Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*

Laurens Kraal*
3130967
Cancer Genomics & Developmental Biology (UU)
Master Thesis
Supervisor: Professor Dr. Frank Holstege

August 10, 2009

Abstract

One of the main goals of systems biology is reverse engineering network structures from experimental data. Using quantitative networks measures, biological processes can be characterized and understood on a systems level. This paper reviews the scientific literature on the transcriptional regulatory networks of *Saccharomyces cerevisiae* to discuss what kind of structures can be identified and what this means biologically. This discussion comprises network characteristics, network dynamics in biological processes, and robustness; an inherent emergent property of networks.

1 Introduction

Systems Biology

Traditionally, biology is a reductionist science. Over the years it has successfully identified many components and interactions of various nature by reducing them to elementary units that can be isolated and studied independently [1]. However, there are many properties of living systems that have no hope of being understood using this method. A system is a holistic entity whose parts have dynamic interactions both with each other, as with the outside world. Its behavior and organization emerges out of this complex interplay. Complex systems like the cell, comprise numerous functionally and structurally diverse components that interact in networks to generate coherent behaviors [2]. The behavior is a function of network properties and the elements involved. Although molecular biology has uncovered many facts about network components, biological systems cannot be understood on this level. A system-level perspective is necessary to understand the function and behaviors of biological systems such

*31 Leroy St, Dorchester, MA 02122, USA
+1-617-606-1949, negnin@gmail.com

as organelles, cells, tissues, organisms, and ecologies. Systems biology is the interdisciplinary field that tries to generate and analyze system-level data and tries to understand biological mechanisms in the context of their system. It aims to develop quantitative mathematical models to gain a formal understanding of biological processes [3].

As biological systems comprise intricate networks, graph theory is employed to analyze and quantify them. For transcriptional regulatory networks, this involves abstracting interactions from experimental data generated by microarray and ChIP-chip to directed graphs with transcription factors and target genes as nodes and their interactions as links between them (see figure 1). Interactions between genes can be identified from experimental data using reverse-engineering methods [3]. This paper deals with the transcriptional regulation networks of the baker's yeast *Saccharomyces cerevisiae* as much work on illuminating the network organization has been done on this eukaryotic organism.

Outline

The first section of this paper discusses the measures that can obtained from network architectures and topology. The next section deals with the integration of these measures with a dynamical analysis of biological processes. The section after this has robustness, an emergent property inherent to complex networks, as its subject. The paper ends with a discussion.

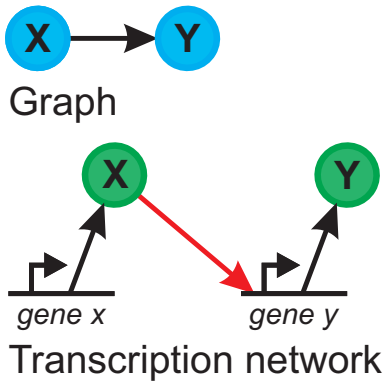

Figure 1: A transcriptional relationship involves a transcription factor binding to a regulatory sequence of a gene to regulate its expression. This relationship is represented in a graph by a node targeting another node (figure adjusted from [4]).

2 Network Measures

The study of complex networks took of more than 50 years ago, with the the work done by the Hungarian mathematicians Paul Erdös and Alfréd Rënyi [5]. They suggested that complex networks could be modeled as regular objects, such as a square or a diamond lattice, or as a random network with random links connecting the nodes of the network. Modeling random networks enabled the study of the structure and properties of complex networks. One of the most elementary characteristics of a node is its degree (or connectivity). The degree component k is a measure for how many links the node has to other nodes and it determines many properties of the system. One of the predictions of random network theory is that such a system is democratic or uniform. As the links connecting the nodes will be almost completely evenly

distributed, the nodes follow a bell shaped Poisson distribution. Nodes that have significantly more or fewer links to other nodes than the average degree, $< k >$ (<> denotes the average), are hard to find in random networks. Random networks are also called exponential, as the probability that a node has connections to k other nodes decreases exponentially for a large k. [6]

However, in the absence of data on large networks, the predictions of random network theory were rarely tested in the real world. In this age of computerized data acquisition, the topological information of many real-world networks have become available and it has been found that the topological properties of real networks, such as biological networks, cannot be explained by random network models [7]. Instead, most real networks are 'scale-free', which means they are categorized by a degree distribution that approximates a power-law ($P(k)$ k^{γ} where indicates 'proportional to') . The degree distribution is characterized by the the probability P(k) that a node in the network interacts with k other nodes [8].

The degree distribution is essential in differentiating between networks. The peaked degree distribution of a random network indicates that the system has no highly connected nodes and most nodes have a typical degree (See figure 2) [8]. In a power law distribution, most nodes have just a few connections while others, the so-called hubs, have a huge number of links. In this sense the network does not have a scale and that is why its topology is termed 'scale-free'. While random networks have a democratic distribution of links, a scale-free network is dominated by the hubs. Hubs are not present in random networks. Using the degree distribution it was found that many complex networks are scale-free, ranging from the internet to transcriptional regulatory networks [9].

In networks characterized by the power law $P(k)$ k^{γ}, the importance of hubs is negatively correlated to the value of γ. For reasons yet unknown k tends to fall between 2 and 3 in scale-free networks. For $\gamma = 2$ there is a hub-and-spoke network with a large fraction of nodes being connected to the largest hub. For $2 < k < 3$ a hierarchy of hubs emerges with only a fraction of all nodes connected to the largest hub. For $k > 3$ the scale-free network behaves like a random one. The unusual properties of scale-free networks only arise for $k < 3$. [6,9]

A feature that many complex networks have in common is that any two nodes can be connected by a path traversing only a few links. This is known as the small-world effect, first discovered by Milgram in the human acquaintanceship network [10]. It was found that any two individuals have an average "six-degrees of separation", meaning that on average they can be connected to each other through 6 intermediate steps.

Although the small-world effect can be found in random networks, the diameters of scale-free networks are ultra-small [11]. The diameter is the average minimal path length between all pairs of nodes in a network. Most biological networks are thought to have a small diameter because it minimizes transition times and is thus more efficient and responsive through enhanced signal-propagation speed [12, 13].

