# Structural and dynamical properties of cellular and regulatory networks

Statistical mechanics of cellular systems and processes

R. Sinatra
*Scuola Superiore di Catania, Via S. Paolo 73, 95123 Catania, Italy*

J. Gómez-Gardeñes
*Scuola Superiore di Catania, Via S. Paolo 73, 95123 Catania, Italy and*
*Institute for Biocomputation and Physics of Complex Systems (BIFI),*
*University of Zaragoza, 50009 Zaragoza, Spain*

Y. Moreno*
*Institute for Biocomputation and Physics of Complex Systems (BIFI),*
*University of Zaragoza, 50009 Zaragoza, Spain*
*To whom correspondence should be sent: yamir@unizar.es

D. Condorelli
*Department of Chemical Sciences Section of Biochemistry and Molecular Biology*
*University of Catania Viale A. Doria 6 95125 Catania Italy*

L. M. Floría
*Institute for Biocomputation and Physics of Complex Systems (BIFI), and*
*Departamento de Física de la Materia Condensada, Universidad de Zaragoza,*
*50009 Zaragoza, Spain*

V. Latora
*Dipartimento di Fisica e Astronomia, Università di Catania, and*
*INFN, Via S. Sofia 64, 95123 Catania, Italy*

# Contents

# 1

# Introduction

Systems that can be mapped as networks are all around us. Recently, scientists have started to reconsider the traditional reductionism viewpoint that has driven science ever since. The accumulated evidence that systems as complex as a cell cannot be fully understood by studying only their isolated constituents, but that rather most of biological characteristics and behaviours are related to complex interactions of many cellular constituents, has given rise to the birth of a new movement of interest and research in the study of complex networks, i.e. networks whose structure is irregular, complex and dynamically evolving in time, with the main focus moving from the analysis of small networks to that of systems with thousands or millions of nodes, and with a renewed attention to the properties of networks of dynamical units. This flurry of activity has seen the physicists' and biologist's communities among the principal actors, and has been certainly induced by the increased computing powers and by the possibility to study the properties of a plenty of large databases of real networks. The regulatory and cellular networks, that would be the subject of study in this chapter, have been among the most studied networks, and the field has benefited from many important contributions. The expectancy is that understanding and modeling the structure of a regulatory network would lead to better comprehend its dynamical and functional behavior. In this chapter, we aim to provide the reader with a glance at the most relevant results and novel insights provided by network theory in this field, discussing both the structure and the dynamics of a number of regulatory networks.

# 2

# Structure

Though in the last years many experimental techniques have been improved and larger and larger amounts of data are available, the determination and construction of the different cellular networks is not an easy task. Nevertheless, many complex interactions which take place at the cellular level, such as metabolic chains, protein - protein interactions and gene - gene regulations, can be represented and then studied using the formalism of graph theory, where each system is reduced to a set of nodes and edges, sometimes also called vertices and links. A node usually represents a cellular constituent, like a protein or a gene. Nodes are linked by edges. An edge can be directed or undirected, whether it has a direction or not (it "goes out" of a node and "gets into" another). The meaning of edges changes case per case: in the cellular field, for instance, it could represent a reaction or an interaction between the nodes it is linking. In the next paragraphs, we describe in more details some networks of fundamental importance for the cell survival. While explaining the meaning of nodes and edges in every particular context, we will also give some basic definitions of graph theory (see Figures 2.1-2.3), in order to better understand the biological roles of graph constituents and the dynamical properties which will be discussed in the last part of this chapter.

## 2.1 Metabolic networks

A metabolic network is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. As such, these networks comprise the chemical reactions of metabolism as well as the regulatory interactions that guide these reactions. These are accelerated, more accurately catalyzed, by enzymes. Therefore a complete metabolic network is constituted by three kinds of nodes: metabolites, re-

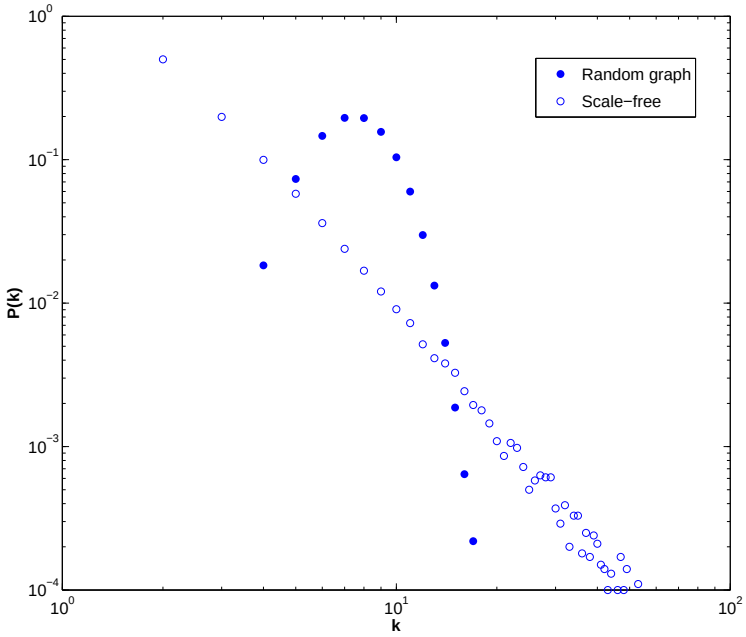


Fig. 2.1. An important graph property is the degree distribution function $P(k)$, that describes the probability to find a node with $k$ edges. A *random graph* is constructed by randomly linking $N$ nodes with $E$ edges, and has a Poissonian degree distribution $P\left(k\right) = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!}$. That means that the majority of nodes have a degree close to the average degree $\langle k\rangle$. A *scale-free graph* is instead characterized by a power-law degree distribution $P(k) = Ak^{-\gamma}$, usually with $2 < \gamma < 3$. A power-law distribution appears as a straight line in a double-logarithmic plot. In a scale-free graph, low degree nodes are the most frequent ones, but there are also a few highly connected nodes, usually called *hubs*, not present in a random graph.

actions and enzymes, and by two types of edges representing mass flow and catalytic regulations. The former kind of edge links reactants to reactions and reactions to products, while the latter connects enzymes to the reactions they catalyze. Needless to say, all these edges are directed.

