Robustness and Evolvability

Laura Langohr

May 14, 2010
Gene Regulatory Networks Seminar
Department of Computer Science
University of Helsinki

Abstract A biological system that retains its function performance under internal or external perturbations is robust. If it acquires novel functions (through genetic change) that advance the surviving and reproduction of the organism it is evolvable. In this seminar report we will show that robustness and evolvability are not contradictory, thereby mostly referring to Aldana et al. work. They used random Boolean networks to model gene regulatory networks and examined the dynamical attractors of the networks, which indicate robustness and evolvability. Their results show that random Boolean networks close to the critical regime have both, maximum robustness and maximum evolvability.

Contents

1 Introduction

A biological system is robust, if it retains its function performance under internal or external perturbations [1]. It is evolvable if it acquires novel functions (through genetic change) that advance the surviving and reproduction of the organism [2]. It has been argued that a consequence of robustness is reduced evolvability [3]. However, Aldana et al. [4] show that robustness and evolvability are not contradictory. They used random Boolean networks with homogeneous random and scale-free topology to model gene regulatory networks. For these the dynamical attractors are studied, which indicate robustness and evolvability of the networks. The results show that random Boolean networks close to the critical regime have both, maximum robustness and maximum evolvability.

Here, we will first introduce random Boolean networks (Section 2), and how gene duplication and divergence can be modeled in these (Section 3). Trajectories, attractors, and attractor landscapes are introduced in Section 4. In Section 5 we present visualisation of robustness and evolvability by attractor landscapes. Experiments and results of Aldana et al. [4] are described in Section 6. We conclude in Section 7 with some more general notes about the results.

2 Random Boolean Networks

Random Boolean networks are a special case of Boolean networks. Therefore, we will first explain Boolean networks and how they can be used to model gene regulatory networks. Afterwards we shortly state the specialty of random Boolean networks.

Boolean networks are simplified models of gene regulatory networks [5]. In Boolean networks expression levels are neglected, genes are either on (active) or off (not active). Alon [6] calls this logic approximation, as smoother Hill functions are approximated with step-functions.

By neglecting all parameters required in other models, information is lost, and detailed simulation might not be possible [7]. However, Boolean networks might still provide valuable information. Kauffman [8] states that this simplification (what he calls Boolean idealization) makes it possible to study enormously complex networks.

A Boolean network is a directed graph $G = (V, E)$ consisting of a node set V and an edge set $E \subseteq V \times V$ [5]. Nodes represent genes[1] and have either value 0 (off) or 1 (on). Edges represent gene regulations. An edge between nodes $u, v \in V$, pointing[2] from u to v means that gene u is regulating the expression of gene v. Each gene is regulated by some genes of the network [8]. To model this, a Boolean function is assigned to each node, which has the node's regulators as inputs and specifies the binary output value [5]. For an exemplary Boolean network see Fig. 1.

In Fig. 1 the Boolean function for v_1 could be, for example the logical equivalent of v_2; for v_2 the negation of v_1; and for v_3 the OR function of v_1 and v_2. Let us denote the set of regulators for gene v_n by $\{v_{n_1}, \ldots, v_{n_{k_n}}\}$, where $k_n \in \mathbb{N}$ [4]. Then, we can write $v_n(t+1) = f_n(v_{n_1}(t), \ldots, v_{n_{k_n}}(t))$, where f_n specifies the Boolean function for node n. For the above examples this would be then $v_1(t+1) = v_2(t)$, $v_2(t+1) = \neg v_1(t)$, and $v_3(t+1) = v_1(t) \vee v_2(t)$ (see Tab. 1).

Applying these Boolean functions, then $v_1(t+1) = 0$ if $v_2(t) = 0$ and $v_1(t+1) = 1$ if $v_2(t) = 1$. In contrast, if $v_1(t) = 0$ then $v_2(t+1) = 1$ and if $v_1(t) = 1$ then $v_2(t+1) = 0$. For $v_3(t+1)$ we have to look at $v_1(t)$ and $v_2(t)$. As we have to apply the OR function, $v_3(t+1) = 0$ only if $v_1(t) = 0$ and $v_2(t) = 0$ and in all the other cases $v_3(t+1) = 1$. In this way we can fill a table, which on the one hand consists of all gene states at time t and on the other hand of their states at time $t+1$ (see Tab. 2).

---

[1] In the following, the terms gene and node will be used indistinguishable.

[2] In terms of graph theory: the edge is adjacent to both, u and v, and the head is on the side of v.

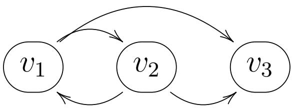

Fig. 1: A Boolean network example with three genes $v_1, \ldots, v_3$ and four edges.

v_n	f_n
v_1	v_2
v_2	$\neg v_1$
v_3	$v_1 \vee v_2$

Tab. 1: The genes of the Boolean network of Fig. 1 are listed with exemplary Boolean functions.

t			t+1		
v_1	v_2	v_3	v_1	v_2	v_3
0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	0	0	1
	⋮			⋮	
1	1	1	0	1	1

Tab. 2: States of the genes of the Boolean network of Fig. 1 at time t and time $t+1$ applying the Boolean functions of Tab. 1.

In *random Boolean networks* the regulators (or the targets) of each gene are chosen randomly, after the number of regulators k_n (targets l_n) has been specified. The Boolean functions are chosen randomly from the set of all 2^{k_n} possible Boolean functions [4]. Here, we will choose the Boolean function such that $f_n = 1$ with probability p, and $f_n = 0$ with probability $p - 1$. The parameter p can be interpreted as the probability of gene expression.

3 Gene Duplication and Divergence

Gene duplication and divergence are the only perturbation considered, here. In this section we will show how gene duplication and divergence can be simulated in a (random) Boolean network.

Gene duplication and divergence are simulated in a Boolean network by randomly choosing one gene, duplicating it, and mutating the duplicate [4] (see Fig. 2). Mutations are simulated by rewiring some of its regulators and targets, as well as changing randomly the Boolean functions for the duplicate and the its targets.

