

Empirical Models for Complex Network Dynamics: A Preliminary Study

Douglas Oliveira and Marco Carvalho

Florida Institute of Technology
Department of Computer Science
Melbourne, Florida, USA
doliveira2011@my.fit.edu, mcarvalho@cs.fit.edu

Abstract. Network analysis has drawn a considerable amount of attention in the last decade, especially after the discovery of common topological characteristics such as Small World or a Power Law degree distribution. Recently our understanding of complex networks has been augmented with the inclusion of a local view of patterns of connectivity, such patterns that are present more often in real networks than in randomized ones have been called motifs. These global and local perspectives equip us with powerful tools to understand the behavior of many real networks. However, an important aspect of complex network analysis is often neglected: the dynamics of the information flows. The structural elements of the network topology are very important, but to fully understand the dynamics of these networks we need to take a closer look at the dynamics of the information flow in a self-regulation perspective. For example, we know that the performance and reliability of a compute network is likely influenced by the dynamics of the packet flows, as much as it is influenced by the network topology. In a biological regulatory network we need to understand the dynamics that control the excitation and the suppression of gene activity and other transcription factors. In this work we introduce a preliminary simulation study of the flow of information in networks with different topological properties and activation functions. The goal is to approach the analysis of network dynamics from a data-driven approach, using simulations to capture, understand, and possibly model the overall dynamics of the network in a self-regulated perspective.

1 Introduction

The study of networks is not recent, it dates back from the eighteenth century, from the studies of Euler in the problem of seven bridges [1]. He used a graph, a mathematical structure that consists of nodes and edges connecting them, to prove that there was no solution to such problem. Many theories, algorithms and concepts have been developed ever since, however what has been called of the new network science is quite different from the old network science [2]. The new network science is concerned with the structure of naturally occurring network, rather than theoretical ones, like social networks [3], biological networks [4] and communications networks [5].

Among the concepts presented with this new network science we can mention Small World networks [6] and Scale Free networks [7]. In a Small World network the distance between any two node is short, precisely it should be $O(\log n)$, where n is the number of nodes of the network. The distance between any two nodes is normally calculated using a shortest path algorithm [8]. Not only the average shortest path must be short but also network must be clustered, which holds the percentage of possible cliques of size three in the network. When the network is called Scale Free, its degree distribution follows a Power Law distribution, formally the probability of a node having degree k is: $p(k) = k^{-\lambda}$.

Moving beyond global properties, the work of [9] presents a concept called motifs. Motifs are basic structural elements, subgraphs, that are present more often in real networks rather than in randomized ones. Depending of the kind of network different motifs arise, for example in information-processing network motifs tend to create multiple paths among different nodes in order to improve robustness of the network. Motifs provide a local view of the network, each node only observes its own connections and the connections of its neighbors to uncover the motifs in its neighborhood. Depending of the types of motifs and their frequency in the network many other properties can be deduced [10].

However, all these properties, either global or local ones, are not enough to properly understand the behavior of many real networks. For example, in a cell regulatory network, that can change its function rapidly depending of the part of the cell cycle that it is in or when there are unexpected changes in the environment due to an stress factor. In these cases, the static properties of the networks remain the same and yet its function changes, this happens because of changes in the information flowing through the network. Such changes can be in the amount of information, in the type of information or in the direction of the information flow. Another characteristic inherent to the dynamics is that each node has its own state that can vary in time, thus any dynamic analysis needs to consider the evolution of the network in time.

According to [11] we are far from comprehend how the collective behavior of thousand of nodes interacting locally contribute to the dynamic behavior that we observe in many real systems. Once we understand how all these nonlinear interactions on the dynamics of real world networks we will have a better understanding of their behavior, even in stressful situations (in a cellular network it could be a cancer state). For this purpose in this work we simulate the flow of information that will be processed by the network, and we measure the various states of the network along the time in many different topological structures. We expect to find the conditions necessary to the self-regulation of the network, in other words, the parameters of the network that will maintain its information flowing without any external interference.

2 Related Work

Several research efforts in the literature focus on the dynamics of the networks from a growth perspective, trying to understand which processes lead to the

appearance of certain characteristics in real networks [12]. For example, the work of [13] proposes a model to generate Scale Free networks based on two main principles: that real networks are constantly evolving by the addition of new nodes and new vertices attach preferentially to already well connected nodes.

Another dynamic process commonly found in literature is the study of diffusion process, like the one that happens in disease spread[14]. In this work the authors claim that the robustness present in Scale Free networks is a disadvantage in this situation once it favors the spreading of viruses. A simulation of the spread of the pandemic disease H1N1 was made by [15] that focused in the role of travel restrictions in halting and delaying the spread of disease. Similar works analyzed diffusion processes in Small World networks [16]. The work of [17] analyze the diffusion pattern of websites visitation and claims that the timing of the browsing process is non-poisson contradicting previous works [18].

The authors of [19] propose a realistic model for dynamics in social networks that takes into account digital and social aspects of online social networks. Their work deals with information propagation, that can be either truthful or not (rumor), and how people reacts by receiving such information, if they believe it or not and if they decide to pass such information to its neighbors or not. This works differs considerably from ours in the kind of information that its propagated in the network. In our work we deal with the amount of information that each nodes passes to its neighbors, thus it is represented as a numerical value. In the work of [19] the information has a dual type, positive or negative, and depending of its type it will propagate more or less.

In the work of [20] the authors present the a statistical analysis of the dynamics in a biological network on a genomics scale by combining gene-expression data with transcriptional regulatory information in many different scenarios. Their results shown that less than one percent of the interactions are retained across four or more conditions and that half of the nodes are uniquely expressed in only one condition. Such findings can be easily incorporated in our model.

In this work each node in the network receives an input and following the behavior of a specific function produces an output that is broadcasted to its neighbors. An analogy can be made with artificial neural networks [21], which also receive inputs and produce outputs based on a given function, however the similarities stop there. In artificial neural networks the structure is organized in layers, which have the purpose of learning a pattern from previous data aiming to adjust several parameters to solve a particular problem. Artificial neural networks have been used in many predicting tasks, like electric load forecasting [22] and in pharmaceutical research [23]. The goal of the model used in this work is to analyze how the information flows in the network, it does not have any direction that the information must flow or try to solve any problem.

3 The Simulation Model

In order to simulate dynamic processes of information flow on a network we create a model where, at each time step t , each node i in the network receives

input signals from each of its incoming neighbor r . The input signal is denoted by $\varphi_r(t)$ and can be either excitatory or inhibitory. Depending of the strength of these signals and the function (γ) used to calculate the output, the node broadcast an output signal to all its output neighbors, which is denoted by $\omega_i(t)$.

At each time step t each node i receives input signals from all its neighbors, the sum of all these inputs, that will be computed, is denoted by $v_i(t)$ and is formally defined as:

$$v_i(t) = \sum_{0 < r < k} \varphi_r(t) \quad (1)$$

Where k is the number of neighbors of node i . The Figure 1 presents an example of a node in the model. The node i receives input signals $\varphi_1(t)$, $\varphi_2(t)$ and $\varphi_3(t)$ from each of its three incoming neighbors (that will compose $v_i(t)$) and thus will produce an output $\omega_i(t)$ that will be broadcast to its outgoing neighbors.

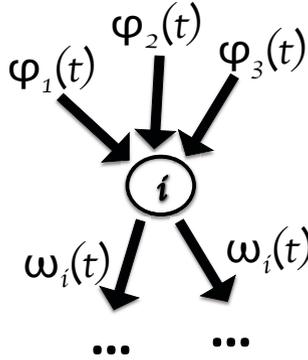


Fig. 1. Example of information flow in a node of the network

To calculate the output, each node uses a function γ on the input $v_i(t)$, we used two well known functions. The first is called Hill function [24]. The Hill function is a common sigmoid function presented in many biological processes, like tumor growth [25]. It consists of a curve that rises from zero and approaches a maximal saturated level δ . Formally, it is defined as:

$$\gamma(v_i(t)) = \frac{\delta}{1 + b^{-v_i(t)}} \quad (2)$$

Where b is the base of calculation, as the value of b increases the function converges faster towards the saturation point δ . At the beginning of the simulation each node defines randomly its own value of b , this variation was made to incorporate the variations of output of different nodes that might receive the same amount of input signals, $v_i(t)$. The second function used is a Poisson function which is widely used and defined by the mean and standard deviation [26].

4 Results

As a result of this model the network's ability to maintain itself 'alive' without any external input is characterized by three main factors: the function (γ) used to calculate the output, the topology of the network and the initial input given to start the simulation. We first analyze the relationship between the topology and γ .

In this work we assume that the network is in a "stable" state if a certain amount of the nodes remains in an state that is neither dead or saturated without any external input. Nodes in a saturated state receive continuously the maximum input, thus producing its maximum output. This situation may occur due to a loop between two nodes that the output of one is the input of the other and the same happens with the other node. A dead state is a state where the node does not receive any input, thus it cannot produce any output, perhaps due to the lack of input in its incoming neighbors. Such "stable" state is an indicative that the network will not reach a point where no node receives any input (in a biological network this means that the cell died) or all the nodes receive the maximum input, which is not realistic. In order to investigate if such behavior occurs in a large network with real world properties we defined the following metric:

$$\Theta(t) = 1 - \frac{(\eta(v_i^0(t)) + \eta(v_i^\delta(t)))}{n} \quad (3)$$

Where the function η computes the number of nodes at time t that have $v_i(t)$ equals to a specific value. In the case of saturation this value is δ and in a 'dead' state this value is zero. Thereby, $\Theta(t)$ holds the percentage of nodes that are in neither states at each time t . From a global perspective of the network, in order to maintain itself functional the network aims to maximize the value of $\Theta(t)$.

We create a network with the same topological characteristics of the regulatory network of *E. coli* [27] that has 126 nodes and 326 edges, which can be seen in Figure 2. In Figure 2 the nodes were colored accordingly to their degree, the majority of nodes only have one or two edges and few nodes connect the entire network, like 'crp' and 'h-ns'. The topological properties of this network were analyzed in more details in [28]. Thus we simulate the flow of transcription factors through the network that can excite or inhibit the production of other transcription factor in other nodes.

We measured the variation of $\Theta(t)$ over the time in order to analyze the collective behavior of the network maintain itself in a 'stable' state without any external interference. As can be seen in Figure 3, when we use the Hill function (left) the nodes soon converges to a dead or a saturated state, the percentage of nodes that died is around 30% and the percentage of nodes saturated is around 70%, which leads to $\Theta(t)$ remain close of zero, during the rest of the simulation with almost no change. When we use the Poisson function we perceive a more dynamical behavior of the network, where a few percentage of nodes are dead, less than 10%, and the percentual of node in a saturated state oscillates around 50 and 70%.

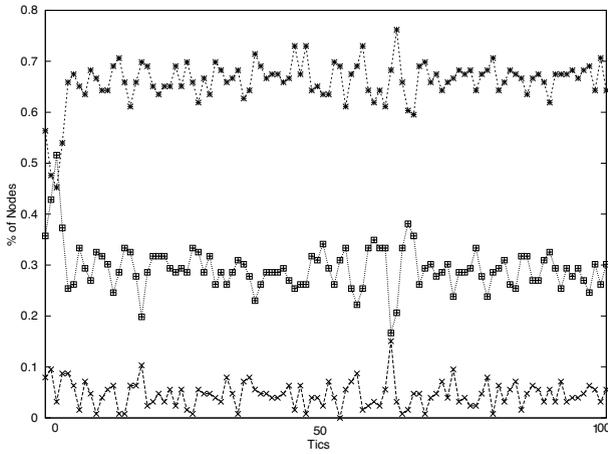


Fig. 4. Variation of the percentage of nodes saturated (X's), dead (asterisk) and $\Theta(t)$ (squares) over time in a Small World network

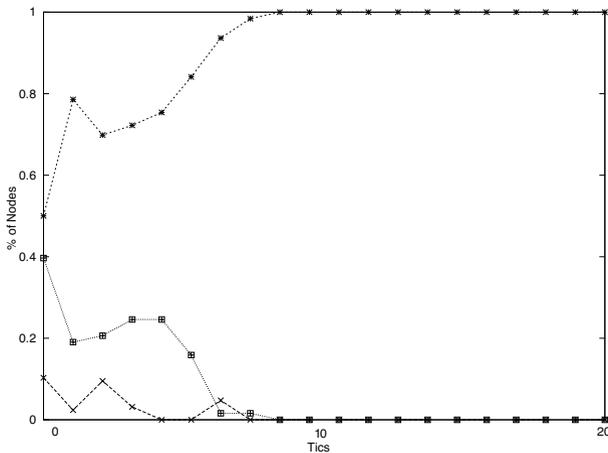


Fig. 5. Variation of the percentage of nodes saturated (X's), dead (asterisk) and $\Theta(t)$ (squares) over time in a Scale Free network

In this case we constructed the network randomly, accordingly to [29]. In order compare the results with a Small-World network, we create an equivalent network (with the same number of nodes and edges) with such characteristics, following the algorithm proposed in [30]. In the Figure 4 we present the results of the simulation of our model in such network that uses a Poisson function. We notice that the results are very similar to the results of the random network Figure 3 (right), although there is a larger percentage of nodes that saturates.

We performed a similar analysis in a Scale Free network. Preserving the dimensions of the network (number of nodes and edges), we rewired the edges aiming that its degree distribution follow a power law distribution with $\lambda=3$, accordingly to [13]. In Figure 5 we see that although the network uses a Poisson function it exhibits a very similar behavior of the randomly created network in Figure 3 (left) where $\Theta(t)$ soon converges to zero. The main difference between these results is that in the Scale Free network the majority of nodes dies and in the random network the majority of nodes saturates.

5 Conclusions and Future Work

In the last few years many advances has been made in network science, including the discovery of Small World properties and its numerous applications, and the characterization of Scale Free networks. Recurrent patterns of local connectivity were also shown to play important role in complex networks, and several algorithms and techniques have been proposed to find and characterize such structures. In this paper we focus on the analysis and understanding of network flow dyanmics. While preliminary, the work proposes a simulation-based model to represent the flow of information. At each time step, each node receives an input from its incoming neighbors and accordingly to a function produces an output that is broadcasted to its output neighbors.

We first introduce two different functions that for node activation: a Hill function and a Poisson function. The Hill function rapidly leads the network to an unrealistic state where almost the totality of nodes are either dead or saturated. In the other hand the Poisson function presented more dynamic results where the percentage of nodes saturated is low and the percentage of nodes dead is not so high, thus the network maintain itself, without any external interference, in such stable state.

We discussed the flow analysis for a Small World network, in contrast to a comparable random network. The results showed, apparently, a similar behavior between the two topologies. The major difference is an increase in the percentage of nodes that saturated, perhaps due to the clustering characteristic of the network that lead the information reach faster between any pair of nodes of the network. We intend to deeper investigate these results aiming to provide a more solid comparison. We also compared the results of such randomly created network with a Scale Free that has the same dimensions. We perceived that due to the majority of nodes only have a couple edges the network rapidly evolves to a completely dead state.

As part of our future work, we intend to further investigate the specific phenomena in Small World networks and Scale Free networks that lead to such behavior presented here. For example, if we change the exponent λ of the power law distribution in Scale Free network the results will remain the same? If the network is more clustered in our Small World version the value of $\Theta(t)$ will increase ou decrease? We also aim to include different motifs in the network and see how if affect the dynamic of the network locally. The possibilities are very

broad, thus we also need to come up with other ways to measure the dynamic flow of information of the network

References

1. West, D.B.: Introduction to graph theory. Prentiss Hall, Upper Saddle River (1996)
2. Newman, M., Barabási, A.-L., Watts, D.J.: The structure and dynamics of networks
3. Rapoport, A., Horvath, W.J.: A study of a large sociogram. *Behavioral Science* 6(4), 279–291 (1961)
4. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L.: The large-scale organization of metabolic networks. *Nature* 407(6804), 651–654 (2000)
5. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.* 29(4), 251–262 (1999)
6. Watts, D., Strogatz, S.: The small world problem. *Collective Dynamics of Small-World Networks* 393, 440–442 (1998)
7. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97 (2002)
8. Dijkstra, E.W.: A note on two problems in connexion with graphs. *Numerische Mathematik* 1(1), 269–271 (1959)
9. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. *Science* 298(5594), 824–827 (2002)
10. Kashtan, N., Itzkovitz, S., Milo, R., Alon, U.: Topological generalizations of network motifs. *Physical Review E* 70(3), 031909 (2004)
11. Argollo de Menezes, M., Barabási, A.-L.: Fluctuations in network dynamics. *Phys. Rev. Lett.* 92, 028701 (2004)
12. Newman, M., Barabási, A.-L., Watts, D.J.: The structure and dynamics of networks. Princeton University Press (2006)
13. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. *Science* 286(5439), 509–512 (1999)
14. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. *Physical Review Letters* 86(14), 3200 (2001)
15. Bajardi, P., Poletto, C., Ramasco, J.J., Tizzoni, M., