," *BMC Syst. Biol.,* vol. 1, pp. 54, Nov. 2007.