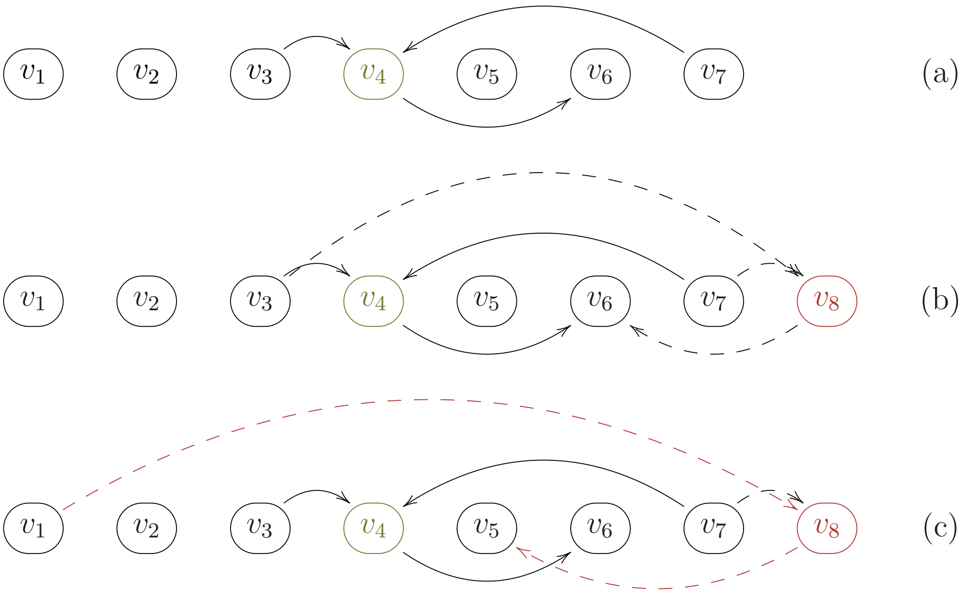

Fig. 2: Gene duplication and divergence in a Boolean network (adapted from [4]). Only regulators and targets of genes v_4 and v_8 are shown. (a) Original network with seven genes. (b) Gene v_4 is duplicated. The result is gene v_8 with same regulators and targets as v_4. (c) Gene v_8 is mutated by rewiring some of its regulators and targets.

Teichmann and Babu [9] showed that 90% of new gene interaction in *E.coli* and *S.cerevisiae* evolved by duplication of either their regulators or targets and following divergence. Here, we will assume divergence always performing directly after duplication. Gu et al. [10] showed that duplicate genes indeed diverge quickly and eventually become divergent in expression.

In case of the latter, the target genes of the duplicate have got an extra regulator [4]. Therefor, their Boolean functions have to be extended. This will be done by copying the values, if the duplicate is inactive (value 0), and may get the same or new values, if the duplicate is active (value 1) (see Tab. 3). In the following, we will consider only cases, where all possible mutations are carried out. This is an extreme case, by which the lower bound of robustness will be observed.

v_2	v_3	f_5
0	0	0
0	1	1
1	0	1
1	1	1

(a)

v_2	v_3	v_8	f_5
0	0	0	0
0	1	0	1
1	0	0	1
1	1	0	1
0	0	1	0
0	1	1	1
1	0	1	0
1	1	1	0

(b)

Tab. 3: Extension of the Boolean function f_5 of gene v_5 that got an extra regulator (adapted from [4]). (a) Gene v_5 has two regulators v_2 and v_3. (b) After gene duplication and divergence it got v_8 as a third regulator. For $v_8 = 0$ f_5 has the same values as before. For $v_8 = 1$ the values of f_5 are randomly assigned.

4 Trajectories and Attractors

In this section we will show how (random) Boolean networks can be used for exploring the evolution of a (gene regulatory) network over time. Therefore, we will describe what trajectories, attractors, basins of attraction and attractor landscape are.

Given a random Boolean network, with regulators and Boolean functions specified, we want to study how the network is evolving. It is important to notice, once regulators and the Boolean function are chosen for each node, they will remain fixed [4]. The interest lies in exploring the temporal evolution of the network. The vector of the node values (or a bit string) is the state (or dynamical configuration) of the network [4, 5]. Starting with an initial state, the network is run over various time steps. Here, we consider the simple case, where all network nodes are updated synchronously [8, 5].

The *state space* consists of all possible $2^{|V|}$ network states [11]. See Tab. 5 for all possible states of the Boolean network in Fig. 3 and their states after one step in time. The visited network states form a *trajectory*, that is, the path of the network states through the state space. The trajectory can be visualized as a state transition graph. For the Boolean network of Fig. 3 the state transition graph is shown in Fig. 4. There, for example, 0010 → 0001 → 0000 is part of a trajectory.

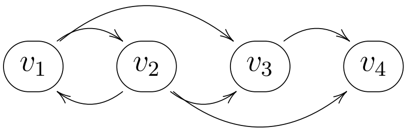

Fig. 3: A Boolean network example with four genes $v_1, \ldots, v_4$ and six edges.

v_n	f_n
v_1	v_2
v_2	v_1
v_3	$v_1 \wedge v_2$
v_4	$v_2 \vee v_3$

Tab. 4: All nodes of the Boolean network of Fig. 4 are listed with exemplary Boolean functions.

Any trajectory must sooner or later encounter a previously visited state [11]. The trajectory then entered an trajectory cycle, called *attractor*, in which it will stay. An attractor cycle may

t				t+1			
v_1	v_2	v_3	v_4	v_1	v_2	v_3	v_4
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1
0	0	1	1	0	0	0	1
0	1	0	0	1	0	0	1
	⋮				⋮		
1	1	1	1	1	1	1	1

Tab. 5: The left four columns show all possible states of the Boolean network of Fig. 3. On the right the states to which they are updated after one time step according to the Boolean functions in Tab. 4 are shown.

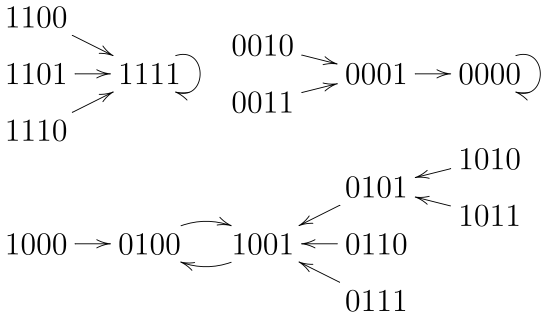

Fig. 4: The attractor landscape of the Boolean network in Fig. 3 as a state transition graph.

consist of several (like 0100 and 1001 in Fig. 4) or only one node (like 0000 or 1111 in Fig. 4).

In a network there might exist several attractors, but at least one must exist [4]. There has to be at least one attractor, as there is only a limited number of states and a trajectory always encounters from one state a next (another or the same) state. Hence, a trajectory sooner or later encounters a previously visited state, that is, it encountered a trajectory circle, called attractor.