The above representation is not always the most suitable. In fact, it sometimes implies the assumption that some virtual intermediate complexes take place. To avoid subjective assumptions in the way the network is built up, Wagner and Fell, for instance, have deduced two distinct networks from the metabolic pathways of *Escherichia Coli*, which do not imply the definition of virtual intermediate complexes. In one of the networks they define, metabolites stand for nodes which are linked to undirected edges if they participate

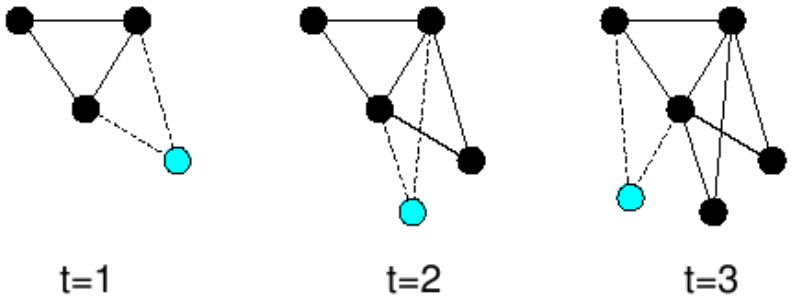


Fig. 2.2. A simple model to grow networks with a power-law degree distribution was proposed by Barabasi and Albert (1999). The model is based on two main ingredients, *growth* and *linear preferential attachment*. That means that the graph grows during time by the addition of new nodes and new links, and that links are not distributed at random, but the probability of connecting to a node depends on the nodes' degree. The algorithm to construct a network starts at time $t = 0$ with a complete graph of $m_0$ nodes (in the example, $m_0 = 3$). Then, at each time step $t$ a new node $n$ is added. The new node has $m \leq m_0$ edges (in our case, $m = 2$), linking $n$ to $m$ different nodes already present in the system. When choosing the nodes to which the new node $n$ connects, it is assumed that the probability $\Pi_{n\to i}$ that $n$ will be connected to node $i$ is linearly proportional to the degree $k_i$ of node $i$, i.e.: $\Pi_{n\to i}(k_i) = \frac{k_i}{\sum_l k_l}$. For large times (or $N$), this corresponds to a graph with a stationary power-law degree distribution with exponent $\gamma = 3$.

in the same reaction. Another network is instead constituted by nodes that represent metabolic reactions linked when sharing a metabolite. The networks are respectively named metabolite network and reaction network. Both these networks exhibit a power-law degree distribution (see Figs. 2.1 and 2.2) and small-world properties (Fig. 2.3). Surprisingly, when considering the whole metabolic pathways of organisms that have evolved differently and consequently show many differences, the metabolic networks share the same topological and statistical properties, namely, those corresponding to scale-free graphs (Jeong et al. (2000)).

## 2.2 Protein - protein interaction networks

The interactions between proteins are crucial for many biological functions. For example, signals from the exterior of a cell are mediated to the inside of that cell by protein-protein interactions of the signaling molecules (Fig. 2.4). This process, called signal transduction, plays a fundamental role in many biological processes. Proteins might interact for a long time to form part of a protein complex, a protein may be carrying another protein, or a protein may

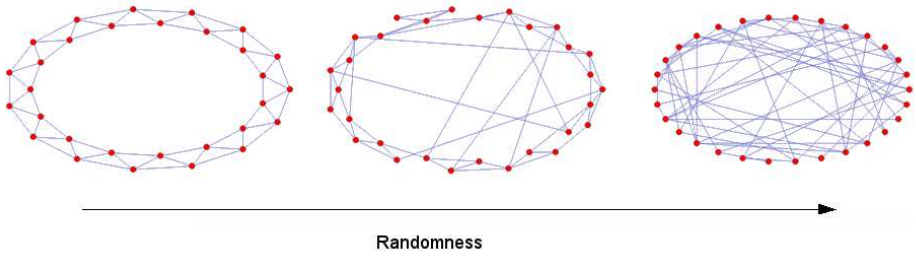


Fig. 2.3. Small-world networks, as defined by Watts and Strogatz (1998), have intermediate properties between regular lattices (such as the first graph in the figure) and random networks (such as the last graph in the figure). A regular lattice has high clustering but also a large average path length, while a random graph is characterized by a short path length together with a low clustering. A small world network (in the middle in the figure) borrows a high clustering coefficient from the former and a short average path length from the latter.

interact briefly with another protein just to modify it. This modification of proteins can itself change protein-protein interactions. Therefore, protein-protein interactions are of central importance for every process in a living cell.

In a protein interaction network, nodes represent proteins while an undirected edge is drawn between two proteins when they physically interact. Though the data may be incomplete and contain a very high number of false positives, the results obtained from databases with very small overlap between them show the same network properties: scale-freeness, high-clustering and small world properties. These topological properties have already been exploited as the network approach allows to look at the system from new points of view and to borrow tools from other fields to solve (or at least to give alternative solutions to) known open problems. For example, this is the case of the *Saccharomyces Cerevisiae* network widely studied in the literature. In particular, using the protein-protein interaction network representation of this organism (Uetz et al. (2000)), it has been possible to suggest or, at least, to guess the function of many unclassified proteins (Vazquez et al. (2003a)).

## 2.3 Gene regulatory networks

Some achievements in experimental techniques during the last few years, like gene chips and microarray, have paved the way to the study of the so-called gene regulatory networks. At the cellular level, the production and degradation of all proteins is supervised by the gene regulatory network,

constituted by those pairs of genes whose products in proteins is so that the former regulates the abundance of the second. That is the case, for instance, of the transcription regulatory network which is also the most studied one. It is well known that the transcription of genes from DNA to RNA is regulated by some particular proteins which are called transcription factors (see Fig. 2.4). These proteins are the products of some genes.

Therefore, the nodes of the network are genes, namely, the DNA sequences which are transcribed into the mRNAs that translate into proteins, while edges between nodes represent individual molecular reactions, the protein-protein, protein-mRNA, and protein-DNA interactions through which the products of one gene affect those of another. These interactions can be inductive, with an increase in the concentration of one leading to an increase in the other, or inhibitory, with an increase in one leading to a decrease in the other.

The method to build the graph representing a gene regulatory network is mostly based on genome-wide gene expression data. Agrawal (2002) has suggested an algorithm to build a gene regulatory network starting from a microarray gene data. For a given number $N$ of genes, data can be represented in a matrix $N \times D$, where $D$ is the number of sampling conditions where the expression level appear. Looking at the rows as vectors, an Euclidean distance among genes can be defined. Each gene is then linked to his $K$ nearest neighbors, where the Euclidian distance has been used to determine them. Many networks are then constructed by treating $K$ as an order parameter.

## 2.4 Overall Leitmotif

What is extremely interesting is that all the biological networks we have so far briefly described, although being very different one from the other, share many topological features. Each of them shows not only a power-law degree distribution, but also high clustering coefficient (numbers of triangles present in the network, or, in other words, the probability that if A is linked to B and B is linked to C, A is also linked to C) and short mean path length (every couple of nodes can be connected with a path of only few links) which lead to the so-called small-worldness.

These characteristics are present not only for networks in the biological field, but also in other complex systems, going from technological to social networks. Such a universality in systems so far one from the other is quite intriguing and suggests to focus on which meanings all these topological features hide.

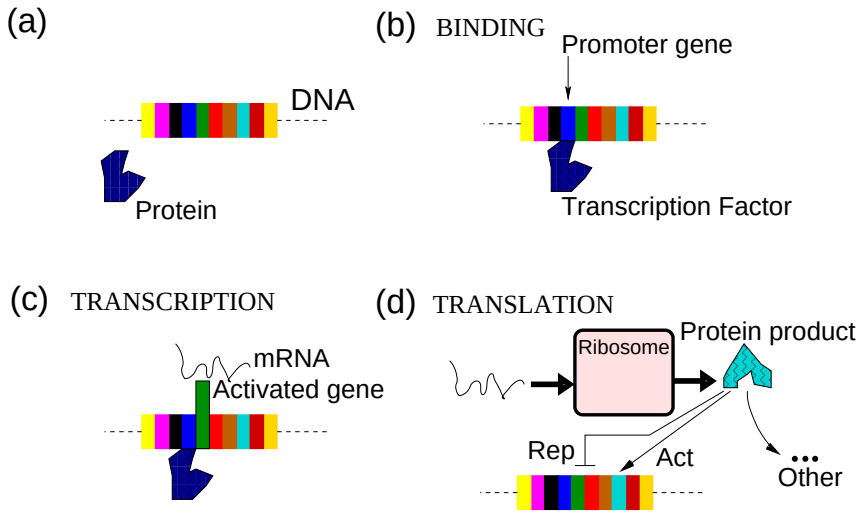


Fig. 2.4. The different stages of gene expression. The basic ingredients are the proteins and specific gene regions in the DNA, such as promoters and transcribed sequences (panel (a)). A specific protein binds to a part of the DNA sequence called the promoter, the protein is known as the transcription factor since it starts the transcription of the genetic information encoded at the specific gene that the complex promoter + transcription factor regulates (panel (b)). After the genetic information is transcribed into the messenger RNA, by RNA polymerase (panel (c)), it is subsequently translated into proteins at the ribosomes (panel (d)). The protein product that emerges after this process can act either as another transcription factor for the expression of other genes or as a repressor of the activity of other genes stopping the synthesis of their protein products. Another possibility is that this protein product participates in the physiological processes of the cell and form protein complexes as enzymes.

The power-law degree distribution involves the presence of highly connected nodes, called *hubs*, even if the small-degree ones are the most abundant (see Figs. 2.1 and 2.2). The existence of hubs seems to be correlated with evolutionary processes. The hubs should represent the oldest cellular constituents, to which new nodes, generated by gene duplication processes, preferentially attach. This growing mechanism has been shown to explain the topology of protein-protein interaction networks (Vazquez et al. (2003b)), but with proper adjustments it seems to explain the scale-free features of regulatory and metabolic networks as well. However, it is quite clear that highly connected nodes are subjected to severe selective and evolutionary constraints and that the cell is vulnerable to the loss of highly interactive hubs, which can result in the breakdown of the network into isolated clusters. A famous example of a hub protein is the tumor suppressor

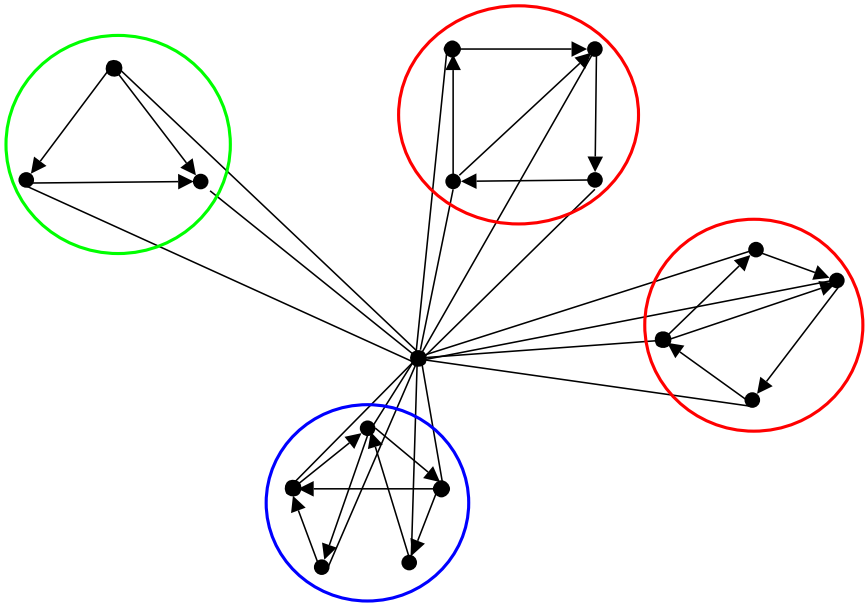

Fig. 2.5. A motif of a graph is a connected subgraph of $n$ nodes which appears over represented if compared to a graph of the same size, number of edges and degree distribution, but with randomized connections. Motifs can group to form clusters or sometimes can interact one with each other if linked to a common hub.

protein p53, which Vogelstein et al. (2005) have demonstrated to be inactive in half of human tumors.

On the other hand, the high clustering feature is related to the existence of modules (Fig. 2.5). Networks are composed by subgraphs of highly interconnected groups of nodes, usually called motifs or modules. Formally, a motif is a connected subgraph of $n$ nodes which appears more frequently than in a graph with the same size, number of edges and degree distribution, but with randomized links. Each real network is characterized by its own set of distinctive motifs. It has been suggested that motifs have specific functions as elementary circuits. For example, in regulatory networks feed-forward loops appear more frequent than expected from randomly connected graphs, while in protein-protein maps there is a high abundance of completely connected subgraphs and short cycles.

The molecular components constituting a motif not only interact with the elements of that motif but can be linked to other motifs giving rise to clusters and modules at a larger scale which are still interconnected to each other. Moreover, the presence of hubs makes the existence of relatively isolated modules unlikely, which ultimately gives to cellular networks a hierarchical

topology. Admittedly, there is a high degree of overlap and crosstalk between modules with small modules forming cohesive communities. Interestingly, these structural modules correlate very well with functional ones, thus providing a way to study systematically the structure-function relationship.

# 3

# Dynamics of Regulatory Networks

After the structural characterization of interaction maps between genes, protein and metabolites the following question turns up into scene: *What is the relationship between the structure of interactions observed in most biological networks and their task-performing ability?* In order to answer this question and shed new light on what is going on at the cellular and molecular levels of organization of biological systems scientists have begun to look for the dynamical evolution of the activity patterns of the constituents of such biological networks. In fact, during the last several years, the available amount of experimental data, obtained with technological advances such as cDNA microarrays, has exploded. This has allowed to face the dynamical characterization of diverse biological processes both on a genome-wide and on multi-gene scales and with fine time resolution. On the other hand, despite the advances in biological engineering, the formulation of compelling models on the dynamics governing metabolic and genetic processes is still a hard issue because the observed dynamical patterns are highly nonlinear and one needs to deal with many degrees of freedom for a proper description of regulatory mechanisms.

