EPJ manuscript No.
(will be inserted by the editor)

Regulatory networks and connected components of the neutral space

A look at functional islands

Gunnar Boldhaus[1] and Konstantin Klemm[1a]

Bioinformatics Group, Dept. of Computer Science, University of Leipzig, Germany

Received: date / Revised version: date

Abstract. The functioning of a living cell is largely determined by the structure of its regulatory network, comprising non-linear interactions between regulatory genes. An important factor for the stability and evolvability of such regulatory systems is neutrality — typically a large number of alternative network structures give rise to the necessary dynamics. Here we study the discretized regulatory dynamics of the yeast cell cycle [Li et al., PNAS, 2004] and the set of networks capable of reproducing it, which we call *functional*. Among these, the empirical yeast wildtype network is close to optimal with respect to sparse wiring. Under point mutations, which establish or delete single interactions, the neutral space of functional networks is *fragmented* into $\approx 4.7 \times 10^8$ components. One of the smaller ones contains the wildtype network. On average, functional networks reachable from the wildtype by mutations are sparser, have higher noise resilience and fewer fixed point attractors as compared with networks outside of this wildtype component.

PACS. 87.16.Yc Regulatory genetic and chemical networks – 87.10.-e General theory and mathematical aspects – 87.17.Aa Theory and modeling; computer simulation

1 Introduction

Neutrality [1] is crucial for robustness and evolvability [2] of biological systems. It describes the fact that the mapping from genotypes to phenotypes is not invertible. A given phenotype can be encoded by more than one genotype. As Wagner [2] writes, *"most problems the living have solved have an astronomical number of equivalent solutions, which can be thought of as existing in a vast neutral space"*.

Computational studies of biopolymers revealed the existence of neutrality in the relation between sequence and spatial structure. RNA molecules and proteins are generated as a chain (sequence) of nucleic bases and amino acids respectively. The number of sequences folding into one and the same functionally relevant spatial structure is found to be large. It is growing exponentially with the size of the molecule [3, 4]. Together with an adjacency given by single mutations, the phenotypically equivalent genotypes form the neutral network (or *neutral graph*). The properties of this graph, in particular its *connectivity*, determine the robustness of the given genotype under mutations and its evolvability towards new phenotypes.

Going from single molecules to the level of the whole organism, the phenotype is not given by the set of its molecule structures alone: The dynamics that arises as the result of activating and suppressing *interactions* between molecules is crucial. This set of interactions is captured as a regulatory network [5] and gives rise to a temporal sequence of chemical concentration vectors that are responsible for the division of a single cell or the development of an embryo. Again, the mapping from genotypes (interaction networks) to phenotypes (temporal sequences) is not injective, i.e. several network structures are able to produce the regulatory dynamics of a given phenotype [6, 7, 8, 9]. Here we apply the neutral graph concept to a dynamical model [10] of cell cycle regulation in the organism of the yeast species *Saccharomyces cerevisiae* (budding yeast). In section 2 we introduce the model dynamics and the wiring of the wildtype network. The ensemble of functional networks that yield dynamics equivalent to the wildtype is analyzed in section 3, finding the neutral graph to be disconnected. In section 4 we focus on statistical properties of the subset of networks that are reachable from the wildtype. After a remark (section 5) on the computation of network statistics, section 6 offers a discussion and open questions.

2 Cell cycle network and Boolean dynamics

During the process of cell division, a eukaryotic cell grows and divides into two daughter cells. A cell cycle consists

[a] E-mail: boldhaus@bioinf.uni-leipzig.de, klemm@bioinf.uni-leipzig.de

of four distinct and separate phases named G_1, S, G_2 and M. In the G_1 ("growth") phase, the cell commits itself for cell division under certain conditions. In particular, a necessary cell size must have been reached. A copy of the genetic information is produced in the S ("synthesis") phase. The G_2 ("gap") phase precedes the actual cell division in the M phase ("mitosis").

Here we are interested in the network of molecules (cyclins, inhibitors and degraders of cyclins and transcription factors) regulating this process. We consider the regulatory network of the mono cellular eukaryotic organism *Saccharomyces cerevisiae* (budding yeast). Its genome comprises 13 million base pairs and 6275 genes, of which approximately 800 are involved in the cell cycle dynamics [11]. The dynamics is controlled by a core of 11 key regulators with 34 directed interactions [10], shown in Figure 1(a), which we denote as the *wildtype* network. Interactions are captured by a matrix A. If node j has an activating effect on node i, the corresponding matrix element is $a_{ij} = +1$, while inhibition is coded as $a_{ij} = -1$. In case of no direct influence from j to i, we have $a_{ij} = 0$. Li et al. [10] model the regulatory dynamics with a Boolean approach [12, 13] where each node i takes state values $S_i(t) \in \{0, 1\}$ when being inactive / active at time t. In the time-discrete dynamics, nodes are updated synchronously, based on their weighted input sum $h_i(t) = \sum_j a_{ij} S_j(t)$. The state at the next time step is obtained by applying the threshold update rule

$$S_i(t+1) = \begin{cases} 1, & h_i(t) > 0 \\ 0, & h_i(t) < 0 \\ S_i(t), & h_i(t) = 0 \end{cases} . \qquad (1)$$

From an initial condition $S(1)$, representing the real starting state of the cell cycle, the dynamics produces the sequence of state vectors $S(1), S(2), \ldots, S(13)$, shown in Figure 1(b). The state $S(13) = G_1$ is a fixed point of the dynamics. The system remains in this state until node Cln3 is externally activated. In the real system the external activation indicates that the cell size is sufficient for another division.

3 Functional networks and the neutral graph

Broadening our treatment of regulatory networks, we consider the set of all networks with interaction matrices over 11 nodes with entries $a_{ij} \in \{-1, 0, +1\}$. We call a network *functional* if it produces the state transitions of the cell cycle sequence in Figure 1(b). Thus, the wildtype network is functional. However, there are further functional networks. Out of the set of all $3^{2^{11}} \approx 5.4 \times 10^{57}$ networks, approximately 5.11×10^{34} are functional [14]. Figure 1(c) shows an example of a functional network different from the wildtype. Figure 2 shows the statistics for the number of interactions (arcs) present in functional networks. The wildtype network is sparse in comparison with the average functional network. However, there are functional networks that are even sparser than the wildtype. These findings analogously hold when activating and inhibiting interactions are counted separately. Interestingly, functional networks have generally more suppressing than activating interactions, as is the case for the wildtype.

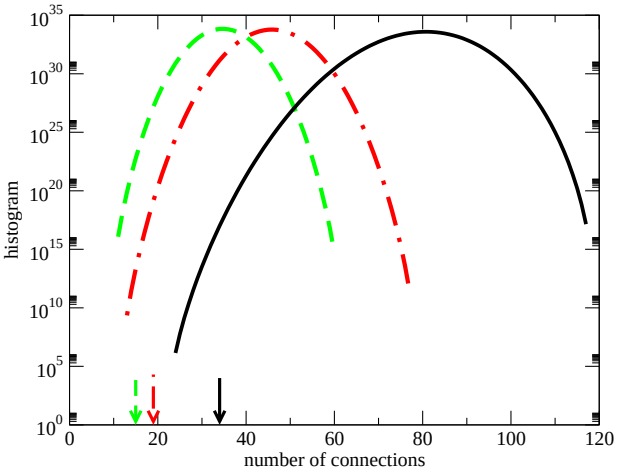


Fig. 2. Histograms of the number of interactions over functional networks. Positive (green dashed curve), negative (red dot-dashed), and total connections (black solid) of almost all functional networks exceed the corresponding counts in the wildtype network (arrows).

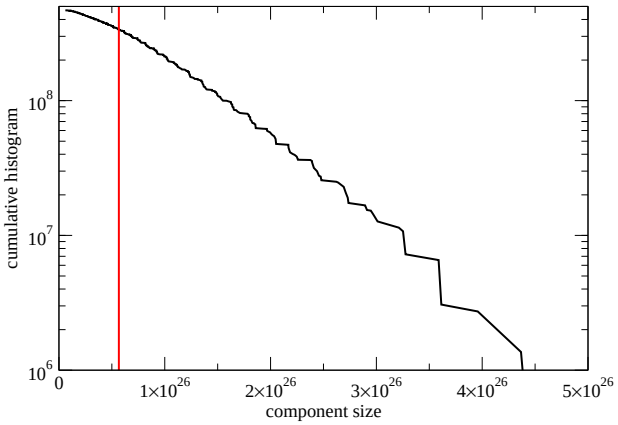
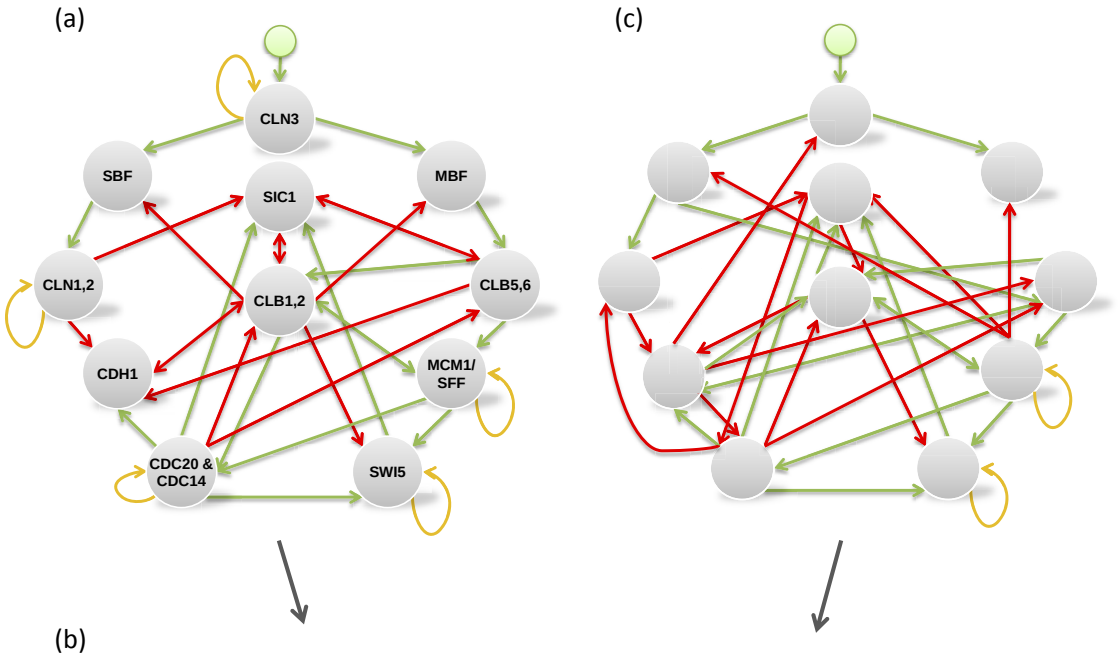


Fig. 3. Cumulative size distribution of connected components of the neutral graph (falling curve). The component containing the wildtype has size 5.66×10^{25} (vertical line).

A structure to reflect mutations on the set of functional networks is the *neutral graph*. Its nodes are the functional networks. Functional networks A and B are adjacent (connected by an edge) in the neutral graph if A is turned into B by a single *mutation*. According to our definition a mutation is a replacement of one entry in the interaction matrix. The Hamming distance between two networks is the number of entries in which their interaction matrices differ. In order to avoid confusion with the networks of interaction we employ the term *neutral graph* as a synonym for the more commonly used *neutral network*. An important property of a neutral graph is its connectedness. A mutational walk from network A to network B is a sequence of single point mutations that turns A into



Time	Cln3	MBF	SBF	Cln1,2	Clb5,6	Clb1,2	Mcm1/SFF	Cdc20&Cdc14	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	0	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	Stationary G_1

Fig. 1. (a) The Cell Cycle Network of the yeast wildtype has 11 nodes connected with activating (green) and inhibiting (red) interactions. Self-suppression is indicated by yellow loops. (b) A sequence of 13 states defines a cell cycle, as produced by the network in (a). (c) A different network (mutant) performs the same sequence of states. As the wildtype, this mutant has 34 interactions. However, 19 entries in the interaction matrix differ from the wildtype.

B without passing through non-functional networks. The neutral graph is connected if such a mutational walk exists for each pair of functional networks. We find that the neutral graph considered here is *disconnected*. One cannot pass from all functional networks to all others by sequences of mutations that preserve functionality. In fact, mutual reachability between functional networks is rare. The neutral graph falls into $\approx 4.7 \times 10^8$ connected components with sizes distributed between $\approx 6.1 \times 10^{24}$ and $\approx 4.4 \times 10^{26}$, as shown in Figure 2. The component of the wildtype comprises around 5.66×10^{25} functional networks.

4 The wildtype component

In this section we extend the analysis of the neutral graph. We focus on a comparison between functional networks in the wildtype component and all functional networks. Figure 4(a-c) shows how the number of (a) negative, (b) positive and (c) all interactions is distributed. All three plots reveal a significant statistical difference between networks in the wildtype component and the set of all functional networks. Networks in the wildtype component are sparse compared with the average functional network.

Geometric information of the neutral graph is provided in Figure 4(d) in terms of the Hamming distance of functional networks from the wildtype. Functional networks in the wildtype component are closer to the wildtype than the average functional network is. Still the most remote

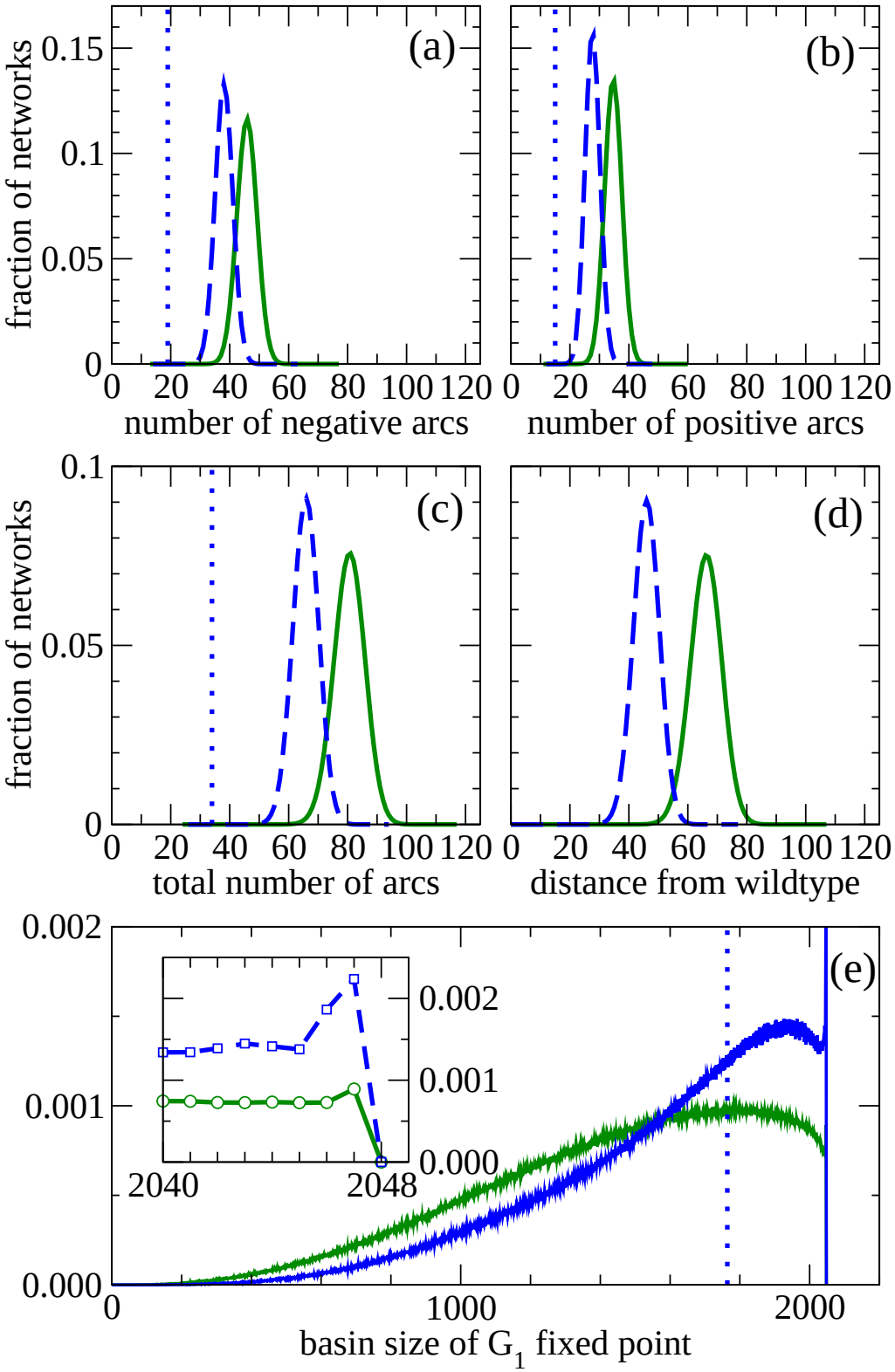


Fig. 4. Comparison of statistics between all functional networks (solid green curves) and functional networks in the wildtype component (dashed blue curves) of the neutral graph. (a) histogram of negative interactions, (b) histogram of positive interactions and (c) histogram of total number of interactions in functional networks. (d) histogram of Hamming distances (minimal number of mutations) from the wildtype. (e) Distribution of basin sizes of the G_1 fixed point. The inset shows a zoom into the histogram for very large basin sizes. For comparison, a vertical dotted line in each panel gives the value of the wildtype network itself. Histograms in panels (a)-(d) are exact. Histograms in (e) were obtained by uniform sampling of 10^7 functional networks each from the whole neutral graph and from its wildtype component, respectively.

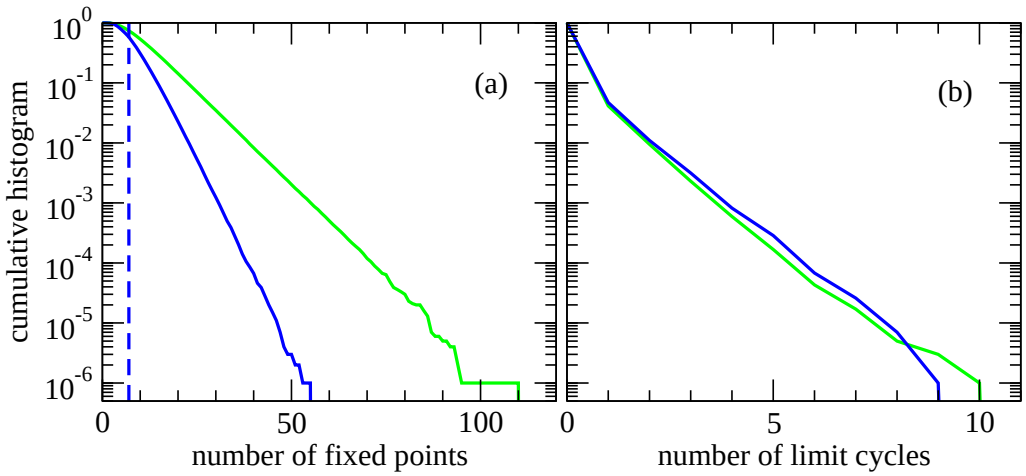


Fig. 5. The number of attractors of all functional networks (green curves) and the functional networks in the neutral graph component containing the wildtype (blue curves). (a) Distribution of the number of fixed points. (b) Distribution of the number of limit cycles (attractors of length at least 2). The wildtype itself has 7 fixed points (vertical dashed line) and no limit cycles.

networks in the wildtype component are found at distance 77 from the wildtype. Despite its moderate size, the wildtype component pervades a large part of the network space.

Shifting attention from the structural to the dynamical properties of the functional networks, let us analyze the resilience of the dynamics against perturbations. As a measure of resilience we use the G_1 basin size [10], i.e. the number of states from which the dynamics eventually reaches the fixed point G_1. Clearly, the basin contains at least the 13 states in the cell cycle sequence. As shown by the distributions in Figure 4(e), actual G_1 basin sizes in functional networks contain many more states. Compared with all functional networks, basin sizes of networks in the wildtype component concentrate at higher values. The most frequently observed basin size is 2047 for networks in the wildtype component, cf. the inset of Figure 4(e). However, we have not found a functional network where the G1 basin contained all 2048 states.

Moreover, the distributions in the number of fixed points of functional networks show a striking difference between the wildtype component and the whole neutral graph. Figure 5(a) displays geometric distributions in both cases. However, networks in the wildtype component have a significantly narrower distribution of fixed points. Interestingly, dynamic attractors (limit cycles) with more than one state show practically the same statistics in the wildtype component as in the whole neutral graph, cf. Figure 5(b).

5 Computational aspects

Computation of histograms over functional networks is facilitated by the fact that rows of functional network matrices combine independently. The set of all functional matrices M is a product over sets M_i of functional row vectors for each node i. In fact, each row of the matrix has its own neutral graph. The Cartesian product [15] of these is the neutral graph of the whole system. As noted by Lau et al. [14], the set of network matrices performing a given state sequence has a simple combinatorial structure. One can check independently for each node i if it takes the required state at each time step t. The states taken by node i only depend on the i-th row and not on the whole matrix. Thus, a functional network can be constructed by independently combined functional row vectors into a matrix. The set M_i of functional row vectors is obtained by testing each of the $3^{11} \approx 2 \times 10^5$ possible vectors over $\{-1, 0, 1\}$ for each node i.

For calculating the histogram of the number of interactions over all functional networks (Figure 4(c)), we first calculate this histogram g_i individually for each node i. The number of functional vectors in row i with exactly x interactions is

$$g_i(x) = |\{r \in M_i | k(r) = x\}| \tag{2}$$

where $k(r)$ is the number of non-zero entries in a row vector $r \in M_i$.

These N histograms are now iteratively combined into the histogram h_i over the first i rows according to

$$h_i(z) = \sum_{x=0}^{z} g_i(x) h_{i+1}(z - x) \tag{3}$$

with initialization $h_1 = g_1$. Then h_N is the histogram over the complete matrices having N rows. Histograms are obtained anlogously for other observables like the number of positive and negative interactions and the Hamming distance. In fact, the iterative combination of histograms is applicable for all matrix observables that are decomposable into observables of single rows.

The number of attractors and the size of basins (Figure 5) do not fall into this class of row-decomposable observables. Thus sampling is used to obtain statistics of these observables. For drawing a matrix A from the uniform distribution on M, we draw the i-th row vector of A from the

uniform distribution on M_i for all $i \in \{1, \ldots, N\}$. Before starting the sampling, the whole sets M_i need to be computed and stored once. In the present case, these sets are sufficiently small with less than 2×10^4 elements for all i.

For each functional matrix A in the sample, basin sizes and attractor numbers are obtained by complete enumeration of the Boolean state space having $2^N = 2048$ elements.

6 Discussion & Outlook

We have analyzed the *neutral graph* (also called neutral network) of discrete regulatory networks reproducing the cell cycle sequence of budding yeast [10]. The neutral graph falls into many connected components. Networks in different components of the neutral graph are not accessible to each other through a sequence of mutations that retains cell cycle functionality. Our finding contrasts with the connected neutral graphs in the work by Ciliberti et al. in a similar type of discrete regulatory networks [7, 8]. There, function is defined as the eventual arrival at a predefined fixed point from a given initial condition. In the present study, the exact sequence of states leading to the fixed point is part of the required phenotype. We hypothesize that the fragmentation of the neutral graph is caused by increasing functional constraints.

Further analysis has revealed that functional networks accessible from the empirical wildtype are structurally and dynamically distinct from other functional networks. Networks in the wildtype component are more sparsely wired and their dynamics is more resilient to perturbations, as compared to the average of all functional networks.

Thus, networks in the wildtype component have properties similar to the wildtype itself. This is remarkable since most networks in the wildtype component are distant from the wildtype, having only a few interactions in common.

Future investigations could establish conditions for the connectedness of the neutral graph. To what extent is the fragmentation of the neutral graph caused by the strong discretization of interaction strengths? Allowing finer adaptations would lead to less fragmented neutral graphs. In the extreme (though chemically unrealistic) limit of continuously evolving interaction strengths, the set of all functional network matrices is convex and thus connected.

Acknowledgments. We thank Anke Busch, Nadine Menzel, and Markus Riester for valuable comments on the manuscript. This work has been funded by the VolkswagenStiftung.

References

1. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK, 1983)
2. A. Wagner, *Robustness and Evolvability in Living Systems* (Princeton University Press, 2005)
3. P. Schuster, W. Fontana, P.F. Stadler, I.L. Hofacker, Proc. Roy. Soc. Lond. B **255**, 279 (1994)
4. A. Babajide, I.L. Hofacker, M.J. Sippl, P.F. Stadler, Fold Des **2**(5), 261 (1997)
5. E. Davidson, M. Levin, Proceedings of the National Academy of Sciences of the United States of America **102**(14), 4935 (2005)
6. S. Bornholdt, K. Sneppen, Phys. Rev. Lett. **81**(1), 236 (1998)
7. S. Ciliberti, O.C. Martin, A. Wagner, Proceedings of the National Academy of Sciences of the United States of America **104**(34), 13591 (2007)
8. S. Ciliberti, O.C. Martin, A. Wagner, PLoS Computational Biology **3**(2), e15 (2007)
9. F. Stauffer, J. Berg, EPL (Europhysics Letters) **88**(4), 48004 (2009)
10. F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, Proceedings of the National Academy of Sciences of the United States of America **101**(14), 4781 (2004)
11. P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, B. Futcher, Mol Biol Cell **9**(12), 3273 (1998)
12. S. Kauffman, Journal of Theoretical Biology **22**(3), 437 (1969)
13. B. Drossel, *Reviews of Nonlinear Dynamics and Complexity, Volume 1* (Wiley-VCH, 2008), chap. Random Boolean Networks, pp. 69–99
14. K.Y. Lau, S. Ganguli, C. Tang, Physical Review E **75**, 051907 (2007)
15. W. Imrich, *Product Graphs: Structure And Recognition* (Wiley Interscience Series in Discrete Mathematics, 2000)