Robustness of Boolean dynamics under knockouts

Gunnar Boldhaus,[1] Nils Bertschinger,[2] Johannes Rauh,[2] Eckehard Olbrich,[2] and Konstantin Klemm[1]

[1]*Bioinformatics Group, Institute for Computer Science,*
University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany
[2]*Max Planck Institute for Mathematics in the Sciences,*
Leipzig, Germany, Inselstraße 22, D-04103 Leipzig, Germany
(Dated: July 28, 2010)

The response to a knockout of a node is a characteristic feature of a networked dynamical system. Knockout resilience in the dynamics of the remaining nodes is a sign of robustness. Here we study the effect of knockouts for binary state sequences and their implementations in terms of Boolean threshold networks. Besides random sequences with biologically plausible constraints, we analyze the cell cycle sequence of the species *Saccharomyces cerevisiae* and the Boolean networks implementing it. Comparing with an appropriate null model we do not find evidence that the yeast wildtype network is optimized for high knockout resilience. Our notion of knockout resilience weakly correlates with the size of the basin of attraction, which has also been considered a measure of robustness.

I. INTRODUCTION

Living systems show a ubiquitous *robustness* against mutations, environmental changes and intrinsic non-determinism [1–3]. In particular, each single cell must control its growth and eventual division by regulating concentrations of proteins in a precise temporal pattern. Using a Boolean state dynamics [4], this cell cycle network has been argued to be *robustly designed* for budding yeast as a model organism [5]. The robustness has been pinpointed as reproducibility of the dynamics in the presence of stochastic perturbations [5–7]. Resilience against mutations, *i.e.* changes of the interactions among the proteins, has been studied as well [5, 8, 9].

A particular type of mutation, either intrinsic or by intervention, is a complete knockout of a single protein. A gene is made dysfunctional such that it is no longer transcribed and its product is effectively removed from the cell. Some knockouts can be tolerated or compensated by the cell, whereas others are lethal. Knockouts are often used to infer the function of specific proteins. This is only appropriate, if the knockout is neither lethal nor fully compensated, but disables a specific function of the cell. Importantly in real experiments, knockout resilient systems are more difficult to analyze and identify because knockout mutants do not exhibit a measurable difference.

In this contribution we study the resilience against knockouts in two scenarios. First, the system under consideration is defined only by a sequence of activation patterns regardless of the specific mechanism producing them ("black box"). Here, resilience with respect to knockout of a node means that the information contained in the activation patterns of all other nodes is still sufficient to unambiguously produce the original sequence. In the second scenario, we consider the sequence together with a given implementation by a Boolean threshold network. Knocking out node j means that we remove its interactions with the other nodes. The network is resilient against this knockout if all other nodes still perform the original sequence of activation patterns.

After defining these notions of knockouts and resilience we apply them to the yeast cell cycle sequence and its network implementations. Significance of the results is assessed by comparison of random sequences as null models with various constraints.

II. KNOCKOUTS AND ROBUSTNESS OF FUNCTION

A. Definitions

Molecular processes within cells are frequently modeled as dynamics with Boolean states [4, 10, 11]. Components of a Boolean state vector correspond to different molecules which can be present in either high or low concentrations. The dynamics is thus described as a temporal sequence of activation patterns $x(0), x(1), \ldots$, where each activation pattern x has n binary components x_i, $i = 1, \ldots, n$. Low and high concentration of molecule i at time t are denoted by $x_i(t) = 0$ and $x_i(t) = 1$, respectively. Each component i computes a Boolean function f_i mapping the present concentration pattern to its activation at the next time step $x_i(t+1) = f_i(x_1(t), \ldots, x_n(t))$. As the interaction between the components may be given in terms of a network (see section III), the components are henceforth called *nodes*.

Let us define what we mean by robustness of a single node i against knockout of another node j. We assume that the input nodes follow a certain dynamics $x_1(t), \ldots, x_n(t)$, where $t \geq 0$. The set of possible activation patterns which the input takes is called the *input support*. It can be represented as a set of binary strings of length n.

Node i is robust against knockout of node j if the state information of node j is either irrelevant for the correct functioning of i or if this information is already contained

in the other available inputs.

Definition: A node i with mapping $f_i : \{0,1\}^n \to \{0,1\}$ is robust against knockout of node j if x_i is independent of x_j given $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots x_n$. With this we mean that

$$f_i(x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots x_n)$$
$$= f_i(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots x_n) \qquad (1)$$

whenever the two states $(x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots x_n)$ and $(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots x_n)$ both lie in the input support. Robustness against simultaneous knockout of multiple nodes is defined in a similar fashion.

It is important to note that this definition depends on the input support. If the input $x_1, \ldots, x_n$ takes each possible value, then robustness against knockout of node j implies that f_i does not depend on x_j. However, if the inputs are highly correlated, then nontrivial robustness may appear: It may happen that x_i can compensate the knockout of a single input by reconstructing the missing information from another input with similar dynamics, but the simultaneous knockout of both inputs leads to a failure. For example, in the sequence of the yeast cell cycle (Table I) it is easy to see that all nodes are robust against knockout of either *MBF* or *SBF*, but in general a simultaneous knockout of both *MBF* and *SBF* cannot be compensated.

Our notion of knockout robustness is based on ideas of Ay and Krakauer [12]. Within the framework defined in [12], resilience means that the *exclusion dependence* of the system with respect to certain knockouts vanishes. Combinatorial conditions characterizing this situation have been found by Herzog et al. [13]. The theory becomes much simpler in the present setting restricted to deterministic dynamics.

Our definition of robustness can be applied to each node of the cell cycle network shown in Figure 2. In our studies of the yeast cell cycle the input support will be the set of activation patterns which appear in this cycle. We then study each node mapping and ask, which inputs are "essential" for the functioning of this node and which inputs can be compensated.

In order to study robustness of the system as a whole there are multiple possibilities: As a measure of *system robustness* we count the number of nodes which are robust with respect to all single node knockouts, *i.e.* how many nodes would remain functional if any one input would be knocked out. Another possibility is to find the knockouts under which the behavior of the whole system is robust. In this case we find the single node knockouts which can be compensated by all (other) nodes of the network. The set of these knockouts is called the *resilience combination.* The cardinality of the resilience combination is called the *knockout resilience.*

B. Null models of state sequence

System robustness and knockout resilience of a specific system may be judged as significantly high or low only in comparison with a null hypothesis from a model. Here we define two null models of state sequences under constraints derived from properties of a cell cycle sequence.

For the *first null model*, the activation states are drawn from $\{0,1\}$ with probabilities $1/2$, independently across time steps and nodes. As an exception, the last activation pattern, is taken as the first activation pattern and then flipping the state of one randomly chosen node. Thus the first and last activation pattern differ at a single node. This constraint is to reflect the fact that the cell cycle is triggered by a single signal protein.

The *second null model* is further constrained as follows.

- *CLN3* acts as an input node. It is active in the first activation pattern and inactive at all later time steps.

- The other nodes are activated and inactivated exactly once during the cycle. Their activity (or inactivity) is therefore constrained to a block of successive time steps.

These constraints are motivated by the observation that switching events (rising and falling edges of activity) are much rarer in the cell cycle than would be expected for uncorrelated random state sequences.

As a natural additional constraint for both models, activation patterns at different steps along the sequence must be different. Sequences violating this condition are discarded.

C. Results

We now apply the ideas of section II to the yeast cell cycle (Table I) and the null models. Starting with the system robustness we find that only the four nodes *CLN3*, *CLN1,2*, *CLB5,6* and *CDC20* are robust against all single node knockouts. For *CLN3* this is trivial since it corresponds to the constant map $f_{CLN3} = 0$. Regarding the knockout resilience, the cell cycle still functions correctly if any one of the five nodes *MBF*, *SBF*, *MCM1/SFF*, *SWI5* or *CDH1* is knocked out.

Summing up these observations, the yeast cell cycle has knockout resilience 5 and system robustness 4. We can now ask whether these values are a special property of the cell cycle or rather expected for a support containing 13 out of $2^{11} = 2048$ possible states. Figure 1 shows that all null models produce a considerable fraction of instances whose resilience and robustness values are above those of the yeast cell cycle sequence.

Formulating this insight in terms of a statistical test, our null hypothesis is that the yeast cell cycle is a random sequence as obtained from the null model considered. The p-value is the probability that the null model

TABLE I: The cell cycle sequence of the yeast species *Saccharomyces cerevisiae* [5]. The last phase *stationary* G_1 is a fixed point of the dynamics, *i.e.* once the system has reached this state, it stays in this state, until some external event flips the first node from 0 to 1 and triggers another cell cycle.

Time	CLN3	MBF	SBF	CLN1,2	CLB5,6	CLB1,2	MCM1	CDC20	SWI5	SIC1	CDH1	Phase
1	**1**	0	0	0	0	0	0	0	0	**1**	**1**	START
2	0	1	1	0	0	0	0	0	0	1	1	G_1
3	0	1	1	1	0	0	0	0	0	1	1	G_1
4	0	1	1	1	0	0	0	0	0	0	0	G_1
5	0	1	1	1	1	0	0	0	0	0	0	S
6	0	1	1	1	1	1	1	0	0	0	0	G_2
7	0	0	0	1	1	1	1	1	0	0	0	M
8	0	0	0	0	0	1	1	1	1	0	0	M
9	0	0	0	0	0	1	1	1	1	1	0	M
10	0	0	0	0	0	0	1	1	1	1	0	M
11	0	0	0	0	0	0	0	1	1	1	1	M
12	0	0	0	0	0	0	0	0	1	1	1	G_1
13	0	0	0	0	0	0	0	0	0	1	1	Stat. G_1

FIG. 1: (Color online) $(a - b)$ The distribution of the number of inputs which can be compensated by all other nodes are shown for the first null model (a) and the more realistic null model (b). The blue (light gray) bars represent the distribution of the random cell cycles and the single green (dark gray) bar represents the knockout resilience of the wildtype network. $(c - d)$ The distribution of the number of nodes which are robust with respect to all single node knockouts for the first null model (c) and the more realistic null model (d). As above, the blue (light gray) bars represent the random cell cycle sequences while the single green (dark gray) bar shows the system robustness for the wildtype network. Note the logarithmic scale of the four diagrams.

generates an instance with an empiricial value greater or equal to the observed value. The first null model yields p-values $P_1^{\mathrm{kr}} = 0.99999$ and $P_1^{\mathrm{sr}} = 0.702$ for knockout resilience and system robustness, respectively. For the second null model we obtain $P_2^{\mathrm{kr}} = 0.808$ and $P_2^{\mathrm{sr}} = 0.226$. It is common practice to reject the hypothesis only for p-values less than 0.05 [14]. Here all p-values are greater than 0.2. Therefore the null hypothesis, stating that the yeast cell cycle sequence is not more optimized for tolerating knockouts than a null model sequence, cannot be rejected.

Another result seen in Figure 1 is that both knockout resilience and system robustness in the second null model are lower than those of the first null model. This is due to the correlations between activation patterns in a sequence generated according to the second null model Fewer state changes lead to more similar patterns along the sequence. Then a knockout is more likely to make activation patterns at different times indistinguishable so the sequence can no longer be performed.

III. ROBUSTNESS OF IMPLEMENTATION

A. Linear threshold networks and neutral graph

The concepts defined up to now apply to the sequence of activation patterns. They do not take into account the interaction network implementing this sequence. In order to investigate the effects of specific network structures we focus on linear threshold networks. Such a network is given as a directed graph on n nodes with weighted edges. We allow loops, *i.e.* edges starting and ending at the same node. For simplicity we only allow weights $w_{ij} = 1$ (*activating* or *positive* edges) and $w_{ij} = -1$ (*inhibiting*
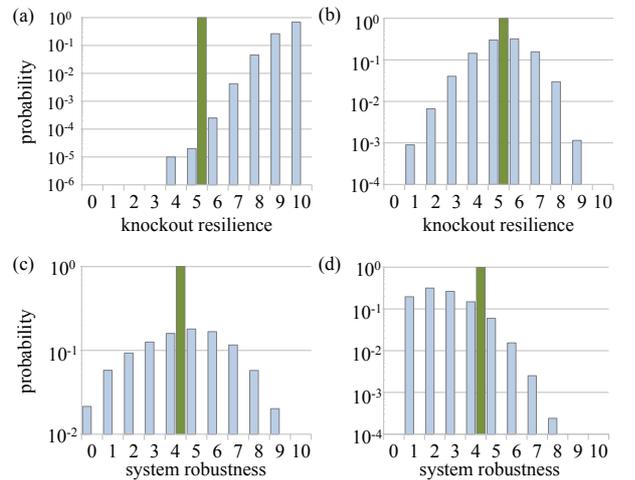
or *negative* edges) for each directed edge $i \leftarrow j$. If there is no edge from node j to node i we write $w_{ij} = 0$. To each node i we associate a time-dependent variable $x_i(t)$ with dynamics given by

$$x_i(t+1) = \begin{cases} 1 & \text{if } \sum_j w_{ij} x_j(t) > 0, \\ 0 & \text{if } \sum_j w_{ij} x_j(t) < 0, \\ x_i(t) & \text{else.} \end{cases} \quad (2)$$

It is possible to refine this model by allowing more values for the edge weights. Furthermore it is customary to assign an activation threshold indiviually to each node i. Then the dynamics is determined by comparing $\sum_j w_{ij} x_j(t)$ to this threshold. In this work we do not make use of these possibilities and restrict ourselves to the simple model.

We focus on networks that perform the sequence of activation patterns of the yeast cell cycle (Table I) when initialized with the START pattern at time 1. One of these *functional* networks called the *wildtype* is shown in Figure 2. The interactions of the wildtype are based on empirical evidence [5]. Note that here the term wildtype does not denote the actual yeast wildtype organism. It is used only to distinguish this network (which is believed to model the actual yeast organism most accurately) from other functional networks on the level of linear threshold models.

As a first null model in the assessment of robustness of an implementation (a particular network) we use a flat distribution on the set of all functional networks, containing 5.11×10^{34} elements [8].

As a second null model we consider a restriction to those networks that may be reached from the wildtype by an evolutionary path. To make this precise, we define the *neutral graph*. Its node set is the set of the functional networks. Two networks A and B are connected by an undirected edge if A can be obtained from B by adding or deleting a single interaction [9, 15]. In our case the neutral graph is disconnected and falls into a large number of connected components [9]. The 5.66×10^{25} networks in the connected component containing the wildype (called wildtype component) serve as another null model, again weighted with a flat distribution.

B. Robustness definition

In the context of linear threshold models we can define a related notion of robustness, which we will call *robustness of implementation*: A knockout of a node j is modeled by removing a node from the network, together with all of the edges involving this node. We can then analyze the dynamics of the changed network and compare it to the dynamics of the unperturbed network. If the dynamics of the remaining $n-1$ nodes is not changed, then we say that the implementation is robust against the knockout of node j. Mathematically this means that the sign of $\sum_{k \neq j} w_{ik} x_k(t)$ equals the sign of $\sum_k w_{ik} x_k(t)$. The *resilience combination* is the set of nodes against whose knockout the network is robust. Now the *knockout resilience* of the network is the cardinality of the resilience combination. This is a property of the implementation and thus depends on the connection structure of the network implementing the cell cycle pattern. Its value cannot be greater than the knockout resilience of the function, as defined above.

C. Results

There are 24 different occurring resilience combinations. Out of these, eight are found in the component containing the wildtype (see Table II).

The majority of networks, *i.e.* $\approx 99\%$ of all the functional networks of the neutral graph and $\approx 90\%$ of the networks in the wildtype component, cannot cope with any single node knockout. However, a maximum of four independent single node knockouts is found. The wildtype itself stays functional after knockout of node *CDH1*, see Figure 3. Networks reachable by a mutational path from the wildtype can manage at most three independent knockouts.

One might speculate that high knockout resilience requires redundant wiring of the network which would be observable as an increased edge density. In Figure 4, we
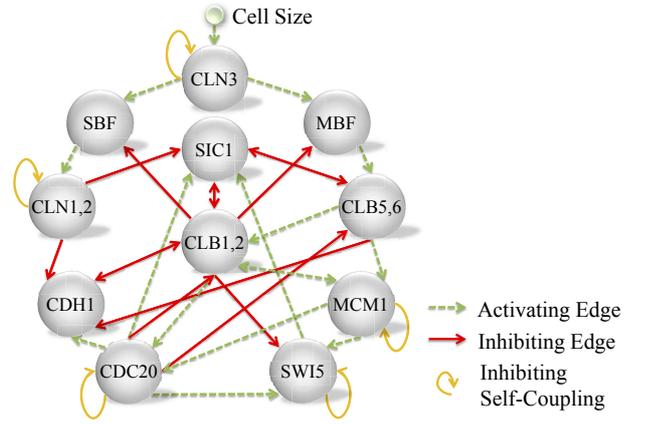

FIG. 2: (Color online) The wildtype network of the cell cycle of the yeast species *Saccharomyces cerevisiae* [5]. The edges of the network are directed and can be activating (dashed green arrow) or inhibiting (solid red arrow). All self-couplings (solid yellow) are inhibiting. The network comprises 34 different interactions, 15 of which are activating and 19 are inhibiting.
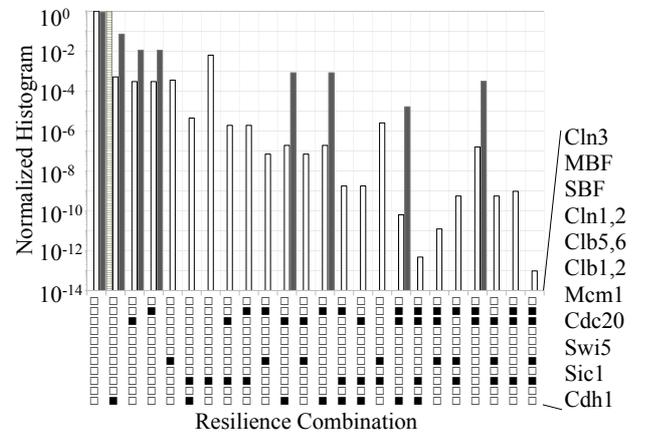

FIG. 3: (Color online) The distribution of the specific knockout schemes is shown. The filled bars represent the wildtype component. The open bars represent all components and the single dashed bar represents the resilience combination of the wildtype itself. Each column represents one of the 24 different resilience combinations which occur among the functional networks. The nodes of the networks are shown as squares (see legend for names). If the knockout of a particular node is functional the square is filled, otherwise it is open. The resilience combinations are ordered such that the average basin size increases.

look at the distribution of the number of positive, negative and all edges for the different resilience combinations. There is no clear correlation between the average number of edges and the knockout resilience.

As suggested by Li et al. [5], the basin of attraction of the G1 fixed point (stationary state) is a measure of robustness. The basin consists of all activation patterns from which the G1 state is eventually reached by following the dynamics (2). The average basin size of networks with a given resilience combination is shown in Figure 5.

TABLE II: The number of networks which allow certain resilience combinations among all networks which implement the cell cycle sequence (column *all components*). Among these networks the number of networks in the wildtype components is shown in the column *wildtype component*.

Number	Wildtype component	All components	CLN3	MBF	SBF	CLN1,2	CLB5,6	CLB1,2	MCM1	CDC20	SWI5	SIC1	CDH1	# Knockouts
1	5.13×10^{25}	5.07×10^{34}	□	□	□	□	□	□	□	□	□	□	□	0
2	3.98×10^{24}	2.59×10^{31}	□	□	□	□	□	□	□	□	□	□	■	1
3	6.22×10^{23}	1.50×10^{31}	□	□	■	□	□	□	□	□	□	□	□	1
4	6.22×10^{23}	1.50×10^{31}	□	■	□	□	□	□	□	□	□	□	□	1
5		1.77×10^{31}	□	□	□	□	□	□	■	□	□	□	□	1
6		2.24×10^{29}	□	□	□	□	□	□	□	□	■	□	■	2
7		3.15×10^{32}	□	□	□	□	□	□	□	□	■	□	□	1
8		9.67×10^{28}	□	□	■	□	□	□	□	□	■	□	□	2
9		9.67×10^{28}	□	■	□	□	□	□	□	□	■	□	□	2
10		3.52×10^{27}	□	■	□	□	□	□	■	□	□	□	□	2
11	4.63×10^{22}	9.62×10^{27}	□	□	■	□	□	□	□	□	□	□	■	2
12		3.51×10^{27}	□	□	■	□	□	□	■	□	□	□	□	2
13	4.63×10^{22}	9.62×10^{27}	□	■	□	□	□	□	□	□	□	□	■	2
14		8.87×10^{25}	□	■	□	□	□	□	□	□	■	□	■	3
15		8.87×10^{25}	□	□	■	□	□	□	□	□	■	□	■	3
16		1.28×10^{29}	□	□	□	□	□	□	■	□	■	□	□	2
17	9.12×10^{20}	3.22×10^{24}	□	■	■	□	□	□	□	□	□	□	■	3
18		2.44×10^{22}	□	■	■	□	□	□	□	□	■	□	■	4
19		6.25×10^{23}	□	■	■	□	□	□	■	□	□	□	□	3
20		2.79×10^{25}	□	■	□	□	□	□	■	□	■	□	□	3
21	1.74×10^{22}	8.04×10^{27}	□	■	■	□	□	□	□	□	□	□	□	2
22		2.79×10^{25}	□	□	■	□	□	□	■	□	■	□	□	3
23		4.82×10^{25}	□	■	■	□	□	□	□	□	■	□	□	3
24		4.95×10^{21}	□	■	■	□	□	□	■	□	■	□	□	4
Sum	$\mathbf{5.66 \times 10^{25}}$	$\mathbf{5.11 \times 10^{34}}$												

TABLE III: The average basin sizes for the different knockout resiliences. Again the wildtype component is considered separately.

Knockout Resilience	Wildtype Component	All Components
0	1663.44	1447.42
1	1629.86	1485.06
2	1675.99	1513.21
3	1779.22	1588.93
4		1594.82

In Table III the average basin size for the different knockout resiliences is shown. With an increasing capability to cope with more single node knockouts the average basin size increases. Additionally, average basins sizes in the wildtype component are larger than their corresponding basin sizes in all components. However, all network implementations in the wildtype component have a network resilience of at most three.

IV. DISCUSSION

We have studied network dynamics with Boolean state vectors (activation patterns) under knockout of single nodes using the example of the yeast cell cycle. The yeast wildtype network is not optimized for knockout resilience, given the sequence of activation patterns. There are networks with significantly larger knockout resilience implementing the same sequence.

The model of the yeast cell cycle studied in this work is far from being a complete description of cell function. For example, a more realistic model could include further nodes as well as graded activations (*i.e.* the concentrations of the molecules at the nodes are considered as real valued activations). An increasing number of nodes would generically increase the robustness of the null models since additional state information is available to compensate node knockouts. Thereby, our conclusion that the yeast cell cycle is not especially robust would not change under this generalization. The analysis of graded concentrations instead of binary activations would require a notion of state distance taking into account which disturbances are considered and how they affect the functioning of the cell. This is beyond reach of the present idealized model.
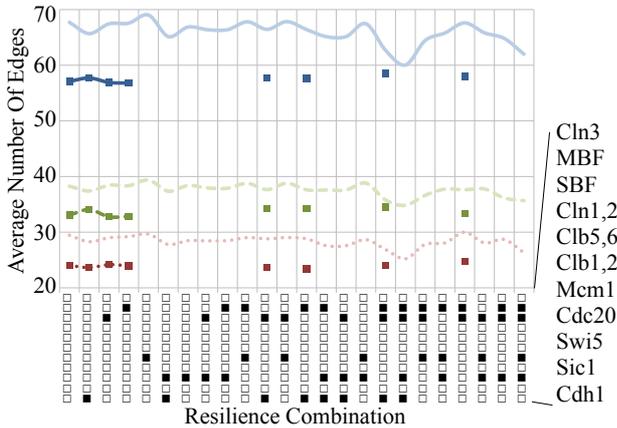
FIG. 4: (Color online) The distribution of positive (dashed lines), negative (dotted lines) and all edges (straight lines) for each of the corresponding resilience combinations are shown. The lighter lines (light gray) represent networks from all components, while the darker lines (dark gray) with squares represent networks only reachable by mutations from the wildtype. The resilience combinations are depicted as in figure 3.
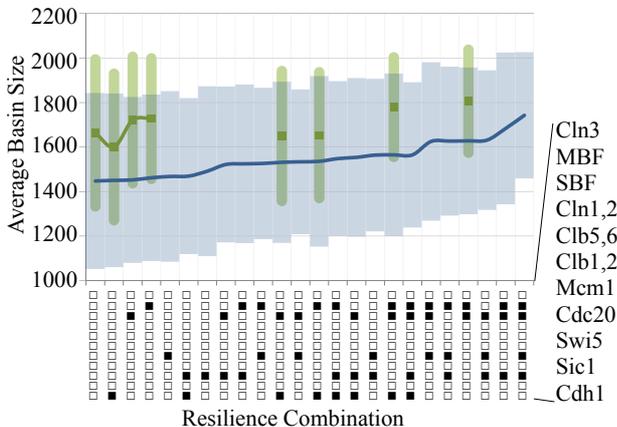

FIG. 5: (Color online) The averages of the basin sizes corresponding to the specific resilience combination are shown. The straight line (dark gray) represents networks from all components, while the line with squares (light gray) represents networks only from the wildtype component. Additionally, the area between the 10% and 90% quantiles is inked. The resilience combinations are depicted as in figure 3. For each resilience combination 10^5 networks were sampled.

Finally we stress that our definition of knockout resilience checks whether a node is dispensable for integrating and transmitting information only in the context of the regulatory network we consider. In reality, the considered node may be involved in other functions indispensable for survival. This would further reduce the knockout resilience.

Acknowledgments

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science (New York, N.Y.) **297**, 1183 (2002), ISSN 1095-9203.

[2] H. Kitano, Nature reviews. Genetics **5**, 826 (2004), ISSN 1471-0056.

[3] A. Wagner, *Robustness and Evolvability in Living Systems* (Princeton University Press, 2005).

[4] S. Kauffman, Journal of Theoretical Biology **22**, 437 (1969).

[5] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, Proceed-

ings of the National Academy of Sciences of the United States of America **101**, 4781 (2004), ISSN 0027-8424.

[6] S. Braunewell and S. Bornholdt, Journal of Theoretical Biology **245**, 638 (2007).

[7] W.-B. Lee and J.-Y. Huang, FEBS letters **583**, 927 (2009), ISSN 1873-3468.

[8] K.-Y. Lau, S. Ganguli, and C. Tang, Physical Review E **75**, 051907 (2007), ISSN 1539-3755.

[9] G. Boldhaus and K. Klemm, European Physical Journal B pp. online first, doi=10.1140/epjb/e2010–00176–4

(2010).

[10] M. I. Davidich and S. Bornholdt, PLoS ONE **3**, e1672 (2008).

[11] I. Albert, J. Thakar, S. Li, R. Zhang, and R. Albert, Source Code for Biology and Medicine **3**, 16 (2008), ISSN 1751-0473.

[12] N. Ay and D. C. Krakauer, Theory in Biosciences **125**, 93 (2007), ISSN 1431-7613.

[13] J. Herzog, T. Hibi, F. Hreinsdóttir, T. Kahle, and J. Rauh, Advances in Applied Mathematics pp. – (2010), ISSN 0196-8858, in Press, Corrected Proof.

[14] R. A. Fisher, *Statistical Methods, Experimental Design, and Scientific Inference* (Oxford University Press, 1990).

[15] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker, Proc. Roy. Soc. Lond. B **255**, 279 (1994).