Although the ultimate goal of systems biology is to describe the cellular processes as a whole by means of a global biochemical regulatory network, the three levels of description (gene expression, protein interaction and metabolic fluxes) are usually studied separately (as we have seen in the previous section) also when studying the dynamics of interactions. The reason behind this compartmentalization of the cellular system is the diverse ability for profiling genes, proteins or metabolites. While current techniques for measuring genome-wide differential gene expression are nowadays widespread, this is not the case for the current methods used to deal with proteins and metabolites. In principle, one cannot get rid of the two higher organization levels of a cell, however, a number of relevant processes of cell

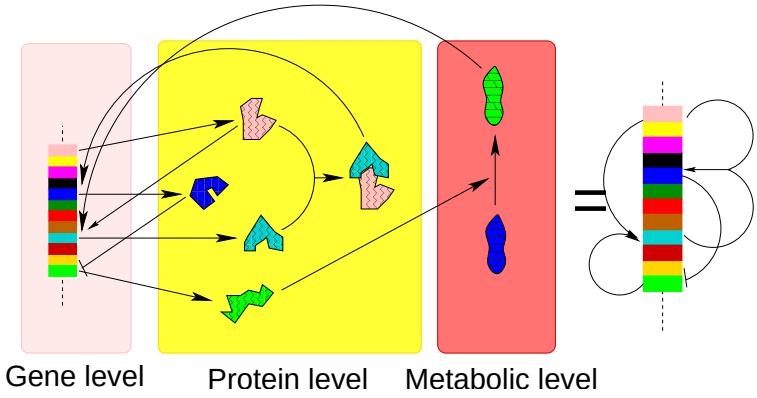


Fig. 3.1. Coarse graining of cellular interactions into a single gene network. The three levels of description (genes, proteins and metabolism) and the interactions between their constituents are embedded on a single map of interactions between genes.

physiology can be mapped into a gene network coarse-grained view of cells (see Figure 3.1). In this part of the chapter we will focus on the current models used to describe the dynamics of gene regulatory networks.

As in other scientific fields, the work on the characterization of gene-gene regulation started by looking at the basic mechanisms and building blocks of the entire biological system. Following this constructionist scheme, concepts such as *operon*, *regulator gene* and *transcriptional repression* were first introduced in the literature by Jacob and Monod (Jacob and Monod (1961)). Their model has settled the basis for more elaborated models as different regulatory mechanisms have been discovered (Wall et al. (2004)). Here, after discussing the basis of the dynamical models, we will also move from small gene circuits (so-called modules), where predictive their power is high, to large scale gene networks, where the goal is to model the global functioning of cells. This bottom-up approach is aimed at addressing several issues of relevance in cellular processes, first on the small scale to analyze the robustness of small circuits under external (environmental) perturbations, and second on large scale networks to test the ability of groups of genes to perform different coordinated tasks.

## 3.1 Mathematical formulation: Boolean models versus Differential equations

Regulatory mechanisms among genes can be translated into mathematical language in various ways. The appropriate choice of the dynamical equations

will depend on the level of description required. In this sense, large scale gene regulatory networks are usually described by a simple mathematical framework that makes use of Boolean functions. On the other hand, when one is interested in describing simple regulatory mechanisms that involve few genes more detailed models such as nonlinear differential equations are best suited. The use of each type of description thus depends on the sum of complexities regarding the structure and the dynamics. In the following we describe the essential ingredients of both mathematical approaches.

### *3.1.1 Boolean modeling of regulatory networks*

Boolean models are based on the assumption that genes can be found in one of a discrete set of states, and account for the different kinds of interactions that appear in gene regulatory networks by means of simple rules. Besides their simplicity, the success of Boolean models relies on the inherent difficulty in obtaining an accurate functional form of the reaction kinetics associated to every gene-gene interaction. Boolean dynamics has been widely used to analyze the importance that the global topological features of a gene network (such as path redundancy or abundance of loops, average number and sign of regulatory inputs, etc...) have on its dynamical organization.

In the usual Boolean framework, a gene $i$ at time $t$ can be in two possible dynamical states: *active* ($g_i(t) = 0$) or *inactive* ($g_i(t) = 1$). The activity of a gene depends on the state of those genes from which it receives a regulatory input (*i.e.* incoming link of the regulatory network). Besides, time is considered as a discrete variable so that at each time step the activity level of every gene $i$ is updated considering its $k_i$ input signals

$$g_i(t + \tau) = f_i(g_{j_1}(t), ..., g_{j_{k_i}}(t)) \; . \tag{3.1}$$

The updating process of the whole network can be synchronous (parallel updating) or asynchronously (sequential updating). The specific form of every function, $f_i$ is constructed by following the specific interactions that gene $i$ receives from its regulators. These functions are always combinations of the basic ("AND", "OR" and "NOT") logical operators so that the results can be either 1 if the statement is true or 0 if it is false. These single functions are expressed by means of truth tables as shown in Figure 3.2.

Once the network and the specific Boolean functions governing the interactions are set, the study focuses on the possible dynamical behaviors. Starting from different initial conditions (in principle $2^N$ different possibilities) one computes how many different final states are reached. There are three possibilities, namely, *(i)* the system becomes frozen in a unique dynam-

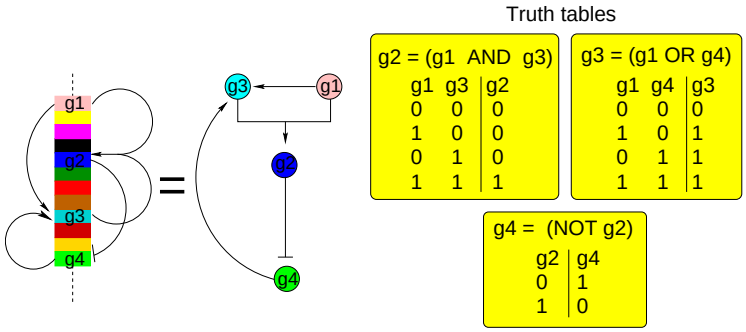


Fig. 3.2. Translation of the regulatory genetic map of Figure 3.1 into Boolean regulatory functions. The three Boolean relations for genes $g_2$, $g_3$ and $g_4$ make use of the basic logical operators "AND", "OR" and "NOT", respectively.

ical state (fixed point of the dynamics), *(ii)* the system explores cyclically a set of states ending in a periodic attractor of a length given by the number of different configurations explored, *(iii)* the dynamics is chaotic and the system explore different configurations without any periodicity. In general, different attractors can coexist in the configuration space, each of them with its own basin of attraction of initial conditions. As an example, we show in Figure 3.3 the typical representation used to characterize the configuration dynamics of a simple regulatory network composed of three nodes, the transition between the different dynamical states reveals one periodic attractor of length 5 and the fixed point $(0, 0, 0)$.