An attractor can be interpreted as a gene expression profile, which the genetic network falls back into after some time [4]. Assume, for example, an *E.coli* that swims in a plate. Normally it will swim straight forward, changing direction approximately every second [12]. If it detects a nutrient gradient it will swim in the direction of that gradient without changing direction anymore. However, when there is no spatial gradient, but the nutrient quantity has increased everywhere with the same amount, the bacterium will (after some time) fall back into changing direction every second - the bacterium's attractor.

All states that encounter the same attractor after some time form the attractor's *basin of attraction* [4]. The basis of attraction might only consist of the attractor itself [11]. Thus, all $2^{|V|}$ possible network states organize into one or several disjoint classes, each consisting of an attractor and its basin of attraction. The set of disjoint classes is called *attractor landscape*, and can be visualized as a state transition graph (see Fig. 4).

Boolean networks have fan-like structures, where several network states have the same successor state [4] (see Fig. 4 and 16). Thus, the information of the initial condition is lost over time. All trajectories that lead to the same attractor can be interpreted as gene expression pathways that produce, for example, the same cell type or cell fate, which is encoded in the attractor [8, 4].

5 Robustness and Evolvability

Now, we can define robustness of a random Boolean network as conservation of all attractors under perturbations, and evolvability as the emergence of new attractors under perturbation [4]. Under perturbation, some network states might jump to another subtree in the same basin of attraction, or it might even jump to another basin [11]. In both cases the attractors remain unchanged, indicating stability against perturbations. This means, differentiation and gene expression pathways might have changed, but the phenotypes are still the same [4]. In contrast, a perturbation might also result in a new attractor, which indicates evolvability [11], and thus changes in the phenotypes [4].

Looking at the attractors and their basins of attractions four different transformations between attractors are possible: identity, expansion, contradiction, and innovation [4]. In Tab. 6 the attractors before and after duplication and divergence of one gene are shown. The original network has three attractors $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$, and the network after duplication and divergence has four attractors $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$, and $\mathcal{B}_4$. In the case of *identity* the attractor and its basin of attraction remained the same after duplication and divergence ($\mathcal{A}_1 = \mathcal{B}_1$), that is, exactly the same network states lead to one attractor. For comparison of two network states, one before, the other after gene duplication and divergence, the value of the duplicate gene is ignored. In case of *expansion* the basin of attraction enlarged ($\mathcal{A}_2 \subset \mathcal{B}_2$), and in case of *contraction* it was reduced ($\mathcal{A}_3 \supset \mathcal{B}_3$). An *innovation* is observed, when a new attractor emerged after duplication and divergence (such as $\mathcal{B}_4$).

Original network	Network after duplication and divergence of a gene	Transfor-mation	Original network	Network after duplication and divergence of a gene	Transfor-mation
Attractor $\mathcal{A}_1$:	Attractor $\mathcal{B}_1$:	Identity	Attractor $\mathcal{A}_3$:	Attractor $\mathcal{B}_3$:	Contraction
00011011011	000110110111		11100010011	111000100110	
01100010011	011000100110		11011010010	110110100100	
11000101110	110001011101		10011011101		
11001101010	110011010100		11011011101	1110110111011	
			01100101101		
Attractor $\mathcal{A}_2$:	Attractor $\mathcal{B}_2$:	Expansion		Attractor $\mathcal{B}_4$:	Innovation
10111011101	101110111011			0001001100101	
01101101011	011011010110			0000110101111	
	000001000101			0100110100101	
01011010001	010110100011			0100110100000	
	001100110100				

Tab. 6: Attractor landscape before and after gene duplication and divergence (adapted from [4]). The original network (first column) consist of 11 genes, and three attractors. After divergence and duplication of one gene (second column) there are four attractors. Network states in the same row are identical, thereby ignoring the 12th value in the second column, which is the value of the duplicated gene. Identity, expansion, contraction, and innovation of attractors are shown.

Notice, that it is possible that all attractors remain with identity, but still new attractors might emerge [4]. In terms of biology this means that all phenotypes are preserved, and still new ones are acquired. In the following it will be demanded that the network after gene duplication and divergence has the same number of attractors as before. Thus, we restrict the network to develop new phenotypes while preserving the old ones over time. The restriction is not very severe as the number of attractors will, on average, grow with the size of genes.

6 Experiments and Results

We will report Aldana et al. [4] experiments and results in this section. Two different network topologies, homogeneous random and scale-free output topology, will be considered. Ordered, critical, and chaotic regimes will be studied for small and critical regimes for larger networks. We will first introduce these concepts, then describe Aldana et al. experiment settings, and finally report on their results.

6.1 Topologies

Aldana et al. [4] considered homogeneous random topology because it has been widely used for studying gene regulatory networks. However, scale-free output topology seem to better reflect the experimentically observed gene regulatory networks [13, 14] and thus Aldana et al. chose this as a second topology for their experiments.

In a network with *homogeneous random topology* each gene has exactly $k_n = k$ regulators, which are randomly chosen. Thus, the regulator probability follows the Kronecker delta function:

$$p_{reg}(k') = \delta_{k',k} = \begin{cases} 1, & \text{if } k' = k \\ 0, & \text{if } k' \neq k \end{cases}$$

The targets are chosen randomly. Thus, the target probability distribution is a Poisson distribution with an average $a = k$:

$$p_{tar}(l) = e^{-a}a^l/l!$$

This means, that most of the genes have $a = k$ targets, and the probability of having more (less) targets is reducing smoothly with decreasing (increasing) a.

In a network with *scale-free output topology* genes have l_n targets, which is a random variable that follows the power law distribution:

$$p_{tar}(l) = Cl^{-\gamma}$$

Hence, with high probability a gene has few targets and with low probability a gene has many targets.

The regulators are chosen randomly (similarly to the targets in the homogeneous random topology). Thus, the number of regulators k_n is a random variable that follows the Poisson distribution with an average a determined by the scale-free exponent γ (and the number of genes $|V|$ in the network):

$$p_{reg}(k) = e^{-a}a^k/k!$$

To show that networks with scale-free output topology do reflect the experimentically observed gene regulatory networks well Aldana et al. [4] plotted the regulator and target distributions for the transcriptional regulatory networks of *E.coli*, *B.subtilis* and *S.cerevisiae* (see Fig. 5 and 6) of previously published data [15, 16, 17]. Aldana et al. then examine which distributions approximate the observed regulator and target distributions well[3].

The regulator distribution of *E.coli* can be approximated by a Poisson distribution with $k = 2.16$ (Fig. 5(a)) and its target distribution by a power law distribution with $\gamma \approx 1.636$ (Fig. 6(a)). For *B.subtilis* the regulator distribution can be approximated by a Poisson distribution with $k = 1.3$ (Fig. 5(b)) and its target distribution by a power law distribution with $\gamma \approx 1.673$ (Fig. 6(b)). For *S.cerevisiae* the regulator distribution is better approximated by a exponential distribution $e^{-\alpha l}$ with $\alpha = 0.5$ (Fig. 5(c)), and its target distribution can be approximated by a power law distribution with $\gamma \approx 0.984$ (Fig. 6(c)).