However, recently it was found that real networks are not smaller than their random counterparts [14]. The diameter of 13 real networks, ranging from linguistic to biologi-

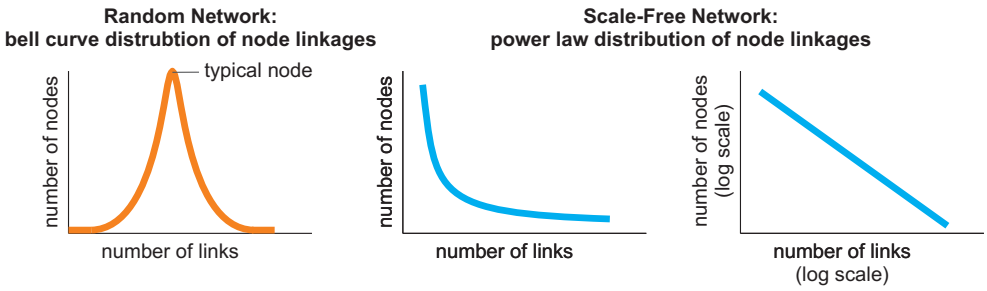

Figure 2: A plot of the distribution of node linkages of a random network follows a bell-shaped curve. Most of the nodes have the same number of links. In scale-free networks the distribution follows a power law. Most nodes have just a few connections, but some have a tremendous amount of links, the hubs. Plotting the distribution on double-logarithmic scale results in a straight line (figure adjusted from [6])

cal, was compared to that of their randomly rewired network wherein the degree of each node remained unchanged, but the connections are randomized. The *S. cerevisiae* transcriptional regulatory network ([15]) that was analyzed comprised 3459 nodes. It had an observed diameter of 3.72 which is about 10% higher than what is randomly expected (3.39). Simulations showed that allowing a slightly greater diameter increases the network modularity significantly. This suggest that complex networks made a trade-off between efficiency and the substantial advantages of modularity (see below).

While the average path length $< l >$ that defines the small-worldness is a global feature of the network, the cliqueness of a node neighborhood, a local feature, is measured by the clustering coefficient $C(p)$ (p denotes probability) [12] (see figure 3. If the neighborhood of a node is fully connected, the clustering coefficient will be 1, while a clustering coefficient closer to 0 means that there are little connections in the neighborhood.

The overall tendency of the nodes of a network to form clusters or groups is characterized by the the average clustering coefficient $< C >$. An important measure of the hierarchical organization of a network is the average clustering coefficient of all nodes with k links, $C(k)$. In many real networks the clustering coefficient follows the scaling law $C(k)\ k^{-1}$, which indicates a hierarchy of nodes with different degrees of modularity [16].

A final characteristic of a network is its directedness. An undirected network with N nodes and L links is characterized by it's average degree given by $< k >= 2L/N$. transcriptional regulatory networks are not undirected as each link has a selected direction. Each node of a transcriptional regulatory network has an incoming degree, k_{in} which denotes the number of links pointing to the node, and an outgoing degree k_{out} which denotes the number of links starting from this node [9]. In transcriptional networks they represent the number of tran-

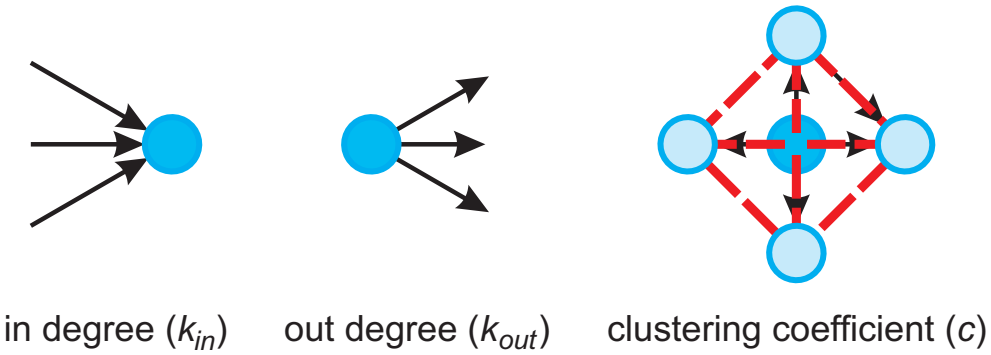

Figure 3: The in-degree k_{in} indicates the number of direct incoming edges per node. The out-degree k_{out} indicates the number of direct outgoing edges per node. In a transcriptional regulatory network this is the number of transcription factors per target gene or the number of target genes per transcription factor. The clustering coefficient c indicates the ratio of the number of edges between a node's neighbors and the maximum number of possible edges. (figure adjusted from [15])

scription factors regulating a target gene and the number of target genes for each transcription factor, respectively [15].

Directedness is especially important when analyzing hierarchies of nodes such as motifs and modules (see below).

While the average degree $< k >$, average path length $< l >$ and average clustering coefficient $< C >$ depend on the number of nodes and links (N and L) in the network, the $P(k)$ and $C(k)$ functions are independent of network size and can therefore be used to classify various networks by characterizing their generic features [9].

Hubs

Two intrinsically related mechanisms are responsible for the existence of hubs in scale-free networks: growth and preferential attachment [7]. Real networks, such as the internet or a gene regulatory network, have a history of expansion. They emerged as new nodes joined existing connections over an extended period of time. The nodes that have been part of the network longer had more opportunities to gain more connections. Nodes also display preferential attachment; the preference to link to nodes that are already have many links (see figure **??**). As well connected nodes gain more and more connections by virtue of network growth and preferential attachment, they become hubs and the scale-free property of the network emerges.