In a large regulatory network, one expect to obtain many different types of coexisting dynamical states. This hypothesis is biologically based on the fact that different network states correspond to different cell types, *i.e.* cells with the same genome developing different functions. Although biochemical data from real gene networks have become available only in recent years, the search of network topologies sustaining a large number of different dynamical states started long time ago with the pioneering work by S.A. Kauffman (Kauffman (1969)). Kauffman considered a random assignment of the boolean functions that governs the dynamical evolution of the gene's activity so that all the nodes receives a constant number of $K$ input signals from other nodes and these signals are activatory or inhibitory with equal probability. The main goal was to analyze the dependence of the type, number and length of dynamical attractors with the system size $N$ and the number of inputs $K$. The results tell that, for $K > 2$ the dynamics is mainly chaotic, the number of cycles scales with the number of genes, $N$,

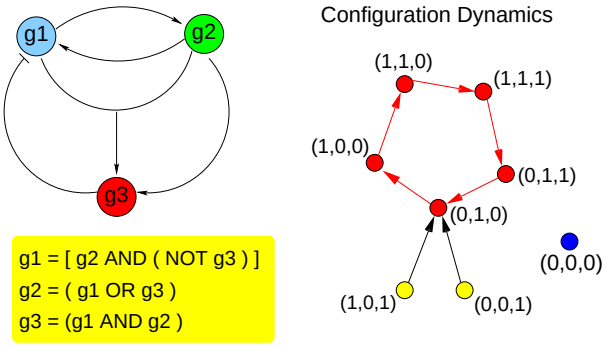


Fig. 3.3. Small regulatory network and its configuration dynamics. The network is composed by 3 genes that interact following the logical rules shown in the box. The configuration dynamics is represented by a network whose nodes are all the possible dynamical states and the directed links are the transitions from one to another (as dictated by the Boolean dynamics). It is shown that a cycle of length 5 (red nodes) and an steady state (blue node) exist. Yellow nodes are in the basin of attraction of the periodic attractor.

and their length scale exponentially with $N$. On the other hand, for the case $K = 1$ the dynamics is frozen and the number of attractors scales exponentially with $N$. Finally, the regime $K = 2$ is the most interesting since both the number and length of attractors scale as $\sqrt{N}$. These findings are very relevant biologically since the cell diversity of a living organism scales approximately with the square root of the gene number, thus pointing out that gene regulatory networks should operate just on the border between frozen and chaotic dynamics, *i.e.* $K_c = 2$. If one breaks the symmetry between activatory and inhibitory inputs the critical value of the network connectivity fulfills the relation

$$2\rho(1-\rho)K_c = 1, \tag{3.2}$$

where $\rho$ is the fraction of activatory inputs.

There has been a burst of research on Kauffman networks in the last thirty years in order to redefine models and make them more accurate (see e.g. (Glass (1975); Derrida and Pomeau (1986); Socolar and Kauffman (2003); Samuelson and Troein (2003); Kauffman et al. (2003); Drossel et al. (2005)) Perhaps, the most important refinement from the network perspective is to abandon the hypothesis of constant number of node's inputs and move to heterogenous networks, *i.e.* scale-free Boolean networks. In this regard, one can consider the value $K$ as the mean value of the number of inputs of a complex network and, in particular, re-express it as a function of the exponent $\gamma$ of the power law degree distribution. The result (Aldana

and Cluzel (2003)) is that an exponent $\gamma > 2.5$ assures robust behavior (*i.e.* the absence of chaotic attractors) of network dynamics. Moreover, the numerical exploration (Fox and Hill (2001)) of the phase space in the regime where chaotic dynamics exist, indicates that the number of observed chaotic attractors is smaller in scale-free networks than in networks with Poisson or delta degree distributions. This result could in principle relate the ubiquity of scale-free networks in regulatory systems to an evolutionary drift towards dynamical robustness.

Synthetic Boolean regulatory networks, although being idealizations, have served as test-beds for the mathematical models that are currently used on large real regulatory networks as we will see in 3.2.2. Besides, most of the results found for the Boolean approximation are robust when moving to more refined piece-wise linear or nonlinear models. On the other hand continuous models do not allow the computation of large-scale statistical properties of their dynamical behavior.

### *3.1.2 Modeling through differential equations*

Now we turn our attention to models where both time and concentrations are modeled as continuous quantities. In this case, the dynamics of the concentration of biochemical products evolves in time following the differential equation

$$\frac{d[x_i]}{dt} = f_i([x_{i_1}], [x_{i_2}], ..., [x_{i_n}]) - \gamma_i [x_i] \ , \tag{3.3}$$

where $[x]$ denote the concentration of product $x$ in units of $\#moles/volume$. The second term in the right hand side of equation (3.3) accounts for the degradation of $x_i$, being $\gamma_i$ the degradation rate parameter. The functional form of $f_i$, *i.e.* the rate of the reaction that produces $x_i$, is dictated by the reaction kinetics of the chemical processes at work. This makes extremely difficult to obtain a general framework since interactions of a gene, its products, and its regulators are reaction-specific, in fact a whole expression should contain various biochemical processes as reaction reversibility, product dimerization, enzyme-catalysis, etc...

The most simple examples of gene regulation are those of a gene $x$ regulated by a transcription factor $y$ acting either as a repressor or as an activator. In these cases it is easy to show (Alon (2007)) that the input functions to be included in Eq. (3.3), are respectively for repression and activation:

$$f_x([y]) \ = \ \frac{\beta}{1 + K^{-n}[y]^n} \qquad (Repression) \ , \tag{3.4}$$

$$f_x([y]) = \frac{\beta[y]^n}{1+K^{-n}[y]^n} \qquad (Activation)\ . \tag{3.5}$$

In both expressions, $[x]$ is the concentration of mRNA transcribed by gene $x$, $\beta$ is the maximal rate of mRNA transcription so that these expressions can be read as $\beta$ times the probability that the gene promoter DNA region is either free (occupied) by the repressor (activator) transcription factor $y$. The parameter $K$ is the dissociation constant of the reaction that describes the binding of $y$ to the gene promoter in DNA. The value of $n$ (so-called Hill coefficient) accounts for the number of transcription factor subunits that binds to the promoter. The repression input function shows that the gene activity decrease to zero as the repressor concentration grows. On the other hand the activatory input is a growing and saturable function, and when $n = 1$, it takes the expression of the Michaelis-Menten equation ubiquitously found for a wide range of biological processes such as enzyme kinetics (Sethna (2006)).

These mechanisms of negative or positive regulation of a gene can be closed in a negative or positive feedback loop if we consider that the transcription factor that regulates gene activity is the same as the protein product of mRNA translation. In this case a second differential equation

$$\frac{d[y]}{dt} = \alpha[x] - \gamma_y[y]\ , \tag{3.6}$$

accounts for the rate production of protein $y$ regulated by the activity of gene $x$. This equation has to be solved together with Eq. (3.3) with the input function given in Eq. (3.4) or (3.5). The solution to the feedback loop depends on whether there is activation or inhibition of $y$ to $x$ as it is shown in Figure 3.4. For the negative regulation a unique steady state exists and it is linearly stable meaning that any small perturbation of the system will return to the original state (homeostasis). On the other hand, when positive regulation occurs three possible steady states are possible, with only two of them being stable. In this case the system can choose between two different cell states.