As the regulator distributions for the three organisms have short tails (approximated by Poisson or exponential distributions) and the target distributions have long tails, it is justified to model gene regulatory networks with the above described scale-free output topology.

---

[3] However, it does not seem to me that Aldana et al. did carry out any broad study of different distributions with different parameters to find the distribution that fits the data best.

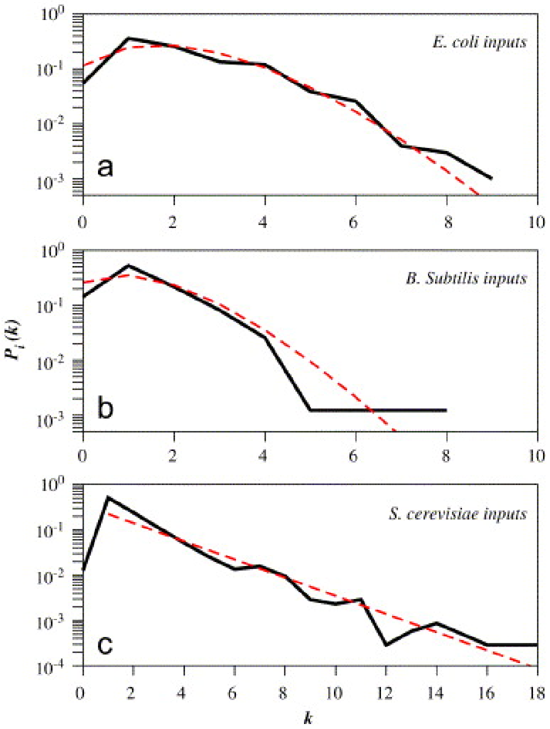

Fig. 5: Regulator distribution for (a) *E.coli*, (b) *B. subtilis*, and (c) *S.cerevisiae*. The solid line corresponds to the experimental data and the dashed lines to distributions that approximate the data well, that is, for (a) a Poisson distribution with $k = 2.16$, for (b) a Poisson distribution with $k = 1.3$, and for (c) a exponential distribution $e^{-\alpha l}$ with $\alpha = 0.5$. All graphs are in log-linear scale. Source: [4].

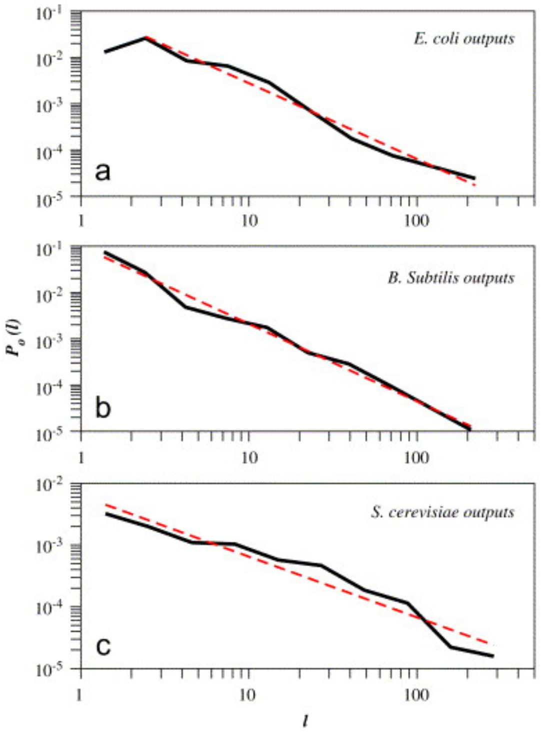

Fig. 6: Target distribution for (a) *E.coli*, (b) *B.subtilis*, and (c) *S.cerevisiae*. The solid line corresponds to the experimental data and the dashed lines to distributions that approximate the data well, that are, power law distributions with (a) $\gamma \approx 1.636$, (b) $\gamma \approx 1.673$, and (c) $\gamma \approx 0.984$. All graphs are in log-log scale. Source: [4].

6.2 Ordered, critical and chaotic regimes

Random Boolean networks with above topologies can be either in a ordered or in a chaotic regime [4]. The phase transition occurs at $S = 2p(1-p)k$. For $S < 1$ the network is in the ordered regime, for $S > 1$ in the chaotic (disordered) regime. The critical regime (phase transition) occurs at $S = 1$. In their work Aldana et al. [4] set $p = 0.5$ and thus the critical regime occurs at $k = 2$. Hence, by varying k, the system behavior can be changed from ordered ($k < 2$) to chaotic ($k > 2$).

One characteristic of a chaotic regime is, that small changes in the initial conditions have big effects on the outcome, that is, the whole system is sensitive to the initial conditions. In a ordered regime, small changes in the initial conditions die out over time, that is, the systems converges over time. Rämö et al. [5] research on microarray data indicates that gene regulatory networks are stable and reside near the critical regime. Aldana et al. [4] shows that in this case the organism obtains maximum robustness and evolvability simultaneously.

6.3 Experiments

For the experiments Aldana et al. [4] used original networks consisting of 20 (or 200) genes, duplicating and diverging one gene in each. A network is realized 20,000 times, and average statistics are studied.

Notice, that for networks with scale-free topology genes are not statistically equivalent. Most of the genes have just few connections. Thus, the duplicate gene will most likely be also poorly connected. Aldana et al. [4] are not aware of experimental evidence, that highly connected genes would be duplicated more likely than poorly connected ones in living organisms. Hence, the gene to be duplicated will be chosen randomly for scale-free networks as well as for homogeneous random networks.

For $|V| = 20$ the whole network state space can be explored. For large networks ($|V| = 200$), only a subset ($\Omega_S = 10^6$) of the whole network state space ($\Omega = 2^{200} \approx 10^{60}$) was explored, due to limitations in computer memory as well as computing time. Aldana et al. sampled a subset of the network state space and hoped most (if not all) attractors were still found. Here, we first report the results for small networks with both topologies, and show afterwards the results for large networks[4].

Aldana et al. [4] restricted the analysis of the large networks to the critical regime ($k = 2$). The ordered regime ($k < 2$) has most likely only one attractor. Thus, the results do not differ from those for smaller networks. The critical regime ($k > 2$) again has extremely long transient times and long attractor length. Hence, it was practically impossible to find their attractors in a reasonable time.