Although little is known about the origin and evolution of gene regulatory networks, gene duplication is thought to be a major force of its growth and emergence in evolutionary history [17, 18]. Gene duplication events will lead to two identical genes with the same regulatory connections. Therefore, each gene that is in contact with the duplicated gene will gain an extra link (see

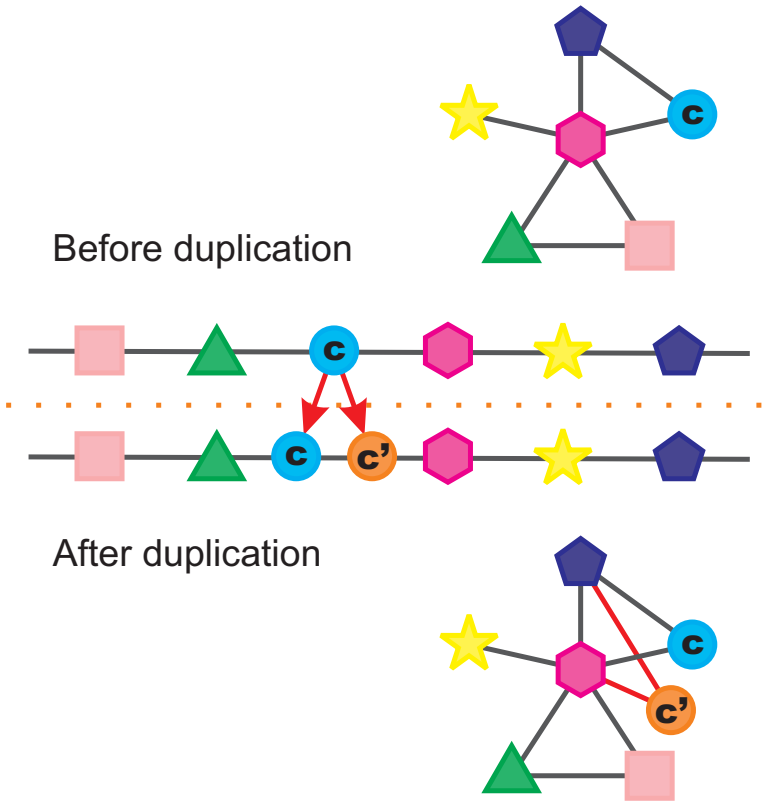

Figure 5: A small gene regulatory network before and after duplication of a gene. The gene duplication event expands the network with an extra node. This node inherits the connections of its parent. As the genes that interact with the parent gain a new link, genes that have many interactions are more likely to gain an interaction after a gene duplication event. This is the mechanism that generates preferential attachment. In the shown figure the hexagon hub will gain a new interaction regardless of which gene is duplicated, while the least connected gene, the star, will only gain a new link if the hub is duplicated. (figure adjusted from [9])

figure 5). Although highly connected genes themselves do not have a higher probability of being duplicated, they are more likely to be connected to a random gene that is duplicated than their less connected peers. Thus, well connected genes have a higher probability of gaining more links. However, this pure duplication model does not support a scale-free distribution of connections [19]. By modeling yeast regulatory network growth, it was shown that gene duplication is not sufficient to yield a scale-free network. Duplication events have to be followed by a second event that breaks the parent-daughter symmetry of the pure gene duplication model. Models that include partial duplication or duplication plus preferential rewiring result in a broader range

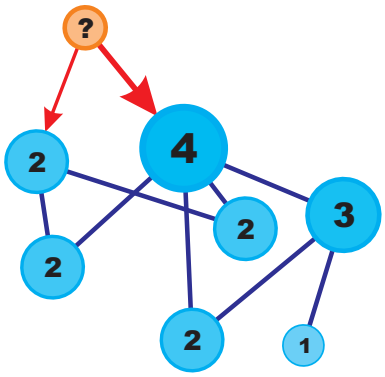

Figure 4: Preferential attachment refers to the preference of nodes to attach to more-connected nodes. The probability that the new node is going to connect to the node with the degree node of four is twice the chance that it will connect to the node with only two links. Network growth and preferential attachment thus generate hubs. (figure adjusted from [9])

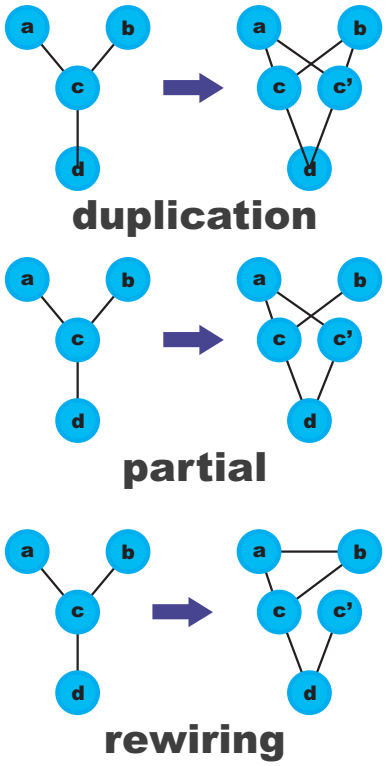

Figure 6: Mechanisms of network growth through gene replication. In the pure duplication event, node c' is created by duplication of the connectivity of the parent node c. In partial duplication, node c' does not retain all the connections of the parent node. In duplication with rewiring, node a gains a new connection to k, while c' retains the connection with b from the parent node. (figure adjusted from [19])

of node connections and the scale-free property (see figure 6. This suggests that the structure of a network is inseparably connected to its history or evolution [20]. However, it is important to note that, although the models show that gene duplication can lead to power-law distributions, there is no direct proof that it is the (only) mechanism that generates the scale-free topology of cellular networks.

Network Modules and Motifs

Next to the global features of a network there are basic structural elements specific to each class of networks. A discrete biological function can only seldom be performed by a single molecule [21]. Cellular functions are likely carried out by modules comprising different kinds of nodes [21]. Most molecules of the cell are connected together in modules to generate a specific and relatively distinct function. A high $< C >$ implies a high amount of clustering in the network suggesting the presence of modules. Modules are per definition relatively isolated from the rest of the system. The

prevalence of hubs that connect many nodes in the system suggest that the existence of modules are an exception, but clustering and hubs are not exclusive as modules interact with other nodes and modules to form a hierarchical network [16].