The above modeling for the gene activity level can be extended to the more realistic case when there is more than one transcription factor regulating its expression. What are the effects of $a$ activators and $r$ repressors coordinated in the gene dynamics? In the simple case where different proteins can bind all together in the promoter region and proteins complexes are not formed after translation, one must consider all the possible configurations for the complex *(Gene Promoter + Transcription Factors)* that allow gene expression and sum up their associated probabilities. A first order approximation to the

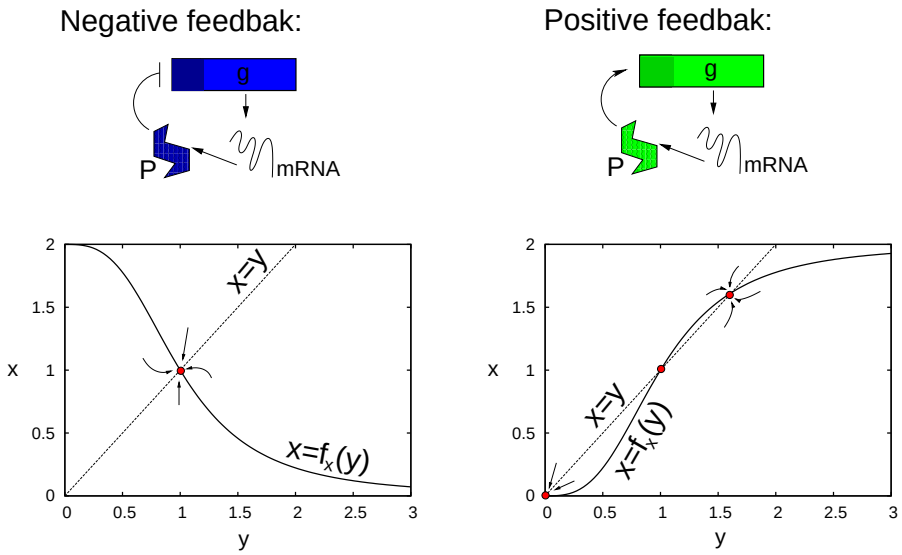


Fig. 3.4. Negative (left) and positive (right) feedback loops. Both systems describe a gene encoding its own transcription (repressor or activator) factor. Solving graphically the two systems of coupled differential equations, Eqs. (3.4)-(3.6) and Eqs. (3.5)-(3.6), one can compute the steady states of the system. In the first case, dynamics of mRNA and protein concentrations reaches a unique stable fixed point. For the positive loop there are three steady states, being stable (and thus biologically reliable) two of them (one corresponding to the rest state of the system). We have set in both cases $\beta = 2$, $n = 3$ and $\gamma_x = \gamma_y = \alpha = K = 1$.

general formula is given by this general input gene function (Alon (2007)),

$$f_x([y_1], ..., [y_a], [y_{a+1}], ...., [y_{a+r}]) = \frac{\sum_{i=1}^{a} \beta_i ([y_i]/K_i)^{n_i}}{1 + \sum_{i=1}^{a+r} \beta_i ([y_i]/K_i)^{n_i}} , \qquad (3.7)$$

where we have ordered the arguments of $f_x(\mathbf{y})$ so that the first $a$ variables correspond to activators and the remaining $r$ are repressors. A more detailed construction (where *e.g.* protein dimerization is taken into account) of a mean-field model for gene regulatory inputs can be found in (Andrecut and Kauffman (2006)).

Although the above mathematical setting is only suited when small network circuits or modules (see section 3.2.1) are analyzed, one can relax the rigidity of regulatory input functions and construct more general equations that incorporate the main features of regulatory dynamics as the saturable character of the gene response under activatory inputs. In reference (Gómez-Gardeñes et al. (2006)) the authors make a coarse-grained formulation of continuous time gene dynamics by writing the following equations for mRNA

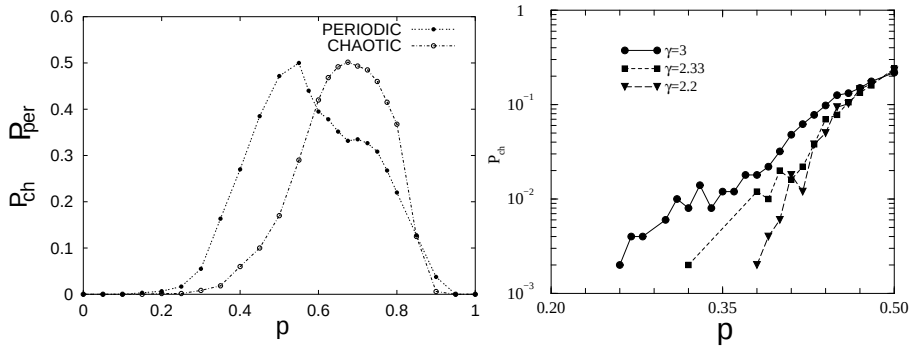


Fig. 3.5. (Left panel) Probabilities that an arbitrary initial condition ends in a periodic, $P_{per}$, and chaotic attractor, $P_{ch}$, as functions of the fraction of inhibitory inputs in the network,$p$. Obviously when $p = 1$ only zero-activity states are achieved thus both probability tend to zero as $p \to 1$. The results show that there is a region where dynamical order (periodic states or steady states) prevails on the network dynamics, and that the threshold of order (no chaotic states) around $p_c \simeq 0.26$. This threshold (right panel) seems to grow slightly when the exponent $\gamma$ of the power law degree distribution decreases (and so the degree heterogeneity in the network grows). After Gómez-Gardeñes et al (2006).

concentrations,

$$\frac{d[x_i]}{dt} = -[x_i] + \beta \frac{\Phi\left(\sum_{i=1}^{N} W_{ij}[x_j]\right)}{1 + K^{-1}\Phi\left(\sum_{i=1}^{N} W_{ij}[x_j]\right)} , \tag{3.8}$$

where $\mathbf{W}$ is the interaction matrix whose entries are $W_{ij} = 1$ if product of gene $j$ activates expression of gene $i$, $W_{ij} = -1$ if product of gene $j$ inhibits gene $i$, and $W_{ij} = 0$ if no regulatory interaction is found. Besides $\Phi(x)$ is defined as $\Phi(x) = x$ if $x > 0$ and $\Phi(x) = 0$ otherwise. This compact and simple form for regulatory continuous dynamics paves the way for a thorough study of two essential ingredients in biological regulatory networks: saturability of the interactions and scale-free character of the interconnections among constituents.

Inspired by the aforementioned works in Kauffman Boolean networks synthetic Barabási-Albert scale-free networks are constructed (see Fig. 2.2) and the sign of every interaction is assigned so that the fraction of inhibitory inputs is equal to $p \in [0, 1]$. The (continuous) phase space of the systems is then explored by analyzing the final dynamical attractor of many different initial conditions. As in the studies of Boolean synthetic networks, the

results here point out that different regimes (steady states, periodic and chaotic attractors) are possible depending on the value of $p$ (which plays here the role of $\rho$ in Boolean networks) as seen in Figure 3.5.

In addition, there are other dynamical features intrinsically related to the continuous character of the equations. In particular, large networks can dynamically fragment so that topologically disconnected subsets of nodes sustain independent dynamics (*e.g.* steady states of non-zero activity and periodic dynamics) while the rest of network's nodes remain in the rest state. The observation of such clusters of active nodes provides an emergent network topology that is defined by those connected nodes sharing the same kind of dynamics and the links between them. The topological analysis of these dynamical clusters reveals remarkable differences from the substrate scale-free network topology such as the existence of a high clustering coefficient. This latter result is much in agreement with the topological analysis of real biological networks, and points out that the observed topology is a result of functional (dynamical) relations between elements and therefore interpretations about the origin of network patterns should not disregard dynamical analysis. In the next section we will see how small network subunits (analogous to these dynamical clusters) with robust and coherent dynamics, termed motifs, have attracted a lot of attention when studying real regulatory networks.

## 3.2 Dynamics of real regulatory networks

We now focus on applications to modeling real gene regulatory networks. As discussed in the previous section, an important issue concerns the most convenient level of description for a particular network (Bornholdt (2005)). While network subunits or modules can be modeled in terms of differential equations, a description of an entire regulatory system need to rely on the coarse-grained picture of logical or Boolean dynamics.

### *3.2.1 Network modules*

Network modules are presented in the literature as small circuits (composed typically of three genes) embedded in large regulatory networks that are able to display autonomous dynamics. Since it is difficult to detect modules by simply looking at the whole network activity, several works have focused on the identification of general building blocks in gene networks by looking for motifs in the network topology (Milo et al. (2002)). Motifs, as previously seen in section 2.4, are those subgraphs whose occurrence in the real network

is significantly higher than in their randomized versions and include auto-regulatory excitatory feedback loops, inhibitory feedback loops, feed-forward loops and dual positive-feedback loops. Once network motifs are identified, different test of dynamical robustness or reliability (*e.g.* synchronizability) in different regulatory networks can be performed (Klemm and Bornholdt (2005); Bradman et al. (2005); Ma'ayan et al. (2005); Lodato et al. (2007)). When regulatory dynamics is implemented, the result is that network motifs show a more robust behavior than other circuits, with the same number of genes, that are not so frequent in the regulatory map. As a consequence, the experimental occurrence of particular structures of regulatory interactions seems to be due to their remarkable dynamical reliability. This suggest a selective process acting on the *pattern of interactions* rather than on *isolated* genes.

The finding that a few basic modules are the building blocks of large real regulatory networks justifies the design and construction of small synthetic regulatory circuits to implement particular tasks. The most salient example of a synthetic gene network is the "repressilator" that has become one of the best studied model systems of this kind. The repressilator is a network of three genes, whose products (proteins) act as repressors of the transcription of each other in a cyclic way (see Figure 3.6). This synthetic network was implemented in the bacterium *Escherichia coli* so that periodically it induces the synthesis of a green fluorescent protein as a readout of the repressilator state (Elowitz and Leibler (2000)). In this regard, the temporal fluctuations in the concentration of each of the three components of the repressilator can be easily reproduced by analyzing a system of six ordinary differential equations, based on Eqs. (3.3), (3.4) and (3.6)), read

$$\frac{d[x_i]}{dt} = -[x_i] + \frac{\beta}{1+[y_j]^n} , \quad (mRNA\ \ dynamics) \tag{3.9}$$

$$\frac{d[y_i]}{dt} = -\alpha([y_i] - [x_i]) , \quad (Protein\ \ dynamics) \tag{3.10}$$

where the couples $(i, j)$ asumme the values $(1, 3)$, $(2, 1)$ and $(3, 2)$. The variable $[x_i]$ is the mRNA concentration encoded by gene $x_i$, and $[y_i]$ is the concentration of its translated protein $y_i$. The parameter $\alpha$ is the ratio of the protein decay rate to the mRNA decay rate, and time has been rescaled in units of the mRNA lifetime. This system of equations has a unique steady state which can be stable or unstable depending on the parameter values. In the unstable region of parameter space, the three protein concentrations fluctuate periodically. Experiments show the temporal oscillations of fluorescence, which were checked to be due to the repressilator, validating the

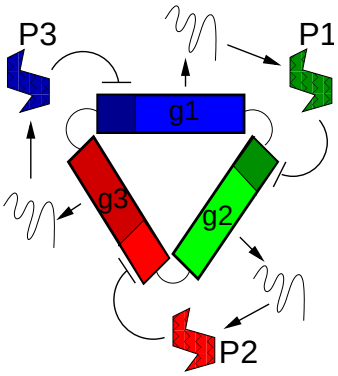


Fig. 3.6. Schematic representation of the *repressilator*. The repressilator is a small network composed of three genes $g_i$ $(i = 1, 2, 3)$, each one inhibiting the activity of the subsequent. That is, the protein product of genes $g_1$, $g_2$ and $g_3$ act as the repressors of the activity of gene $g_2$, $g_3$ and $g_1$ respectively.

model predictions. In particular, the previous mathematical model served to identify possible classes of dynamical behavior and to determine what experimental parameters should be adjusted in order to obtain sustained oscillations.

The repressilator is an illustrative example of the experience gained by identifying network modules and modeling its dynamical behavior in real networks. Not surprisingly, the repressilator called attention from experts on (biological) synchronization, for it offers good prospectives for further insights into the nature of biological rhythms, whose mechanisms remain to be understood. In this respect, a simple modular addition of two proteins to the repressilator original design has been recently proposed (Ojalvo et al. (2004)). This extension is made so that one of the new proteins can diffuse through the cell membrane thus providing a coupling mechanism between cells containing repressilator networks. This inter-cell communication couples the dynamics of the different cell oscillators (with different repressilator periods) and thus allows the study of the transition to synchronization of coupled phase oscillators in a biological system. The result reproduces the phase transition from uncorrelated to coherent dynamics as the cell dilution decreases (increasing cell-cell interaction).

### *3.2.2 Large regulatory networks*

The study of elementary gene circuits certainly provides answers to intriguing questions about the regulatory mechanisms at work and their organiza-

tional patterns. On the other hand, the description of how the organization at the system level emerges is far from being trivial. Despite the enormous complexity differences between living organisms as worms (*e.g. C. Elegans*) and humans the difference between their genome length is too small to explain the species' evolutionary gap. The difference should be unveiled by looking at the complexity of each gene network that relies on the variety of collective responses or phenotypes it displays. It is therefore clear that genome-wide approaches will allow to discover new higher-order patterns.

The characterization of the dynamical complexity of real regulatory networks is a very recent issue. The most successful approach is to use Boolean dynamics to characterize the regulatory interactions and thus to construct an oversimplified (free of parameters) model. Recent studies on this direction have addressed different regulatory networks, some examples are found in references (Mendoza et al. (1999); Li et al. (2004); Davidich and Bornholdt (2007); Albert and Oltmer (2003); Faure et al. (2006)). Although the number of gene networks that are currently analyzed from the dynamical point of view is growing, it will still take a long time to follow the evolution of complexity in living organisms by comparing gene networks and their dynamical behavior. Up to now the models focus on reproducing the sequential expression patterns observed experimentally. This is the case for the segment polarity gene network in *Drosophila Melanogaster* (Albert and Oltmer (2003)), and the cell-cycles of *Saccharomyces Cerevisiae* (Li et al. (2004)) and *Schizosaccharomyces Pombe* (Davidich and Bornholdt (2007)). These two latter studies are particularly important since the two cells are well-studied eukaryotic organisms and their network dynamics show remarkable differences. In particular, it has been observed (Davidich and Bornholdt (2007)) that although both cycles are similar in terms of the length of the basin of attraction, the overall dynamics observed when external signals are introduced is qualitatively different.