6.4 Results for homogeneous random networks

The probability that the number of attractors after duplication and divergence is equal or increased is shown in Fig. 7 as a function of k. It is higher for networks in ordered ($k < 2$) and critical regime ($k = 2$) than for networks in chaotic regime ($k > 2$). The probability is always larger than 1/2, thus, it is most likely that the network after duplication and divergence has at least the same number of attractors as before. Note that for the network in an ordered regime ($k < 2$) it is more likely that the number of attractors remains the same, and for chaotic regime ($k > 2$) it is more likely that the number of attractors increases after gene duplication and divergence.

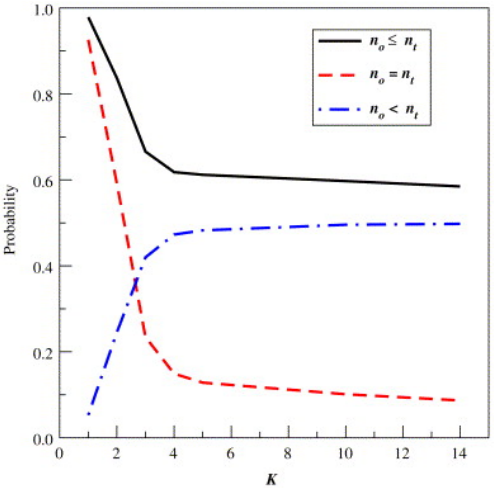

Fig. 7: Probability that the number of attractors after duplication and divergence is equal or increased (solid line, $n_0 \leq n_t$) as a function of k. It is split up in the probabilities that the number of attractors is equal (dashed line, $n_0 = n_t$) and increased (dotted-dashed line, $n_0 \geq n_t$). Source: [4].

In Fig. 8 the percentages q of attractors, which are conserved after gene duplication and divergence are shown as probabilities over 20,000 network realizations. For networks in ordered ($k < 2$) and critical regime ($k = 2$) the maximum occurs at $q = 100\%$, indicating, all attractors are conserved with high probability. For networks in chaotic regime ($k > 2$) the peak at $q = 100\%$ decreases and is almost vanished already for $k = 4$. Moreover, the maximum for networks in chaotic regimes is at $q = 0\%$, which indicates, that attractors are not conserved at all. Hence, networks in ordered and critical regimes are much more robust that networks in chaotic regimes. Thereby, networks in ordered regime seems to be the most robust.

Most likely, when attractors are conserved, they are identically conserved, and seldom expanded (see Fig. 9). In Fig. 10 the probability of new attractors occurring after gene duplication

---

[4] Note, that for the rest of this section I will leave out citations as all results are from Aldana et al. [4].

and divergence given that all original attractors are conserved is shown. For a network in critical regime ($k = 2$) it is maximum, thus, networks at the critical regime have maximum evolvability.

6.5 Results for scale-free networks

In Fig. 11 the percentages q of attractors, which are conserved after gene duplication and divergence are shown as probabilities over 20,000 network realizations. For networks in ordered ($k < 2$) and critical regime ($k = 2$) the maximum occurs at $q = 100\%$, indicating, all attractors are conserved with high probability. The values are similar to those shown in Fig. 8 for homogeneous random networks. For networks in chaotic regime ($k > 2$) the peak at $q = 100\%$ decreases. However it is decreasing more slowly than for homogeneous random networks (Fig. 8). Hence, for chaotic regimes scale-free networks seem to be more robust that homogeneous random networks.

The reason for this may lies in the random selection of genes to be duplicated and the high heterogeneity in the connectivity of the genes. Duplicated genes with a large number of regulators and targets do more likely change the dynamical attractors of the network than duplicated genes with a small number of regulators and targets. See Fig. 12 for the probability that all original attractors are conserved as a function of l, the number of targets of the duplicate gene. The probability decreases rapidly and therefore duplicate genes with few targets are more likely to conserve the attractor landscape. Hence, as we chose the genes to be duplicated randomly, they will have most likely few regulators and targets, which results again in the high probability of conserving all attractors.

In Fig. 13 the probability of new attractors occurring after gene duplication and divergence given that all original attractors are conserved is shown. As for homogeneous random networks, it is maximum for a network in critical regime ($k = 2$). Thus, also scale-free networks have maximum evolvability at the critical regime.

6.6 Results for large networks

For networks with 200 genes only a randomly chosen subset of 10^6 network states the critical regime ($k = 2$) was studied. In Fig. 14 percentages q of attractors, which are conserved after gene duplication and divergence are shown as probabilities over 10^6 network states of 20,000 network realizations. For both, homogeneous random (Fig. 14 (a)) and scale-free topology (Fig. 14 (b))

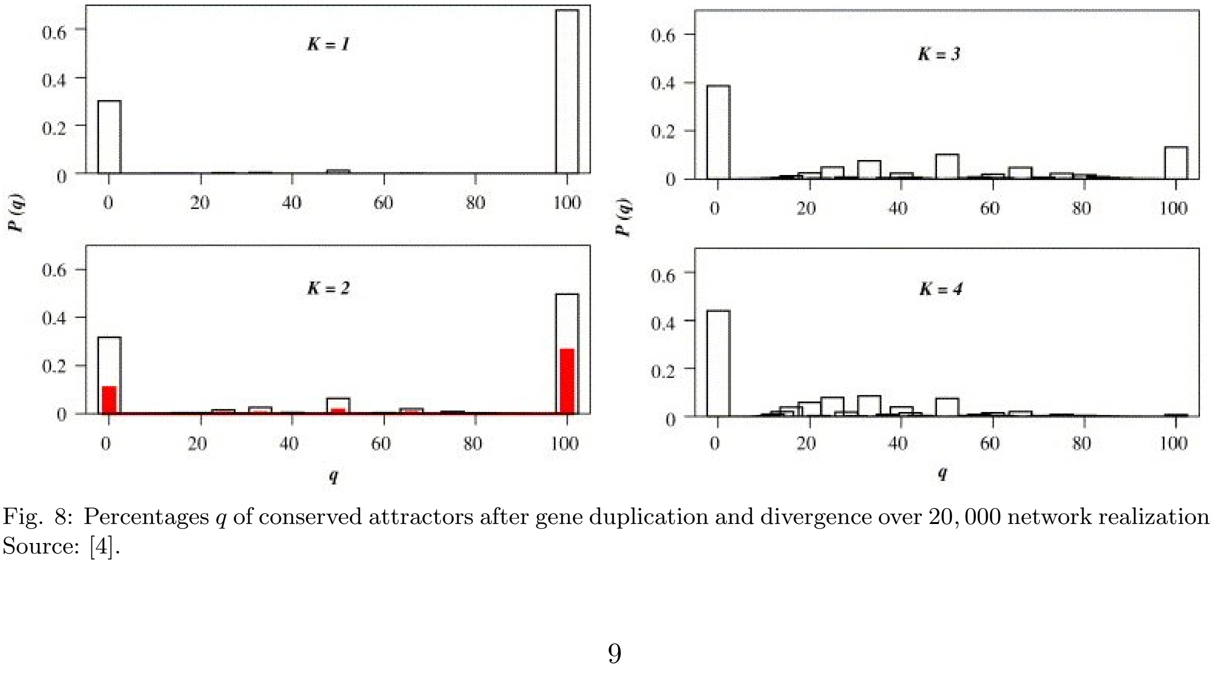

Fig. 8: Percentages q of conserved attractors after gene duplication and divergence over 20,000 network realizations. Source: [4].