High clustering of a network indicates the presence of various subgraphs. The specific patterns of interconnections characterize the network at a local level. Some subgraphs are however more significant than others. Motifs are patterns of connections that are overrepresented in certain networks. To find the motifs of the yeast transcriptional regulatory network, Milo et al. [4] scanned the directed networks for all possible n-node subgraphs (with $n = 3$ and $n = 4$) and compared the results with randomized networks. In the randomized network the number of the incoming and outgoing links are kept the same, but the connections are randomized. When the number of subgraphs with $n = 3$ and $n = 4$ found in the gene regulatory network of *S. cerevisiae* were compared with those in the randomized networks based on the same data, the three node "feed forward loop" motif and the four node "bi-fan" were the only two motifs found to be significantly overrepresented (see figure 8. Out of 13 possible subgraphs with $n = 3$ (see figure 7) and 199 possible subgraphs with $n = 4$ only these two are significant, occurring 10 standard deviations times more often in the transcriptional regulatory network than in randomized networks. The feed-forward loop was found 70 times, while the bi-fan was found 1812 times in the yeast genome. Interestingly, the feed-forward loop and the bi-fan are also found numerous times in neuron connectivity and electronic circuit networks as well as in the transcriptional regulatory network of the bacterium *Escherichia coli*. As these networks have information processing in common, it is likely that this motifs play a functional role in information processing networks and therefore characterize them.

The second most prominent network motif, the feed-forward loop, consists of three genes (nodes X, Y and Z) and three transcriptional interactions (the edges) shown in figure 8. In the feed-forward loop, X regulates Z both directly (X $\rightarrow$ Z) as indirectly through Y (X $\rightarrow$ Y $\rightarrow$ Z). The effect of the three transcriptional interactions in both pathways is either activating or repressing. The output of Z therefore depends on both the regulatory effects of the interaction in the direct pathway as on the effect of both interactions in the indirect pathway [22]. The function of a feed forward loop is to act as a buffer to only allow output if the signal is persistent and to facilitate rapid deactivation when the signal ceases [23].

The function of the bi-fan is less clear and depends critically on the nature of the regulation [24]. If all interactions (P $\rightarrow$ R, P $\rightarrow$ S, Q $\rightarrow$ R, and Q $\rightarrow$ S in figure 8) are positive regulations, the function of the bi-fan may be the coordinated expression of a set of genes dependent on two different inputs and thus function as a logical AND gate. However, if only one positive regulation event is necessary for expression, the bi-fan acts as a logical OR gate. The dynamic behavior of the bi-fan with different kinds of regulation (positive and negative) is therefore extremely varied, even within the range of biologically plausible parameters and configurations. It is therefore diffi-

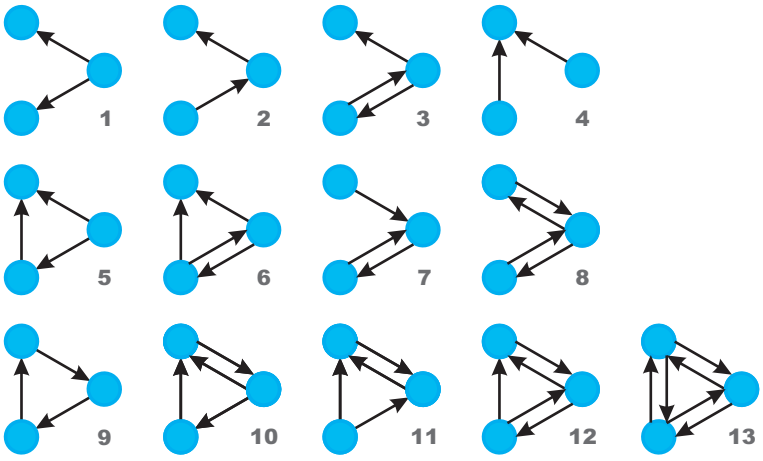

Figure 7: All 13 possible types of connected subgraphs with three nodes. (adjusted from [4])

cult to tell the function of the bi-fan based its structure and there is no characteristic behavior of this motif. Additional information about the transcriptional relations and kinetic parameters are necessary to understand the biology of individual bi-fan motifs [24]. However, the structure of the bi-fan implies at least that the combination of different input factors in important in coordinating gene expression.

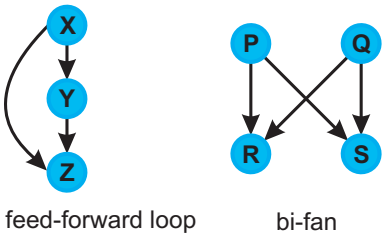

Figure 8: The feed-forward loop and the bi-fan, the two most common n-node motifs ($n = 3$ and 4) in the yeast transcriptional regulation network. (adjusted from [4])

While the study by Milo et al. was done on microarray data from the Yeast Database Proteome [25], Lee and colleagues took another approach to identify network motifs [26]. A genome wide location analysis (ChIP-on-chip) was performed to determine where 106 transcription factors bind to promoter sequences across the genome. Almost 4000 interactions between regulators and promoter regions were observed with 37% of all yeast genes being bound by at least one transcription factor. From the genome-wide location data the following six regulatory network motifs were identified: auto-regulation, multi-component loop, feed-forward loop, single-input, multi-input, and regulator chain (see figure9).