More promising, from the point of view of predicting power, are the findings for the flower morphogenesis regulatory network in *Arabidopsis thaliana* (Mendoza et al. (1999)). In this case the 5 different phenotypes (dynamical attractors) of the flower (petals, sepals, stamen, carpels, and flower inhibition) are reproduced, plus a new sixth phenotype that does not correspond to any previously found cell type. The above examples summarize the large predicting power of dynamical models in large real regulatory networks. Despite the use of a coarse-grained view as Boolean dynamics, the qualitative aspects of cellular dynamics seem to be already captured in this framework.

# 4

# Conclusions

Finally, we would like to emphasize that the studies mentioned here and others available in the literature are only the tip of the iceberg. It is expected that new tools come into play and that the universal behavior observed in the topology of many diverse phenomena from physical, social, technological and biological systems will allow a cross-fertilization between different disciplines, the ultimate goal being to tackle the complex structural and dynamical relationships in living systems such as the a cell. If this is achieved, then we would be able to use that consistent framework to make predictions and to develop alternative experimental techniques and practical applications such as targeted drugs.

# References


Barabási, A.L. and Oltvai, Z. (2004). Network biology: understanding the cell's functional organization, *Nature Review Genetics* **5**, 101–113.

Albert, R. (2005). Scale-free networks in cell biology, *Journal of Cell Science* **118**, 4947–4957.

Barabási, A.L. and Albert, R. (1999). Emergence of scaling in random networks, *Science* **286**, 509–512.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of "small-world" networks, *Nature* **393**, 440–442.

Wagner, A. and Fell, D. A. (2001) *Proc. R. Soc. London Ser. B* **268**, 1803-1810.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabasi, A.L. (2000). The large-scale organization of metabolic networks, *Nature* **407**, 651.

Uetz, P. et al (2003). A comprehensive analysis of protein-protein interactions in *Saccharomyces Cereviasiae*, *Nature* **403**, 623.

Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. (2003a). Global protein function prediction from protein-protein interaction networks, *Nature biotechnology* **21**, 697.

Agrawal, H. (2002). Extreme Self-Organization in Networks Constructed from Gene Expression Data, *Phys. Rev. Lett.* **89**, 268702.
bibitemalexei A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, (2003b) Modeling of protein interaction networks, ComplexUs 1, 38.

Vogelstein, B., Lane, D. and Levine, A. J. (2000). Surfing the p53 network, *Nature* **408**, 307–310.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks, *Science* **298**, 824–827.

Jacob, F. and Monod, J. (1961). Genetic Regulatory Mechanisms in the Synthesis of Proteins , *J. Mol. Bio.* **3**, 318–356.

Wall, M.E., Hiavacek, W.S. and Savageau, M.A. (2004). Design of gene circuits: lessons from bacteria, *Nature Rev. Genetics* **5**, 34–42.

Kauffman, S.A. (1969). Metabolic stability and epigenesis in randomly constructed genetic sets, *J. Theor. Biol.* **22**, 437–467.

Glass, L. (1975). Classification of biological networks by their qualitative dynamics *J. Theor. Biol.* **54**, 85–107.

Derrida, B., and Pomeau, Y., (1986) Random networks of automata: a simple annealed approximation, *Europhys. Lett.* **1** 45–49.

Socolar, J.E.S, and Kauffman, S.A. (2003). Scaling in Ordered and Critical Random Boolean Networks *Phys. Rev. Lett.* **90**, 68702.

Samuelsson, B., and Troein, C. (2003). Superpolynomial growth in the number of attractors in Kauffman networks *Phys. Rev. Lett.* **90**, 98701.

Kauffman, S.A., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random Boolean network models and the yeast transcriptional network *Proc. Natl. Acad. Sci. USA* **100**, 14796–14799.

Drossel, B., Mihaljev T., and Greil, F. (2005). Number and length of attractors in a critical Kauffman model with connectivity one *Phys. Rev. Lett.* **94**, 88701.

Aldana, M., Cluzel, P. (2003). A natural class of robust networks, *Proc. Nat. Acad. Sci. USA* **100**, 8710–8714.

Fox, J.J., and Hill, C.C. (2001). From topology to dynamics in biochemical networks, *Chaos* **11** 809–815.

Alon, U. (2007). An introduction to Systems Biology. Design Principles of Biological Circuits, Chapman & Hall.

Sethna, J.P. (2006). Statistical Mechanics: Entropy, Order Parameters, and Complexity, Oxford University Press.

Andrecut, M., and Kauffman, S.A. (2006). Mean-field model of genetic regulatory networks, *New J. Phys.* **8**, 148.

Gómez-Gardeñes, J., Moreno, Y., and Floría, L.M. (2006). Scale-Free topologies and Activatory-Inhibitory interactions *Chaos* **16**, 15114.

Bornholdt, S. (2005). Less is More in Modeling Genetic Networks, *Science* **310**, 449–451.

Klemm, K., and Bornholdt, S. (2005). Topology of biological networks and reliability of information processing, *Proc. Natl. Acad. Sci. USA* **102**, 18414–18419.

Bradman, O., Ferrel Jr., J.E., Li, R., and Meyer, T. (2005). Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions *Science* **310**, 496.

A. Ma'ayan et al. (2005). Formation of Regulatory Patterns During Signal Propagation in a Mammalian Cellular Network *Science* **309**, 1078–1083.

Lodato, I., Boccaletti, S. and Latora, V. (2007). Synchronization properties of network motifs *Europhys. Lett.* **78** 28001.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators, *Nature* **403**, 335–338.

García-Ojalvo, J., Elowitz, M.B., and Strogatz, S.H. (2004). Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing *Proc. Natl. Acad. Sci. USA* **101**, 10955–10960.

Mendoza, L., Thieffry, D., and Alvarez-Buylla, E.R. (1999). Genetic control of flower morphogenesis in Arabidopsis Thaliana: a logical analysis *Bioinformatics* **15**, 593–606.

Li, F., Long, T., Lu, Y., Quyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proc. Nat. Acad. Sci. USA* **101**, 4781–4786.

Davidich, M.I., and Bornholdt, S. (2007). Boolean network model predicts cell-cycle sequence of fission yeast *arXiv/q-bio:0704.2200.*

Albert, R., and H.G. Oltmer, H.G. (2003). The topology of the regulatory interactions predicts the expression pattern of the Drosophila segment polarity genes *J. Theor. Biol.* **223**, 1–18.

Faure, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle *Bioinformatics* **22**, e124–e131.