Organizational Structure of the Transcriptional Regulatory Network of Yeast: Periodic Genes

Frank Emmert-Streib[1] and Matthias Dehmer[2]

[1] Queen's University Belfast, Computational Biology and Machine Learning, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, 97 Lisburn Road, Belfast BT9 7BL, UK
v@bio-complexity.com
[2] Institute for Bioinformatics and Translational Research, UMIT, Eduard Wallnoefer Zentrum 1, 6060, Hall in Tyrol, Austria
Matthias.Dehmer@umit.at

Abstract. In this paper we investigate the organizational structure of the transcriptional regulatory network of *S. cerevisiae* with respect to the connectivity structure of periodic genes. We demonstrate that the giant strongly connected component plays a prominent role serving as central connector for genes experimentally found to be periodically expressed during the cell cycle of yeast. Numerically, we find by randomization of the gene labels that this organizational structure is unlikely to be formed by chance.

Keywords: graph theory, transcriptional regulatory network, causality, randomization, periodic genes.

1 Introduction

The analysis of complex network has gain much attention during the last decade [2,5,14,20,19]. This interest comes in part from the fact that many natural phenomena can be cast into a network framework that enables an analysis of the problem. Especially, in molecular biology such approaches have been used frequently [8,15,18]. In contrast to the theoretical analysis of general complex networks and their properties in biology that major interest consists in understanding the functional organization of gene networks [4]. So far, however, it is largely unknown how to connect, e.g., graph theoretical network properties meaningfully to the biological function of a molecular biological system.

In this paper we use the transcriptional regulatory network of yeast to investigate the organizational structure of periodic genes. Genes are called periodic if they are expressed periodically during the cell cycle [1,12,21,23] that means if they are not just switched on or off but alternate periodically between activation states. Traditional approaches studying periodic genes use, e.g., time series data from DNA microarray experiments trying to identify periodic pattern. Here we do not aim to identify periodic genes but are interested instead in their structural organization in the transcriptional regulatory network. That

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 140–148, 2009.

means, we use the transcriptional regulatory network [13,22], which is a directed, unweighted network, and a list of genes known to be periodic [23] to perform a structural analysis of this network. Our analysis is based on the observation that a (general) network may contain one or more strongly connected components. A strongly connected component is a subnetwork connecting each pair of nodes in this subnetwork bidirectionally. With other words, the strongly connected component has a cyclic structure allowing to connect nodes on closed paths (cycles). The transcriptional regulatory network of yeast contains such strongly connected components [9]. Due to the fact, that only nodes in the strongly connected component can occur on cycles we hypothesis that only these genes can be directly activated periodically [9]. All other gene that are periodic need to be triggered by these genes. For this reason we hypothesis that the strongly connected component plays a prominent role in the organization of the cell cycle and the activation of periodic genes. We calculate the shortest paths [6] from the strongly connected component to all periodic genes in the network (if possible) and investigate the observed structure. More strictly, due to the fact that the transcriptional regulatory network is a curated network its structure can be considered as *causal* representing molecular interactions instead of just some form of association between the genes. This implies that our graph theoretical approach is causality based because a path connects only genes that (potentially) influence each other causally.

This paper is organized as follows. In the next two sections we present the method we apply to the transcriptional regulatory network and the data we use for our analysis. In section 4 we present numerical results and this paper finishes in section 5 with conclusions.

2 Methods

We use a graph theoretical approach to study the structural organization of periodic genes in the transcriptional regulatory network G of yeast. In addition to the transcriptional regulatory network we use a list of genes known to be periodically expressed during the cell cycle.