If a regulator binds to the promoter region of its own gene, it is called an autoregulation motif. 10 autoregulation motifs were found in the yeast genome, suggesting that 10% of the transcriptional regulators are autoregulated [26]. Autoregulation is thought to generate faster response times and an increased stability of gene ex-

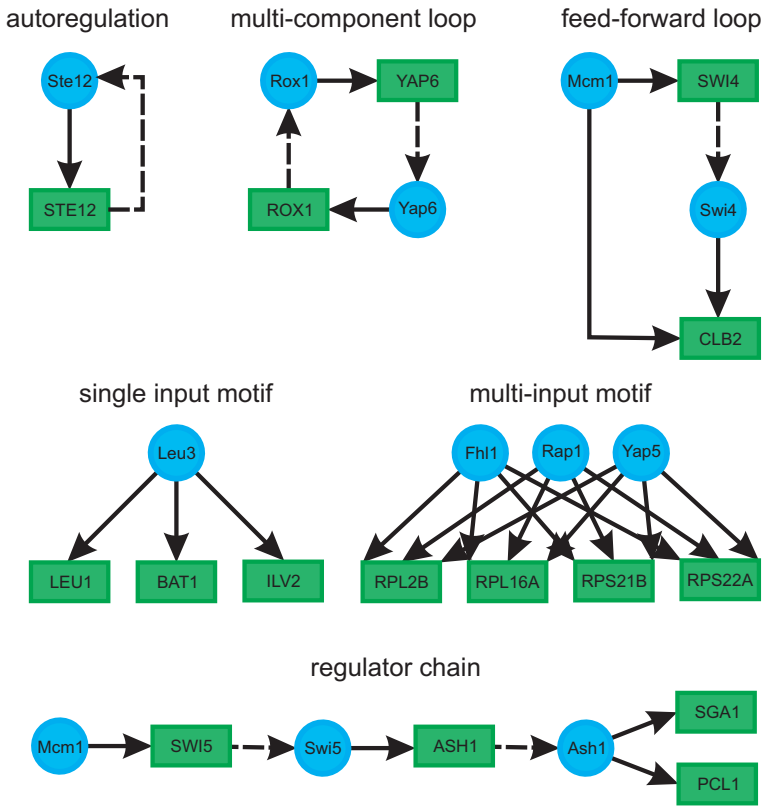

Figure 9: The network motifs found in the yeast regulatory network, with example genes. The regulators are represented by blue circles, while the target genes are represented by green rectangles. A solid arrow indicates the binding of a regulator to a promoter. The genes encoding regulators are connected to their regulators with dashed arrows. (adjusted from [26])

pression [27].

The multi-component loop is a closed circuit with multiple factors, where regulators indirectly regulate themselves. The regulation steps in between give additional opportunities for control and regulation. In this way they are capable of feedback control and generating bistable systems that can switch between two discrete, alternative steady states [28]. In the location data for the 106 transcription factors, only three multi-component loops were found [26].

The single-input motif contains a single regulator that binds to a set of condition under a specific condition. They are instrumental in controlling modules performing a discrete biological function by coordinating their expression [21]. The multi-component motif is an expanded bi-fan motif [4] (see above). It may enable additional control of the expression of genes that is dependent on multiple different inputs. 295 combinations of two or more regulators that bind a common set of promoter regions were found

(this includes bi-fan motifs) [26].

The regulator chain motif is defined as a chain of three or more regulators where the first regulator binds to the promoter sequence of the second regulator and the second regulator binds to the promoter sequence of the third regulator and so forth [26]. This sequence enables the expression of genes in a temporal order. This is critical in a controlled temporal events such as the cell cycle, where regulators of one stage of the cell cycle regulate the expression of genes that are necessary to enter the next stage (see below) [29].

The studies done by Lee et al. and Milo et al. are complementary, but also very different in nature. Milo and colleagues used microarray data from a multitude of experiments to infer regulatory relationships based on expression levels [4]. This means that they did not focus on transcription factors and included indirect relationships. However, they did constrain their analysis by only searching for motifs with 3 or 4 nodes. In contrast, Lee and colleagues focused only on physical interactions between transcription factors and promoter regions and thus excluded indirect regulation [26]. They also only grew the yeast strains in rich medium, thereby ignoring regulatory events in other conditions, and unfortunately, because of experimental difficulties, analyzed only 75% of all transcription factors. However, they did not constrain the number of nodes of the motif. This is one of the reasons why they find more motifs.

Identifying regulatory motifs helps us to understand the structure of the network architecture. Overrepresentation and convergent evolution of certain motifs make a strong case for the biological relevance of certain motifs [9]. However, there remains a distinct possibility of functional relevant, but statistically insignificant motifs. Also, as mentioned in the discussion of the bi-fan, the specific regulatory interactions are important in determining the specific function of each motif. It is also important to realize that there is a significant amount of cross-talk between network motifs [22]. Network motifs are not independent units functioning apart from the rest of the network. They form complex structures by interacting or cross talking to each other, which happens when they share a node.

3 Biological Dynamics

[15] The networks discussed above were only analyzed statically [4, 9, 18, 23, 26, 30]. This may oversimplify the actual organization, as cells have inherent dynamics by virtue of being alive. In this section, the network measures are integrated with the analysis of the dynamical network organization of biological processes. By integrating transcriptional regulatory information and gene-expression data for multiple conditions, Luscombe et al. were able to analyze the network dynamics of *S. cerevisiae* on a genomic scale [15]. First a static representation of the known regulatory interactions was generated comprising 7074 regulatory interactions, 142 transcription factors and 3420 target genes. For the dynamic perspective, the data from 240 ChIP-chip experiments was integrated, representing 5 conditions: cell-cycle, sporulation, diauxic shift, DNA damage, and stress response.

This analysis showed great differences in the network wiring in the different conditions.

Although most transcription factors are used in multiple conditions, half of the target genes are only expressed in one of the conditions [15]. More than half of the active regulatory interactions are completely rewired in the different sub-networks. There are only 66 interactions that are retained over four or more conditions. These perpetually active 'hot links' [9] are mostly regulating housekeeping functions.

The five condition-specific sub-networks could be divided into two categories based on their network characteristics: exogenous and endogenous [15] (see figure 10). The endogenous conditions, cell cycle and sporulation, are regulated by an internal transcriptional program and are organized in multiple stages. The exogenous conditions, diauxic shift, DNA damage, and stress response, are responding to binary external events and are characterized by a rapid turnover of expressed genes. While a static analysis may suggest that the global topological measures that quantify the network architecture are constant in most biological networks [9], examining these measures for exogenous and endogenous conditions shows significant changes between them [15].

