Networks of interactions between feed-forward loop transcriptional motifs in gene-regulatory networks

Michael Mayo
Environmental Laboratory, US Army Engineer Research and Development Center
Vicksburg, MS 39180
Michael.L.Mayo@usace.army.mil

Ahmed Abdelzaher
Department of Computer Science
Virginia Commonwealth University
Richmond, VA, USA
abdelzaheraf@vcu.edu

Bhanu K. Kamapantula
Department of Computer Science
Virginia Commonwealth University
Richmond, VA, USA
kamapantulbk@vcu.edu

Edward J. Perkins
Environmental Laboratory, US Army Engineer Research and Development Center
Vicksburg, MS 39180
Edward.J.Perkins@usace.army.mil

Preetam Ghosh*
Department of Computer Science
Virginia Commonwealth University
Richmond, VA, USA
pghosh@vcu.edu

ABSTRACT

Transcriptional motifs are smaller subnetworks found within the gene-regulatory networks of many organisms in larger abundance than can be explained by chance alone. The feed-forward loop is one such three-node motif, wherein one top-level protein regulates the expression of a target gene either directly, or indirectly through an intermediate regulator protein. However, no systematic effort has yet been made to understand how individual feed-forward loops interconnect. Here, we address this problem by examining embedded transcriptional motifs that interact topologically by sharing one (vertex-share graphs), two (edge-share graphs), or three (triad-share graphs) nodes. Using transcriptional networks of the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, we constructed networks of feed-forward loops based on these interaction patterns, and termed them "motif networks." In view of these motif networks, we show that, on average, feed-forward loops connect primarily to others similarly connected–a phenomenon termed assortativity or homophily and often attributed to social networks. We fit these correlations to a power-law equation, which exhibits a sublinear exponent indicative of an "economy of scale" in the FFL connectivity. We show that connectivity distributions of the motif networks (similar to degree distributions in complex networks) appear approximately uniform, but with a large variance. Although assortative mixing may arise from a scale-free degree distribution, we conclude that assortativity observed here arises by alternative means.

*To whom correspondence should be addressed.

BICT 2014, December 01-03, Boston, United States

DOI 10.4108/icst.bict.2014.257926

Categories and Subject Descriptors

[**C.2.1**]: Network Architecture and Design

General Terms

Theory; Algorithms

Keywords

Assortativity; Degree correlations; Graph transformations; Feed-forward loop

1. INTRODUCTION

In a transcriptional network, the genes of an organism are the nodes, while the links between them denote whether the expressed proteins of one gene regulates the activity of another, either positively or negatively. Despite being

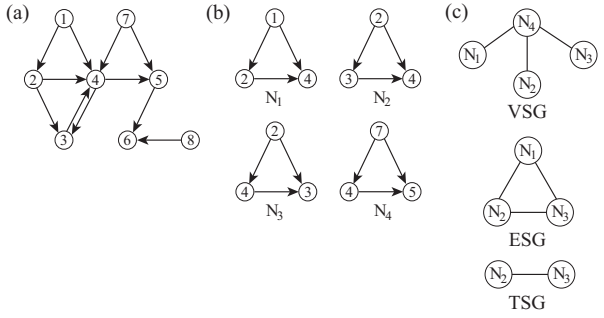

Figure 1: Steps to form a vertex-based motif network. (a) An exemplary network; (b) Feed-forward loops taken from the example network are contracted into nodes labeled N_1-N_4; (c) Networks generated if feed-forward loops share exactly one (VSG), two (ESG), or three (TSG) vertices.

sparse, transcriptional networks posses many types of repeating subnetworks, termed network motifs [15]. One of them, the feed-forward loop (FFL) transcriptional motif, has been identified using transcriptional networks of the bacterium *Escherichia coli* (*E. coli*) and the common baker's yeast *Saccharomyces cerevisiae* (*S. cerevisiae*), wherein they occur in abundances higher than those of randomized versions of these same networks [9]. In *E. coli* the FFL is the most common of all 13 distinct directed three-node patterns [15]. In addition, theoretical and experimental investigations have shown them to delay, accelerate, or dynamically pulse transcriptional signals, depending on the level of the upstream gene stimulus [9, 5, 10, 6, 8, 12]. This knowledge has been used, for example, to predict interaction partners of proteins in protein-interaction networks [1]; to assist in classification of networks [14]; and to analyze structural network properties [21].

Although much work has so far focused on dynamical properties of the various types of individual feed-forward loops (e.g., see Ref. [9] and references therein), less is known relating how one feed-forward loop generally interacts with another. Whether or not these transcriptional motifs are functionally buffered from the influence of the greater network remains unclear [25], despite that some examples of serially-coupled transcriptional motifs have been identified, such as in cell-cycle regulation [18, 23, 24]. Some progress has been made in this area through a topological analysis of similar networks [3, 11].