The transcriptional regulatory network can be partitioned by the presence or absence of cycles connecting genes. In mathematical terms a part of the network that is cyclic is also called a strongly connected component (SCC) [7]. For example, for a SCC containing at least three genes, A_i, A_j, A_k there exists a cycle $A_i \to \ldots \to A_j \to \ldots \to A_k \to \cdots \to A_i$. The dots indicate that there are possibly other genes involved. However, the important point is that there exists a cycle on which all three genes appear. This observation is important because the presence of a cycle in a network is a necessary condition that truly periodic behavior can be observed because these genes have the ability to interact (activate/inhibit) each other consecutively and, hence, can form a limit cycle [17]. This leads us to the separation of the genes in two classes. The first class consists of genes that belong to the SCC. The genes in the second class do not belong to the SCC. Further the two classes are not equal but the information should

flow in one direction namely from $SCC \rightarrow G/SCC$. The reason is that only genes in the SCC can establish a periodic behavior, as explained above, while genes in G/SCC can not. Based on this classification and hierarchy we raise the following hypothesis [9].

Hypothesis 1. *Given a causal path from a gene in the SCC to a gene in G/SCC, obtained from the transcriptional regulatory network, connecting two genes known to be periodic than all genes on this path are periodic if: First, the connecting path is a shortest path. Second, there is just one shortest path connecting the periodic genes.*

In this paper we will not analyze the predictions of our hypothesis but instead we focus on the structural organization of the obtained subnetwork. More precisely, we will analyze if properties of the observed subnetwork are formed by chance or significant with respect to gene label randomization.

3 Data

For our analysis we use the transcriptional regulatory network (TRN) of yeast [13,22] which is a directed, unweighted network. From this network we extract the weakly connected component (WCC) consisting of 3357 genes and 7230 interactions. The weakly connected component of a network is defined as the subnetwork that connects every pair of nodes by at least one directed path [7]. In contrast, the strongly connected component (SCC) is defined as subnetwork that connects each pair of genes in both directions. That means there exists a path connecting, e.g., gene A with gene B but there exists also a path connecting gene B with gene A. The TRN consists of two strongly connected components. One consists of 36 and the other of just 2 genes. When we speak in the following of the SCC of the TRN we speak always about the larger subnetwork also called the giant strongly connected component [24]. The strongly connected component is part of the weakly connected component, $SCC \subseteq WCC$. We use a list of Zhao et al. as reference for periodic genes [23]. They categorized 260 genes as periodic from which 179 are in the subnetwork (WCC) considered in our analysis.

4 Results

In Fig. 1 we show a subnetwork of the transcriptional regulatory network. This subnetwork contains the strongly connected component and all periodic genes that can be reached from there. The SCC is shown as one node only (red) because we are here not interested in the connectivity of the SCC but the connectivity from the SCC to periodic genes. The periodic genes are shown in orange and genes that are non periodic are shown in blue. We want to emphasize that we included only edges that occur on shortest paths from the SCC to periodic genes. This does not only simplify the situation but corresponds also to an assumption frequently employed in the context of gene networks in general [3,11,16] assuming that interactions follow shortest paths. The resulting network looks

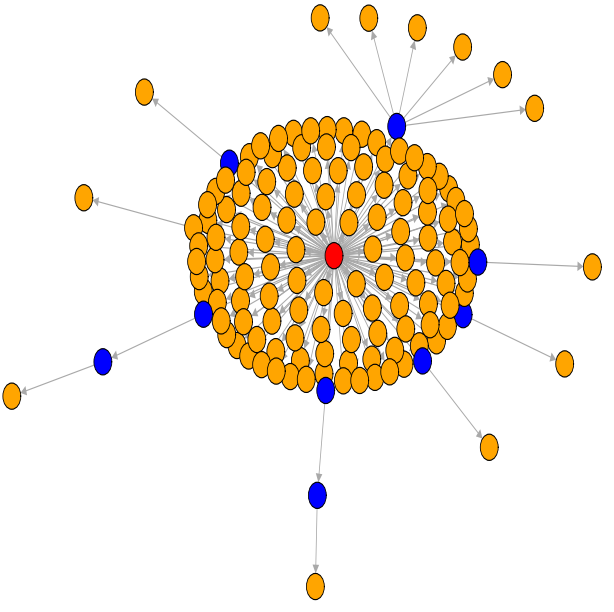

Fig. 1. Subnetwork of the TRN of yeast. Shown are 141 genes and the strongly connected component represented as one red node. Nodes in orange correspond to periodic genes [23], blue nodes are genes not categorized as periodic. The connections shown are shortest paths connecting the periodic genes to the strongly connected component. All other connections are omitted.

remarkably simple containing many periodic genes and only very few non periodic genes. More precisely, we find among all 179 periodic genes in the WCC 132 are connected to the SCC. This corresponds to 73% of all periodic genes. Considering the fact that the SCC contains nine more periodic genes our model view covers 78% of all periodic genes. Another interesting result from Fig. 1 is that only 9 non periodic genes (blue nodes) are necessary to accomplish the shown connected subnetwork.