Specifically, the nodes of the exogenous sub-networks have on average a 20% lower 'in degree' (k_{in}). In contrast, in the endogenous conditions the average k_{out} and path length (l) are twice as high as for the endogenous conditions. The clustering coefficient (c), the interconnectivity of nodes, in the exogenous conditions is almost half that of nodes in the endogenous sub-networks. [15]

Biologically, the small k_{in}, the larger k_{out} and the short path lengths indicate that in exogenous conditions, the transcription factors are regulating their target genes in simpler combinations, target more genes simultaneously, and propagate their regulatory signals faster than in endogenous conditions [15] (see figure **??**). In contrast, the longer paths in the endogenous sub-networks indicate the existence of regulator chains (see figure **??**) that generate slower actions and high control of intermediate phases [15, 26]. The higher level of control is also signified by the high clustering coefficient indicating greater inter-regulation of transcription factors in endogenous conditions. In short, exogenous sub-networks have evolved to generate rapid large-scale responses, while endogenous sub-networks are responsible for carefully coordinated and controlled processes.

The endogenous and exogenous sub-networks were also analyzed for the presence of network motifs [15]. The most common motifs present were the single-input, multi-input and feed-forward loops (see figure 9. Although the relative multi-input motif usage is not different between both set of conditions, the exogenous networks favor the use of single-input motifs, while the relative occurrence of feed-forward loops is significantly higher in endogenous conditions.

This is in line with the biological functions of the endogenous and exogenous processes [15]. By conferring similar regulation to groups of genes, the single- and multi-input motifs function as orchestrators of large-scale gene activation in exogenous conditions. Because of the buffer capacities of feed-forward loops that make them

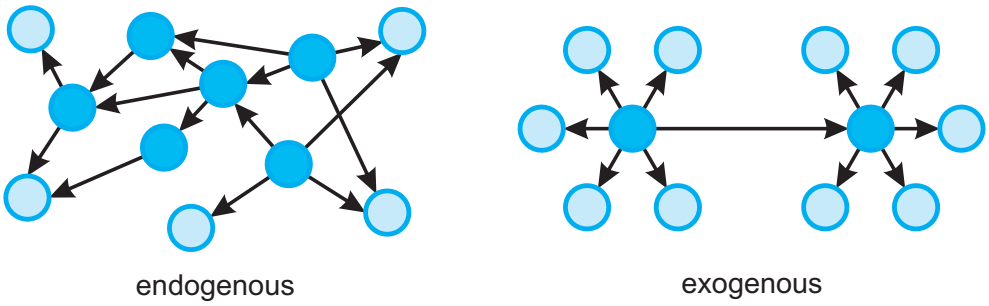

Figure 10: Endogenous processes have complex transcription factor combinations, few targets per transcription factors and long path lengths. The transcription factors are highly inter-connected and have many feed-forward loops. In contrast, exogenous processes have simple transcription factor combinations, many targets per transcription factor, short path lengths and few inter-connected transcription factors. The most common motif of exogenous processes is the single input motif. (figure adjusted from Luscombe04)

only responsive to persistent signals, they are ideal to ensure that stages of endogenous processes have finished before allowing entry to the next stage.

All the analyzed sub-networks display scale-free characteristics, signifying the importance of regulatory hubs that target a disproportionally large fragment of genes [15]. As hubs have the most interactions, they are extremely influential in the behavior of the network [9]. Two groups of hubs were identified in the sub-networks. Permanent hubs that are responsible for regulating house-keeping functions are the smaller group. The bigger group, comprising almost 80% of the hubs, are transient, meaning that they only are influential in some conditions. Transient hubs are less prevalent in exogenous sub-networks (as signified by a lower γ) which suggest a more centralized command hierarchy (see figure 10). Transient hubs are able to change their interactions between conditions. The amount of interchange differs for all transcription factors. 27 transcription factors rewire all their interactions when conditions change, many of which only actively regulate genes in a single condition. However, most transcription factors rewire only a fraction of their interactions, including most hubs. Interestingly the permanent hubs rewire their connections as much as transient hubs, only over more conditions. Permanent hubs take on roles in addition to their core function in different (exogenous) conditions. This rewiring allows transcription factors to be active in many conditions and generates significant overlap in transcription factor usage for different functions. This shows that the cell dynamics are critical in determining functions and topology.

Cell cycle

Dynamic transitions also take place within conditions. The cell cycle is a multi-stage endogenous process. It comprises five phases: early G1, late G1, S, G2, and

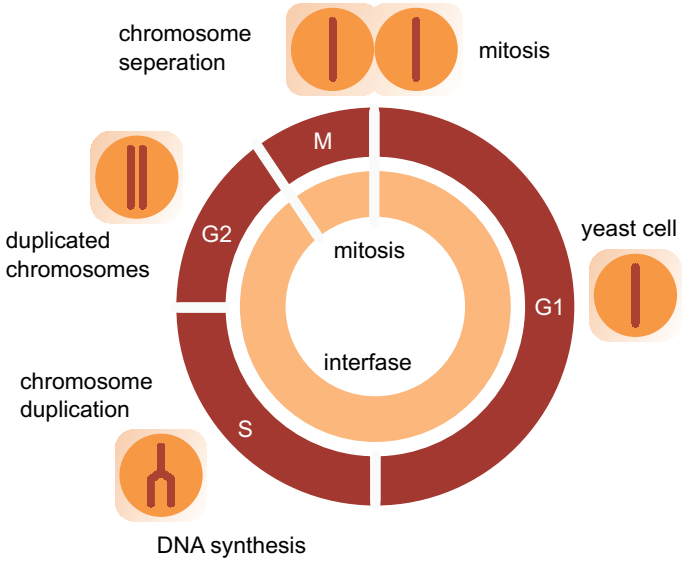

Figure 11: The yeast cell cycle comprises five phases: early G1, late G1, S, G2, and M. DNA is synthesized in the S phase and mitosis happens in the M phase. The G1 and G2 are gap phases in this sequential process.