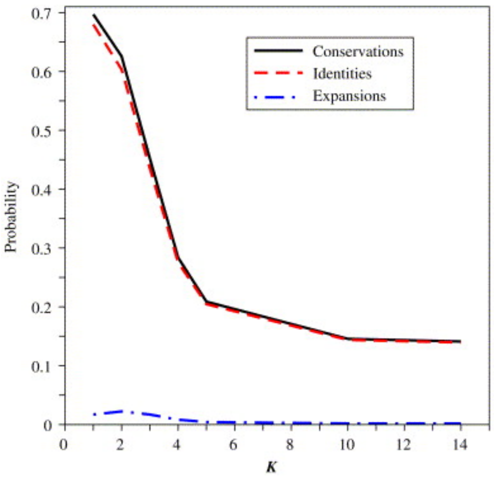

Fig. 9: Probabilities of conservations (solid line) decomposed in identical conservations (dashed line) and expansions (dotted-dashed line). Source: [4].

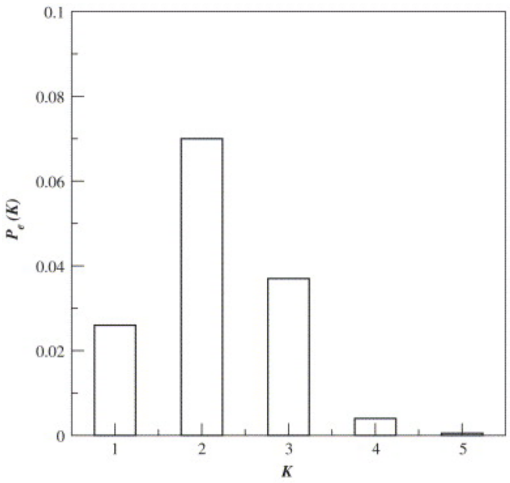

Fig. 10: Probability of new attractors occurring after gene duplication and divergence given that all original attractors are conserved. Probabilities for $k = 1, \dots, 5$ are shown. Source: [4].

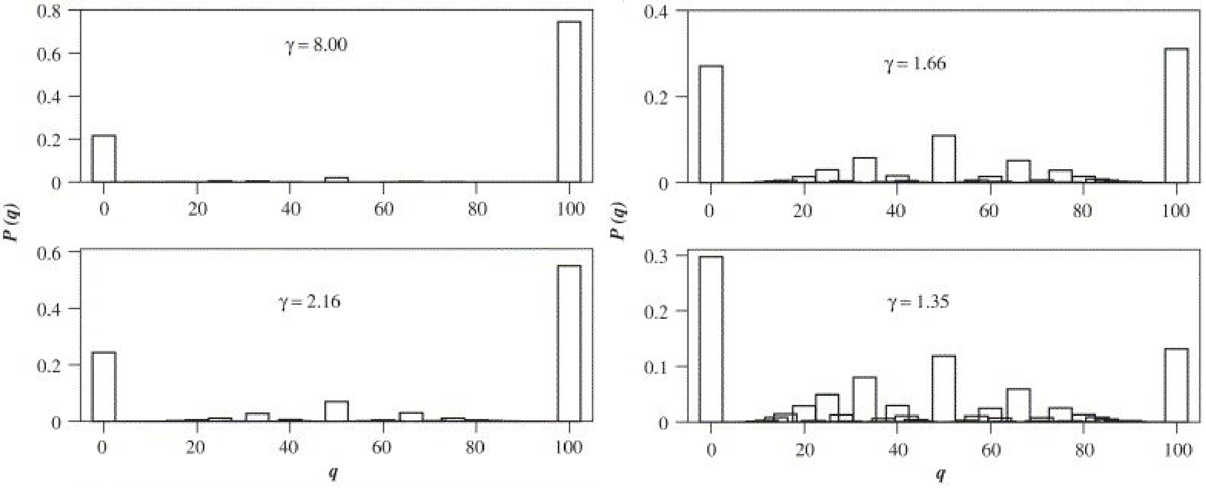

Fig. 11: Percentages q of conserved attractors after gene duplication and divergence over $20,000$ network realizations. Note, that $k = 1$ for $\gamma = 8.00$, $k = 2$ for $\gamma = 2.16$, $k = 3$ for $\gamma = 1.66$, and $k = 4$ for $\gamma = 1.35$. Source: [4].

the plots are very similar to those with smaller networks (Fig. 8 and 11). As due to the random selection of network states some attractors might not have been recognized as conserved, even though, they are conserved. Thus, it seems that all, or almost all, attractors are still found.

As can be seen in Fig. 14 (and in Fig. 8 and 11), with a high probability either all or none attractors are conserved. It could be presumed that the reason for this is that the original network has only one attractor, which is either conserved or not. However, this is mostly not the case: see Fig. 15(a) for the probabilities of having $n = 1, \dots, 100$ attractors. In Fig. 15(b) the probability of attractor conservations is shown as a function of the number of attractors in the original network. The conditional probability $p(100|n)$ of conserving all attractors given that there were n attractors in the original network is a slowly decaying function. Moreover, the probability of conserving one attractor $p(100|n = 1) \approx 0.6$ is only twice as big as conserving 35 attractors $p(100|n = 1) \approx 0.3$. Therefore, the all-or-none behavior in Fig. 14 can not only be described by the number of attractors.

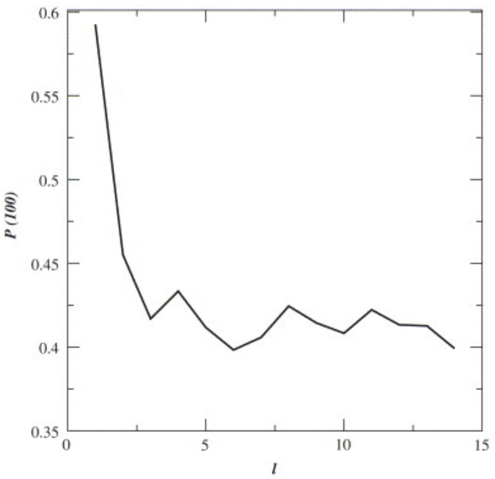

Fig. 12: Probability that all original attractors are conserved as a function of number of targets l of the duplicate gene, $l = 1, \ldots, 14$. Source: [4].

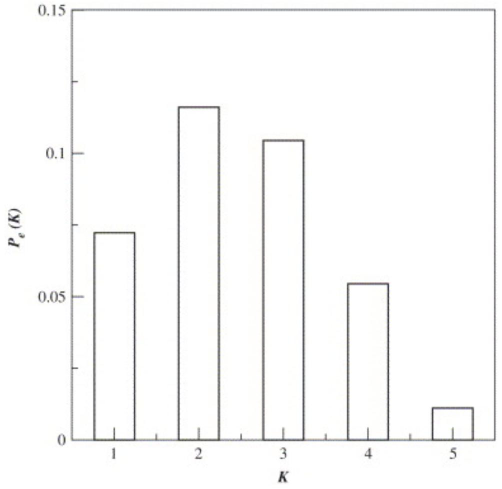

Fig. 13: Probability of new attractors occurring after gene duplication and divergence given that all original attractors are conserved. Probabilities for $k = 1, \ldots, 5$ are shown. Source: [4].

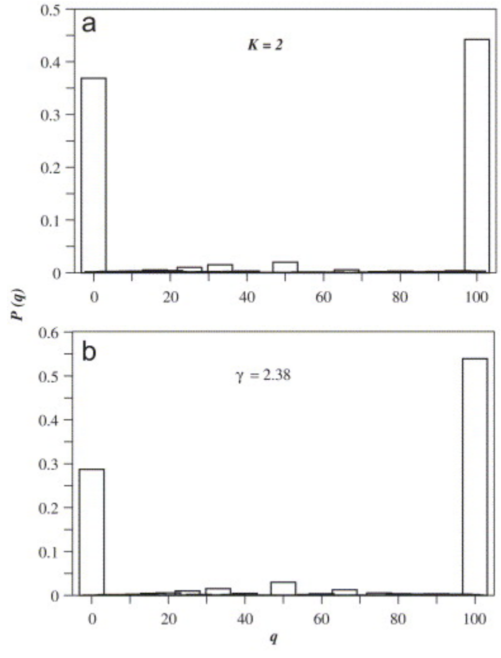

Fig. 14: Percentages q of conserved attractors after gene duplication and divergence over 10^6 network states of $20,000$ network realizations for networks with (a) homogeneous random topology and (b) scale-free topology. Note, that $k = 2$ for $\gamma = 2.16$. Source: [4].

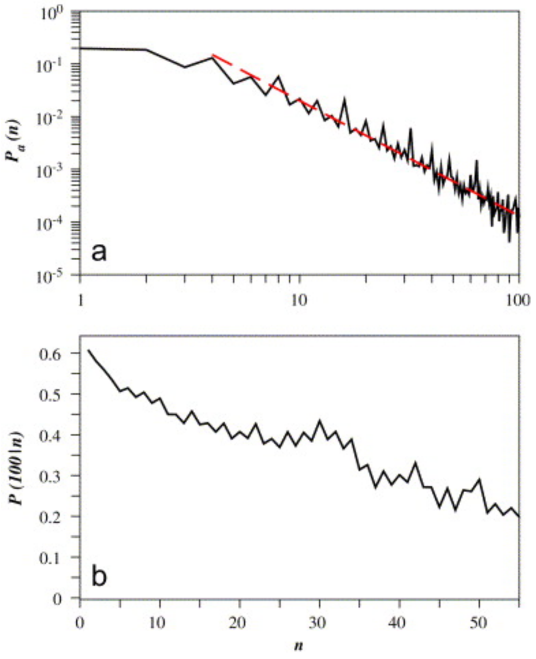

Fig. 15: (a) Probability $p_a(n)$ that the original network has n attractors. (b) Conditional probability $p(100|n)$ that all attractors are conserved given that there were n attractors in the original network. For both, (a) and (b) a scale-free topology, 200 genes, $k = 2$, and $20,000$ network realizations were used. Source: [4].

6.7 Attractor landscape

In Fig. 16 an exemplary attractor landscape for a network with $k = 2$ and $|V| = 15$[5] are shown. The network is robust as all three attractors $(\mathcal{A}_1, \ldots, \mathcal{A}_4)$ are conserved (now named $\mathcal{B}_1, \ldots, \mathcal{B}_4$) after duplication and divergence of one gene. At the same time the network is evolvable as a new

---

[5] Aldana et al. do not specify what topology was used.

attractor ($\mathcal{B}_5$) emerged.

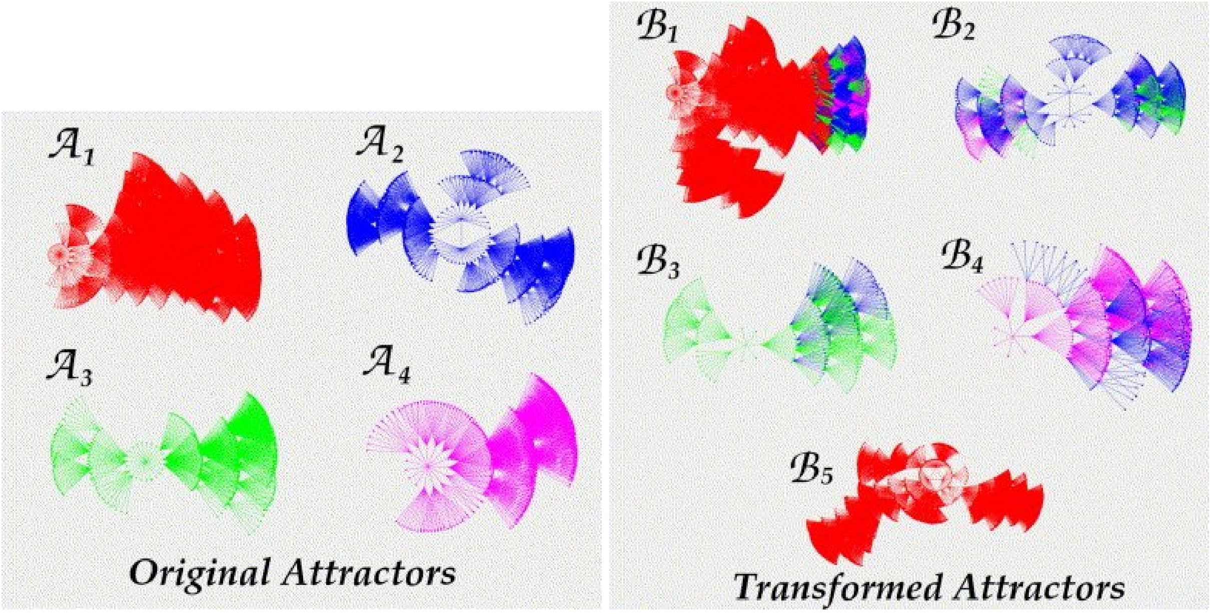

Fig. 16: Attractor landscapes of an exemplary network with $k = 2$ and $|V| = 15$ before (original attractors) and after (transformed attractors) duplication and divergence of one gene. The original attractors ($\mathcal{A}_1, \ldots, \mathcal{A}_4$) are conserved ($\mathcal{B}_1, \ldots, \mathcal{B}_4$)and a new attractor ($\mathcal{B}_5$) emerged. The network states kept their color after the gene duplication and divergence, thus, visualizing the "exchange" of network states between the basins of attraction. Source: [4].

Identical network states (ignoring the value of the 16th gene) have the same color after gene duplication and mutation. Thus the "exchange" of network states between the basins of attraction is visualized. It can be observed that each attractor after gene duplication and divergence is dominated by the same color as before. Though, they also contain network states of different colors. Even though only one gene is duplicated and mutated, the whole attractor landscape is reconfigured. In a biological point of view this means that duplication and divergence of a single gene changes gene expression and differentiation pathways.

7 Conclusion

We have described that simultaneous existence of robustness and evolvability in a gene regulatory network is not contradictory. We based our seminar report mainly on Aldana et al. [4] work, who modeled gene regulatory networks with random Boolean networks for 20 and 200 genes, with two different topologies, and studied the attractor landscapes to analyze the robustness and evolvability of the networks.

Their result show that networks in the ordered and critical regime are robust against gene duplication and divergence (Fig. 8, 11, and 14). Networks in the chaotic regime are much less robust, even though, the change is not so drastic for scale-free networks, as there exist many low connected genes. As real gene regulatory network are robust, these result indicate that they operate in the critical or near-critical regime.

Further, the results show that networks in the critical regime exhibit the maximum probability that new attractors and thus new phenotypes occur (Fig. 10, and 13). These phenotypes might be beneficial for adaption to new environments or might not, but they anyway serve as the pool of phenotypes on which natural selection acts.

Hence, networks in critical regime exhibit robustness and evolvability at the same time, schematically visualized in Fig. 16. Thus, this seminar report describes how a gene regulatory network can simultaneously exhibit both, robustness and evolvability, under the process of gene duplication and divergence.

References

[1] H. Bolouri. **Computational Modeling of Gene Regulatory Networks – a Primer.** *Chapter 18: Robustness Analysis.* Imperial College Press, London, UK, 2008.

[2] A. Wagner. **Robustness, Evolvability, and Neutrality**. *FEBS Letters*, 579(8):1772 – 1778, 2005.

[3] J.A.G.M. de Visser, J. Hermisson, G.P. Wagner, L.A. Meyers, H. Bagheri-Chaichian, J.L. Blanchard, L. Chao, J.M. Cheverud, S.F. Elena, W. Fontana, G. Gibson, T.F. Hansen, D. Krakauer, R.C. Lewontin, C. Ofria, S.H. Rice, G. von Dassow, A. Wagner, and M. C. Whitlock. **Perspective: Evolution and Detection of Genetic Robustness**. *Evolution*, 57(9):1959–1972, 2003.

[4] M. Aldana, E. Balleza, S. Kauffman, and O. Resendiz. **Robustness and Evolvability in Genetic Regulatory Networks**. *Journal of Theoretical Biology*, 245(3):433 – 448, 2007.

[5] P. Rämö, J. Kesseli, and O. Yli-Harja. **Perturbation Avalanches and Criticality in Gene Regulatory Networks**. *Journal of Theoretical Biology*, 242(1):164–170, 2006.

[6] U. Alon. **An Introduction to Systems Biology.** *Chapter 2: Transcription Networks: Basic Concepts.* CRC Press, Boca Raton, FL, USA, 2007.

[7] S. Bornholdt. **Systems Biology: Less Is More in Modeling Large Genetic Networks**. *Science*, 310(5747):449–451, 2005.

[8] S.A. Kauffman. **The Origins of Order: Self Organization and Selection in Evolution.** *Chapter 5: Self-Organization and Adaption in Complex Systems.* Oxford University Press, 1993.

[9] S.A. Teichmann and M.M. Babu. **Gene Regulatory Network Growth by Duplication**. *Nature Genetics*, 36(5):492–496, 2004.

[10] Z. Gu, D. Nicolae, H.H.-S. Lu, and W.-H. Li. **Rapid Divergence in Expression Between Duplicate Genes Inferred from Microarray Data**. *Trends in Genetics*, 18(12):609–613, 2002.

[11] A. Wuensche. **Modularity in Development and Evolution.** *Chapter 13: Basins of Attraction in Network Dynamics: A conceptual framework for biomolecular networks.* eds. G. Schlosser and G.P. Wagner. Chicago University Press, Chicago, IL, USA, 2004.

[12] U. Alon. **An Introduction to Systems Biology.** *Chapter 7: Robustness of Protein Circuits: The Example of Bacterial Chemotaxis.* CRC Press, Boca Raton, FL, USA, 2007.

[13] N. Guelzim, S. Bottani, P. Bourgine, and F. Kps. **Topological and Causal Structure of the Yeast Transcriptional Regulatory Network**. *Nature Genetics*, 31:60–63, 2002.

[14] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann. **Structure and Evolution of Transcriptional Regulatory Networks**. *Current Opinion in Structural Biology*, 14(3):283–291, 2004.

[15] A. Martínez-Antonio and J. Collado-Vides. **Identifying Global Regulators in Transcriptional Regulatory Networks in Bacteria**. *Current Opinion in Microbiology*, 6(5):482—489, 2003.

[16] Y. Makita, M. Nakao, N. Ogasawara, and K. Nakai. **DBTBS: Database of Transcriptional Regulation in Bacillus Subtilis and its Contribution to Comparative Genomics**. *Nucleic Acids Research*, 32:D75—77, 2004.

[17] 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*, 431:308—312, 2004.