The crucial question arsing from these observations is if these results are an effect caused by evolution or if these results are merely random structures. To investigate this we randomize the transcriptional regulatory network in the following way. We keep all genes from the SCC fixed. All other node labels, which correspond to gene names, are randomized by permuting these node labels. According to this randomization we generate an ensemble of $N_E = 1000$ networks and repeat our analysis for Fig. 1. The results of these randomizations are shown in Figs. 2–4.

In Fig. 2 we show histograms of the number of periodic genes N^p_{sc} that are directly connected to the SCC (top) and of the number of periodic genes N^p_{tot} that are reachable via a shortest path from the SCC (bottom). It is interesting to see that in these randomized networks the number of periodic genes is much smaller compared to the results for the (normal) TRN shown in Fig. 1.

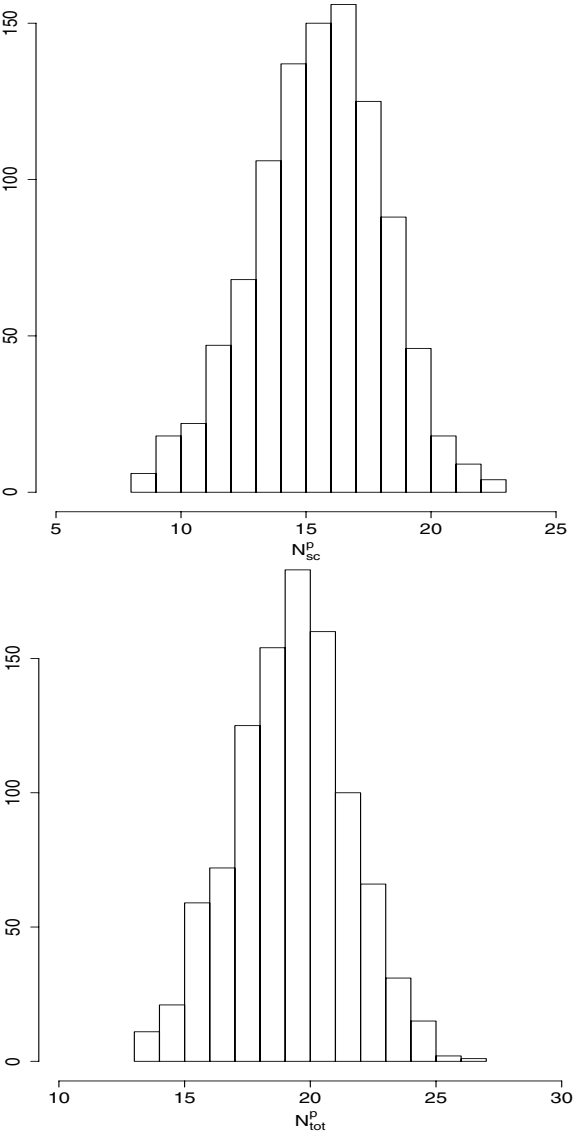

Fig. 2. Top: Histogram of the number of periodic genes N^p_{sc} that are directly connected to the SCC as a result of gene label randomization. The mean value of N^p_{sc} is 16.02. Bottom: Histogram of the number of periodic genes N^p_{tot} that are reachable from the SCC as a result of gene label randomization. The mean value of N^p_{tot} is 19.75.