M (see figure 11). Because it is known which genes are differentially expressed in different phases [31], the sub-networks of the different phases could be studied and compared. It was found that only a minority transcription factors operate throughout all stages. Most transcription factors are only active in one particular stage. Two major forms of transcription factor interregulation are instrumental in coordinating the cell cycle. In serial inter-regulation, the active transcription factors of one phase sequentially regulate the expression of transcription factors that function in the next stage [29]. This forms regulatory network cascades that ultimately run in a circle. These cascades are responsible for the long path length network measure of the endogenous cell cycle process. Next to serial transcription factor interregulation there is also parallel interregulation present in the cell cycle. Using a two tiered system, the the ubiquitously active transcription factors control the phase-specific transcription factors to provide them with a stable signal to aid them through the phase-transitions [15]. Because many of the ubiquitously active transcription factors are hubs, the progress of the cell cycle can be coordinated with house-keeping functions. This is part of the aforementioned hub rewiring, as the cell cycle is not always active. This analysis suggests that the network of sequentially expressed transcription factors controls the temporal program of transcription of the cell cycle with an emergent oscillating property [32].

As there is no such thing as condition unspecific measurements, analyzing the network dynamics is essential in uncovering the system-level adjustments cells make for different condition. Analyzing the dynamics of biological networks by network measures provides a way to quantify, characterize and categorize the cell's behavior. Combining the network analysis with biological knowledge opens up a way to put the emergence of sub-networks, rewiring, transcriptional transitions, preferential motif usage etc. in a biological context. However, there are some emergent properties that are thought to be intrinsic to most biological networks. One such property is the topic of the next section.

4 Robustness

A fundamental property of all biological systems is robustness [33]. Robustness is a typical system-level phenomenon that cannot be understood at a level of individual components. It refers to the systems ability to maintain its functions against internal and external perturbations [34]. Robustness doesn't mean unchanged regardless of stimuli or mutations and is therefore distinct from homeostasis and stability. Homeostasis comprises the coordinated physiological processes that maintain the steady state of biological systems. It therefore refers to a stable condition and is a property that maintains the state of the system rather than its functions. In contrast, robustness of the system can mean transitioning states if it's instrumental to maintaining the functionality of the system while coping with perturbations. Changing states is a major coping mechanism of cells facing exogenous stress conditions [15]. Stability also differs from robustness, as robustness can be gained by increasing instability. Viruses, for example, use genome instability and high mutations rates to increase survivability [35]. Various mechanisms that generate robustness actually facilitate evolution, while evolution itself selects for robust traits [33].

Increased robustness comes with a cost. Being robust to certain perturbations generates trade-offs by causing the system to collapse when encountering unexpected perturbations [33]. Systems that have evolved to be robust to general perturbations are fragile against certain types of rare perturbations because of evolutionary optimizations and selective pressure [36]. Trade-offs are inherent to the system and cannot be avoided. Robustness can also come at the cost of performance.

There the robustness of a system is ensured in four ways: system control, alternative mechanisms, modularity, and decoupling [33].

System control involves using feedback to generate robust, dynamic behavior in regulatory networks [33]. Robust responses are mainly enabled by negative feedback, facilitating the perfect adaptation to a range of input signals. Positive feedback is used to amplify signals, creating discrete responses and bistability. This way the activation of downstream pathways is controlled and states are clearly distinguished. Amplification also makes the system robust to noise and fluctuations of stimuli.

Alternative mechanisms are used a failsafe for when components fail to function

properly [33]. When the probability that a function with a single component fails is p, the probability that the function with two components that back each other up fails is $(1-p)^2$ [33]. This usually involves a certain degree of redundancy within the system, but can also be attained by a heterogeneous population of components fulfilling a same function by other means. Although multiple identical components could make the system more robust by virtue of redundancy, the occurrence of this is very rare in biological systems. Fail-safe functions are usually performed by heterogeneous components with some functional overlap. Gene duplication is instrumental in generating such components in evolutionary history (see discussion above) [17–19]. Although gene duplications are rather common, so far there have been no reported duplications of network circuits [33]. Although redundancy through functional overlap of subnetworks could be a powerful mechanism in enabling network robustness, this seems to be difficult to obtain in evolution. System control mechanisms coordinate alternative mechanisms to maintain system behavior and function.

Related to alternative mechanisms is modularity. Modules are present in biological systems on different levels with a hierarchical organization [33]. An organelle is an example of a biological module that can be part of a cell module of a multi-cellular organism, but can also comprise lower-level modules itself. Modules can be organized on a functional, spatial and temporal level as seen in the discussion about the cell cycle. Modules can contain perturbations locally so they cannot propagate to the system thereby facilitating robustness. Furthermore, by using modular organization, the failure of a module does not necessarily affect the core functionality of the biological system. The downside of modularity is that it comes at a cost of increased resource demands.

Decoupling creates buffers between the input and output of a system [33]. This can mean encoding a certain environmental input to a signal that can be propagated or allowing fluctuations of signals by correcting them or leveling them out. Buffering is thought to be an intrinsic property of complex networks.

As robustness is an emergent property of the system, it is embedded in the network structure. When removing a substantial number of nodes of a random network, the system will collapse leaving islands of non-communicating islands of nodes [9]. In contrast, complex networks with a scale-free nature are extremely resilient against component failure. The topology of the network is instrumental in generating this robustness [37]. There is no threshold of disintegration for scale-free networks. When as many as 80% of randomly selected nodes are removed the remaining 20% still form a connected network wherein any two nodes can still be connected with a path. Hubs are the key to this feature, as removing random nodes will mainly affect the numerous less-connected nodes. Only removing hubs significantly interrupts the network's integrity. The downside of this is that scale-free networks are extremely vulnerable to a coordinated attack on the hubs. Biologically, this means that in *S. cerevisiae* 60% of proteins with more than 15 interactions are essential, while only 10% of the proteins with 5 interactions or less are vital for survivability

as shown by deletion analysis [38].

However, it is important to remember that in biological networks each node has its own biological function [9]. How the system deals with a perturbation of a node therefore doesn't merely depend on the degree of the node. This shows that a network analysis has to be supplemented by biological knowledge and experiments to properly understand the system.

5 Discussion

Identifying the structures and dynamics of yeast transcriptional regulatory networks, shows the power of quantitative analysis. The networks models that are generated are instrumental on understanding how biological processes and behaviors are organized and coordinated. However, the models still represent extremely simplified versions of the intricate inner-workings of the cell. For one, it doesn't take pre- and posttranscriptional and translational modifications into account [39]. The nature of the transcriptional relationships is also ignored in graphing the nodes. Although directional, the distinction between positive and negative regulation is seldom made and the strength of interactions is not quantified. It is also not clear how the transcriptional regulatory networks relate to protein-interaction networks, and metabolic networks. This relates to the question of how 'real' the generated models are.

However, some progress has recently been made to benchmark the reverse-engineering approaches [3]. Taking a synthetic biology approach, a synthetic gene network was developed in *S. cerevisiae* that can be used as a gold standard. The network comprises the five transcription factors (ASH1, CBF1, GAL4, GAL80, and SWI5) and includes regulator chains, single-input motifs, and multiple feedback loops. Time-series and steady state expression profiles were collected for the synthetic circuit after perturbations and overexpression experiments. Because the topology of the network is known, the validity and power of reverse engineering methods could be assessed. This approach greatly facilitates the evaluation and development of inference tools [39].

This integration of synthetic and systems biology also shows the promise of increased understanding of network functions. The goal of synthetic biology is to engineer organisms to display specific functions or to optimize the expression of certain compounds [40, 41]. Understanding the network structure and distilling the rules that generate desired behaviors is essential to this goal.

As our understanding of network biology grows, the possibilities increase. Quantifying biology is an important step in harnessing the potential of biology, but only an early step in fully understanding cells and organism on a system-level.

References

[1] L. Chong and L.B. Ray. Whole-istic biology, 2002.

[2] H. Kitano. Systems biology: a brief overview, 2002.

[3] I. Cantone, L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, and M.P. Cosma. A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches. *Cell*, 2009.

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks, 2002.

[5] P. Erdos and A. Renyi. On random graphs. *Publ. Math. Debrecen*, 6(290-297):156, 1959.

[6] A.L. Barabasi and E. Bonabeau. Scale-free networks. *Scientific American*, 288(5):50–59, 2003.

[7] A.L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[8] R. Albert and A.L. Barabasi. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47–97, 2002.

[9] A.L. Barabási and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[10] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[11] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):58701, 2003.

[12] DJ Watts and SH Strogatz. Small world. *Nature*, 393:440–442, 1998.

[13] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proceedings: Biological Sciences*, 268(1478):1803–1810, 2001.

[14] Zhihua Zhang and Jianzhi Zhang. A big world inside small-world networks. *PLoS ONE*, 4(5):e5686, 05 2009.

[15] 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*, 431(7006):308–312, 2004.

[16] E. Ravasz, AL Somera, DA Mongru, ZN Oltvai, and A.L. Barabasi. Hierarchical organization of modularity in metabolic networks, 2002.

[17] A. Rzhetsky and S.M. Gomez. Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome, 2001.

[18] S.A. Teichmann and M.M. Babu. Gene regulatory network growth by duplication. *Nature Genetics*, 36(5):492–496, 2004.

[19] A. Bhan, D.J. Galas, and T.G. Dewey. A duplication growth model of gene expression networks, 2002.

[20] A.L. Barabasi. Scale-Free Networks: A Decade and Beyond. *Science*, 325(5939):412, 2009.

[21] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402(6761):47, 1999.

[22] S. Ishihara, K. Fujimoto, and T. Shibata. Cross talking of network motifs in gene regulation that generates temporal pulses and spatial stripes. *Genes to Cells*, 10(11):1025–1038, 2005.

[23] S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1):64–68, 2002.

[24] P.J. Ingram, M.P.H. Stumpf, and J. Stark. Network motifs: structure does not determine function. *BMC genomics*, 7(1):108, 2006.

[25] M.C. Costanzo, M.E. Crawford, J.E. Hirschman, J.E. Kranz, P. Olsen, L.S. Robertson, M.S. Skrzypek, B.R. Braun, K.L. Hopkins, P. Kondu, et al. YPDTM, PombePDTM and WormPDTM: model organism volumes of the BioKnowledgeTM Library, an integrated resource for protein information. *Nucleic Acids Research*, 29(1):75, 2001.

[26] 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, et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

[27] D. Thieffry, A.M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays*, 20(5), 1998.

[28] J.E. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in cell biology*, 14(2):140–148, 2002.

[29] I. Simon, J. Barnett, N. Hannett, C.T. Harbison, N.J. Rinaldi, T.L. Volkert, J.J. Wyrick, J. Zeitlinger, D.K. Gifford, T.S. Jaakkola, et al. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106(6):697–708, 2001.

[30] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nature genetics*, 31(1):60–63, 2002.

[31] R.J. Cho, M.J. Campbell, E.A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T.G. Wolfsberg, A.E. Gabrielian, D. Landsman, D.J. Lockhart, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell*, 2(1):65–73, 1998.

[32] D.A. Orlando, C.Y. Lin, A. Bernard, J.Y. Wang, J.E.S. Socolar, E.S. Iversen, A.J. Hartemink, and S.B. Haase. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, 453(7197):944–947, 2008.

[33] H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004.

[34] H. Kitano. Towards a theory of biological robustness. *Molecular systems biology*, 3(1), 2007.

[35] S.F. Elena, P. Carrasco, J.A. Daròs, and R. Sanjuán. Mechanisms of genetic robustness in RNA viruses. *EMBO reports*, 7(2):168–173, 2006.

[36] JM Carlson and J. Doyle. Complexity and robustness, 2002.

[37] R. Albert, H. Jeong, and A.L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[38] H. Jeong, SP Mason, A.L. Barabasi, and ZN Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

[39] D.M. Camacho and J.J. Collins. Systems Biology Strikes Gold. *Cell*, 137(1):24–26, 2009.

[40] D. Sprinzak and M.B. Elowitz. Reconstruction of genetic circuits. *Nature*, 438(7067):443–448, 2005.

[41] C. Khosla and J.D. Keasling. Metabolic engineering for drug discovery and development. *Nature Reviews Drug Discovery*, 2(12):1019–1025, 2003.