Here, we address this problem at the topological level by first defining what it means for two feed-forward loops to interact without reference to the underlying state of the biochemical molecules, such as phosphorylation or complexation states. Next, we carry out a systematic computational study to identify these interaction partners within sampled transcriptional networks from the well-studied bacterium *E. coli* and the yeast *S. cerevisiae*. This analysis results in a new network that expresses the individual feed-forward loops within a given transcriptional network as a single node, and the links between them signify connectivity between loops. Collectively these data form new networks, termed here "motif networks."

2. METHODS

2.1 Transcriptional networks

Transcriptional regulatory networks of several simple organisms have been experimentally and exhaustively interrogated. Among these is the transcriptional regulatory network of the bacterium *E. coli*, which we work with here primarily due to the ready availability of its dataset. Model transcriptional networks of varying size were obtained for *E. coli* using the GeneNetWeaver tool [22], which provides directed subnetworks of user-defined size from the full *E. coli* dataset, which is 1565 nodes and 3758 edges. The *S. cerevisiae* network was similarly derived, and composed 4441 nodes and 12873 links.

2.2 Identifying transcriptional motifs in model networks

Feed-forward loops were extracted from the topology of the transcriptional networks using the mFINDER software [7],

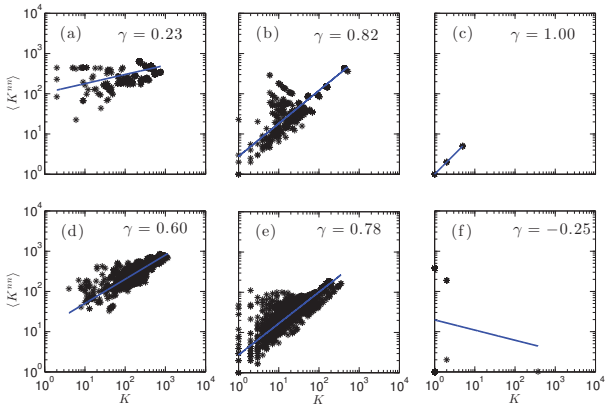

Figure 2: Degree assortativity measured by the average degree of nearest neighbors, $\langle K^{nn} \rangle(K)$, in the motif networks derived from either *Escherichia coli* (panels (a)-(c)) or *Saccharomyces cerevisiae* (panels (d)-(f)). Degree assortativity is shown for VSGs (panels (a) & (d)), ESGs (panels (b) & (e)), and TSGs (panels (c) & (f)), with empirical fits to a power-law equation: $\langle K^{nn} \rangle(K) \propto K^{\gamma}$ (blue lines).

which employed the following detection scheme. An algorithm counts all n-node motifs based on locating edges; it first begins from a single edge, and then searches for all of the n-node subgraphs to which it belongs. An array of hash tables stores all sets of nodes that have been visited during detection, which reduces the search time when compared to other methods, primarily by terminating the search tree if a set (or subset) of nodes have already been visited/recorded. This step is repeated for all network edges, and counts for each subgraph type are recorded.

Randomized networks were generated to resolve motifs within the input network. By default, mFINDER generates such networks to locate significant subgraphs by using a "switching" method: the number of incoming, outgoing, and shared edges of each node of the input network are fixed, while the search method switches between candidate edges. The number of switches is random, chosen from a range 100-200 fold the number of edges in the candidate network.

2.3 Constructing motif networks

The mFINDER software package [7] was used to identify feed-forward loop transcriptional motifs within the sampled transcriptional regulatory networks. Next, each of these loops was represented as a single node of a new network. This concept is illustrated in Fig. 1, where the feed-forward loops have been labeled as nodes N_1-N_4, and the links between placed according to the following rules.

A link was drawn between two motifs if (i) they shared exactly one vertex between them; (ii) they shared exactly two vertices (i.e. shared a link); or (iii) shared all three of their vertices. Item (iii) is possible if two feed-forward loops are embedded within the same subnetwork; given that transcriptional networks are sparse, we should expect that

such links are rare. The networks built using these items are collectively termed vertex-based "motif networks" (VMNs). Specifically, networks built from (i) are termed vertex-share graphs (VSGs); those built from (ii) are termed edge-share graphs (ESGs); and those built from (iii) are termed triad-share graphs (TSGs).

3. RESULTS

3.1 Feed-forward loops couple assortatively

The structure of many complex networks are degree correlated, a property termed assortativity [16] or homophily, wherein nodes tend to connect with others of similar degree. In particular, social networks have been widely shown to be assortatively connected [13], whereas technological networks are not [17]. Counterexamples to these general rules have been presented, such as disassortativity in social networks [4], or assortative mixing in biological networks [2, 20]. That some networks exhibit assortative correlations is more mysterious, given that the "natural" state of a complex network appears to be the disassortative state [19]: given a scale-free network, nodes with higher degree are less abundant, so the pool of available connecting nodes is mostly of lower degree, hence a disassorative outcome.