In Fig. 3 we show the percentage of non-periodic genes involved to connect the SCC to N^p_{tot} periodic genes,

$$p^p_{np} = \frac{N_{np}}{N^p_{tot}}. \tag{1}$$

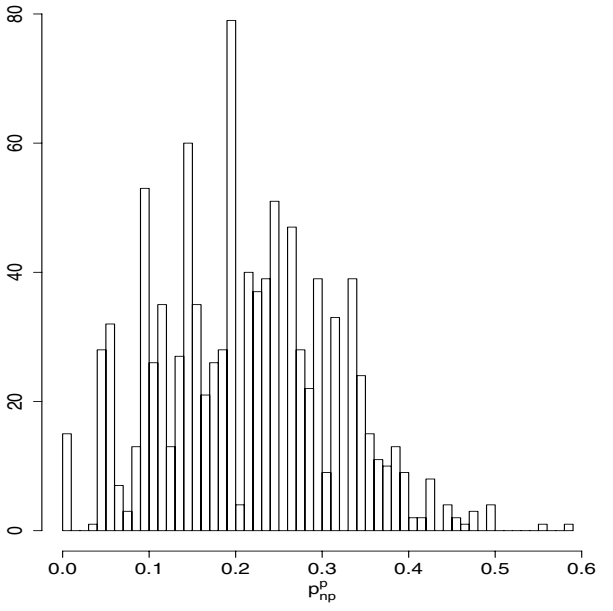

Fig. 3. Histogram of p^p_{np}, the percentage of non-periodic genes involved to connect the SCC to N^p_{tot} periodic genes

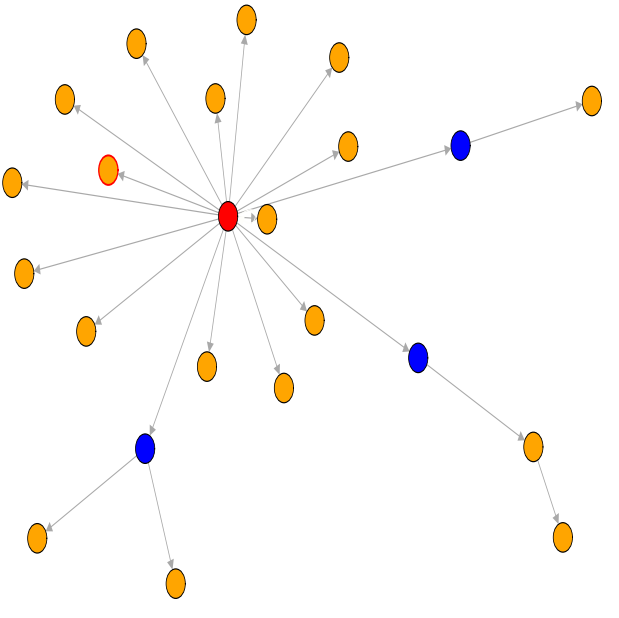

Fig. 4. Same as Fig. 1, however, for a gene label randomized network

That means, N_{np} is the number of non-periodic genes necessary to connect genes from the SCC with periodic genes outside the SCC. Figure 1 visualizes this for the transcriptional regulatory network and Fig. 4 for a randomized version thereof (see figure caption). For example, for Fig. 1 $N_{np} = 9$ (number of blue nodes). The histogram in Fig. 3 is again the result of $N_E = 1000$ randomizations. From Fig. 1 follows that $p^p_{TRN,np} = 0.068$. Using this value as threshold to calculate

$$p_b = \frac{1}{N_E} \sum_{p^p_{np}} I(p^p_{np} \leq p^p_{TRN,np}) \tag{2}$$

gives $p_b = 0.083$[3]. We want to remark that despite the fact that the largest cluster consists for the randomized networks of less than 30 periodic genes (see Fig. 3), in contrast to the TRN which consists of 132 periodic genes, this is a quite low number indicating that by chance the expected number of non-periodic genes necessary to obtain a connected cluster is larger. The fact, that non of the randomized networks is capable connecting close to 132 periodic genes from the SCC is even more striking and a strong indicator that our hypothesis is sensible unraveling possibly an evolutionary mechanism underlying the yeast cell cycle.

5 Conclusions

In this paper we investigated the organizational structure of the transcriptional regulatory network of yeast with respect to periodic genes. We started by assuming that the strongly connected component is playing a prominent role in the regulation of periodic genes because only genes from the SCC can be found on cycles - closed paths - and, hence, can be activated or deactivated cyclically forming a kind of trigger for genes that occur not on cycles in the TRN. Applying our hypothesis we find a surprisingly compact subnetwork that spans almost 80% of all periodic genes (Fig. 1). By the randomization of gene labels we could numerically demonstrate that this observed subnetwork and its constituting parts is unlikely to be formed by chance. This might be an indicator that this connectivity pattern has been forged by evolution rather than accidentally.

We want to remark that the property *cyclicity* of a network has been already previously used to study molecular networks meaningfully by separating proteins in their structural domains [10].

Acknowledgments

We would like to thank Galina Glazko for fruitful discussions and two anonymous reviewers for their comments that helped to improve this paper.

[3] In Eqn. 2 $I()$ is the indicator function which is 1 if the argument is true and 0 else.

References

1. Ahnert, S., Willbrand, K., Brown, F., Fink, T.: Unbiased pattern detection in microarray data series. Bioinformatics 22(12), 1471–1476 (2006)
2. Albert, R., Barabasi, A.: Statistical mechanics of complex networks. Rev. of Modern Physics 74, 47 (2002)
3. Arita, M.: The metabolic world of escherichia coli is not small. Proc. Natl. Acad. Sci. USA 101(6), 1543–1547 (2004)
4. Barabasi, A.L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. Nature Reviews 5, 101–113 (2004)
5. Bornholdt, S., Schuster, H. (eds.): Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH, Chichester (2003)
6. Dijkstra, E.: A note on two problems in connection with graphs. Numerische Math. 1, 269–271 (1959)
7. Dorogovtesev, S., Mendes, J.: Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, Oxford (2003)
8. Emmert-Streib, F., Dehmer, M. (eds.): Analysis of Microarray Data: A Network Based Approach. Wiley-VCH, Chichester (2008)
9. Emmert-Streib, F., Dehmer, M.: Analyzing and predicting periodic genes of the transcriptional regulatory network of yeast: A causality based approach (submitted) (2008)
10. Emmert-Streib, F., Mushegian, A.: A topological algorithm for identification of structural domains of proteins. BMC Bioinformatics 8, 237 (2007)
11. Jeong, H., Tombor, B., Albert, Z.N., Olivai, R., Barabasi, A.L.: The large-scale organization of metabolic networks. Nature 407, 651–654 (2000)
12. Luan, Y., Li, H.: Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data. Bioinformatics 20(3), 332–339 (2004)
13. Luscombe, N., Badu, M., Snyder, Y.H.M., Teichmann, S., Gerstein, M.: Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312 (2004)
14. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)
15. Palsson, B.: Systems Biology. Cambridge University Press, Cambridge (2006)
16. Rahman, S., Schomburg, D.: Observing local and global properties of metabolic pathways: load points and choke points in the metabolic networks. Bioinformatics 22(14), 1767–1774 (2006)
17. Schuster, H.G.: Deterministic Chaos. Wiley VCH Publisher, Chichester (1988)
18. Shmulevich, I., Dougherty, E.: Genomic Signal Processing. Princeton University Press, Princeton (2007)
19. Watts, D.: Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton (1999)
20. Watts, D., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998)
21. Wichert, S., Fokianos, K., Strimmer, K.: Identifying periodically expressed transcripts in microarray time series data. Bioinformatics 20(1), 5–20 (2004)

22. Yu, H., Kim, P., Sprecher, E., V., T., Gerstein, M.: The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. PLoS Computational Biology 3(4), e59 (2007)
23. Zhao, L., Prentice, R., Breeden, L.: Statistical modeling of large microarray data sets to identify stimulus-response profiles. Proc. Natl. Acad. Sci. USA 98(10), 5631–5636 (2001)
24. Zhu, D., Qin, Z.: Structural comparison of metabolic networks in selected single cell organisms. BMC Bioinformatics 6, 8 (2005)