We measured the degree correlations in the three types of motif networks (Fig. 1), built from both *E. coli* and *S. cerevisae* transcriptional networks, by measuring the average connectivity of feed-forward loops, $\langle K^{nn}\rangle(K)$, neighboring one with connectivity K. As shown in Fig. 2, we found that feed-forward loops connect with others assortatively, in the sense of a strong positive power-law relationship: $\langle K^{nn}\rangle(K) \propto K^{\gamma}$. In other words, for VSGs and ESGs, we found that more highly-connected feed-forward loops tended to connect with others of higher motif connectivity.

The sublinear nature of the exponent γ is indicative of an "economy of scale" in feed-forward loop connectivity. If we relate the average neighbor connectivity to each FFL connection, we will find that it drops off with increasing K for $\gamma < 1$: $\langle K^{nn}\rangle/K \propto K^{\gamma-1}$. Put another way, the more lightly coupled FFLs support, on average, more FFLs per unit connection than a more highly coupled one. This observation suggests, counter-intuitively, that information sharing (e.g. propagation of a stimulus effect) between FFLs in dense, highly-connected clusters may be less effective than FFLs inhabiting the more loosely coupled regions outside of the cluster.

3.2 Connectivity distributions

Figure 3 illustrates connectivity distributions among feed-forward loops embedded within *E. coli* and *S. cerevisae* transcriptional networks. We measured the probability, $p(K)$, that a feed-forward loop in these networks is connected to K-many others, and calculated these distributions for all three types of vertex-share motif networks: VSG, ESG, and TSG.

For the vertex-share graphs, the degree distribution is nearly uniform, as can be seen for both *E. coli* (Fig. 3(a) and *S. cerevisiae* (Fig. 3(d)). Thus, nearly every feed-forward loop that shares a vertex with another loop connects to roughly

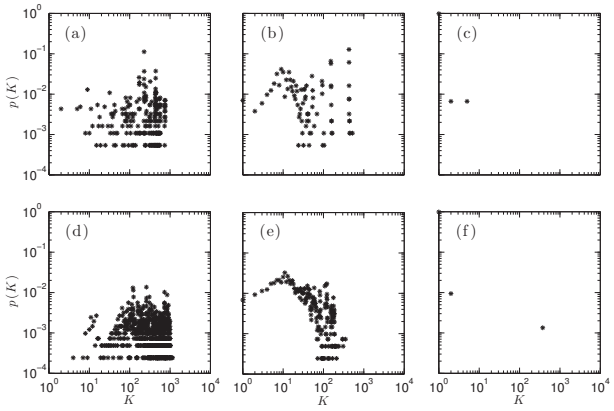

Figure 3: Connectivity distributions of embedded feed-forward loops calculated from either *Escherichia coli* (panels (a)-(c)) or *Saccharomyces cerevisiae* (panels (d)-(f)) transcriptional networks. Here, the probability, $p(K)$, to find a feed-forward loop connected to K-many others is computed for either VSGs (panels (a) & (d)), ESGs (panels (b) & (e)), or TSGs (panels (c) & (f)).

the same number of other such loops by one vertex. Thus, there are roughly an equal number of feed-forward loops with high connectivity as there are those of lower connectivity. This is unexpected, because fewer numbers of individual transcription factors or genes in such networks participate in multiple feed-forward loops [11].

This behavior differs for the edge-share graphs, which show a more correlated connectivity distribution by exhibiting a power-law tail at high connectivity. While this feature is present in the connectivity distributions for the *E. coli* ESG, the distribution for *S. cerevisae* exhibits much less variance in the trend.

Finally, triad-share graphs display different connectivity distribution patterns. Figures 3(c) and 3(f) indicate that very few embedded feed-forward loops share all three vertices, which may result from the lower availability of fully or near-fully connected three node subnetworks.

4. CONCLUSIONS

We examined correlations in the connectivity among feed-forward loop transcriptional motifs in both *E. coli* and *S. cerevisiae* networks, which exhibited strong assortative tendencies. Specifically, feed-forward loops were observed to couple assortatively for both VSGs and ESGs. This observation is consistent with previous reports that feed-forward loops tend to cluster into dense clusters [3], from which it could be hypothesized that highly connected motifs connect similarly to others within the cluster. We have tested this hypothesis, demonstrating assortative coupling between feed-forward loops throughout the whole of a transcriptional network–not just those residing in large clusters. A power-law fit showed the assortative FFL mixing followed a sublinear trend representative of an "economy of scale" in the

FFL connectivity; thus, less coupled FFLs supported more FFLs per unit connection than the highly coupled ones.

The natural state of a complex network, such as a transcriptional network, may be skewed toward disassortativity due to the underlying nature of the degree distribution [19]. For scale-free networks, highly-connected nodes are rare, so the pool for new connections is dominated by lower-degree nodes, leading to disassortative degree mixing. We tested whether this mechanism could be responsible for the observed assortative connections between feed-forward loops by measuring the motif connectivity distributions, which are similar to degree distributions in complex networks (Fig. 3). We observed that these distributions were mostly uniform, but with a large variance. Based on this evidence we reject the above hypothesis, and conclude that an alternative mechanism is responsible for the assortative trends we observed. Further investigations will be required to better understand what biological phenomena manipulates a potentially random network of gene connections into an assortative mixture of feed-forward loop modules.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their informative comments. This work was supported by grant number NSF-1143737, and the US Army's Environmental Quality and Installations 6.1 Basic Research program. Opinions, interpretations, conclusions, and recommendations are those of the author(s) and are not necessarily endorsed by the U.S. Army.

5. REFERENCES

[1] I. Albert and R. Albert. Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics*, 20:3346–3352, 2004.

[2] G. Bagler and S. Sinha. Assortative mixing in protein contact networks and protein folding kinetics. *Bioinformatics*, 23:1760–1767, 2007.

[3] R. Dobrin, Q.K. Beg, A.L. Barabasi, and Z.N. Oltvai. Aggregation of topological motifs in the escherichia coli transcriptional regulatory network. *BMC Bioinformatics*, 5:10, 2006.

[4] H.B. Hu and X.F. Wang. Disassortative mixing in online social networks. *Europhys. Lett.*, 86:18003, 2009.

[5] S. Kalir, S. Mangan, and U. Alon. A coherent feed-forward loop with a sum input function prolongs flagella expression in escherichia coli. *Mol. Syst. Biol.*, 1(2005.0006), 2005.

[6] S Kaplan, A Bren, E Dekel, and U Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol. Syst. Biol.*, 4:203, 2008.

[7] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. Mfinder tool guide, 2002.

[8] Dongsan Kim, Yung-Keun Kwon, and Kwang-Hyun Cho. The biphasic behavior of incoherent feed-forward loops in biomolecular regulatory networks. *Bioessays*, 30:1204–1211, 2008.

[9] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100(21):11980–11985, 2003.

[10] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *J. Mol. Biol.*, 356:1073–1081, 2006.

[11] M. Mayo, A. Abdelzaher, E.J. Perkins, and P. Ghosh. Motif participation by genes in *E. coli* transcriptional networks. *Front. Physiol.*, 3:357, 2012.

[12] M. Mayo, A. Abdelzaher, E.J. Perkins, and P. Ghosh. Top-level dynamics and the regulated gene response of feed-forward loop transcriptional motifs. *Phys. Rev. E*, 90:032706, 2014.

[13] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.*, 27:415–444, 2001.

[14] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.

[15] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[16] M.E.J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[17] M.E.J. Newman and J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 68:036122, 2003.

[18] B. Novak, A. Csikasz-Nagy, B. Gyorffy, K. Chen, and J.J. Tyson. Mathematical model of the fission yeast cell cycle with checkpoint controls at the g1/s, g2/m and metaphase/anaphase transitions. *Biophys. Chem.*, 72(1):185–200, 1998.

[19] J. Park and M.E.J. Newman. Origin of degree correlations in the internet and other networks. *Phys. Rev. E*, 68:026112, 2003.

[20] M. Piraveenan, M. Prokopenko, and A. Zomaya. Assortative mixing in directed biological networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 9:66–78, 2012.

[21] N. Przulj, D.G. Corneil, and I. Jurisica. Modeling interactome: scale-free or geometric. *Bioinformatics*, 303:3508–3515, 2004.

[22] T. Schaffter, D. Marbach, and D. Floreano. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, 2011.

[23] A. Sveiczer, A. Csikasz-Nagy, B. Gyorffy, J.J. Tyson, and B. Novak. Modeling the fission yeast cell cycle: Quantized cycle times in *wee1*$^-$ *cdc25*δ mutant cells. *Proc. Natl. Acad. Sci. USA*, 97(14):7865–7870, 2000.

[24] J.J. Tyson, K. Chen, and B. Novak. Network dynamics and cell physiology. *Nat. Rev. Mol. Cell Biol.*, 2(12):908–916, 2001.

[25] D.M. Wolf and A.P. Arkin. Motifs, modules and games in bacteria. *Curr. Opin. Microbiol.*, 6(2):125–134, 2003.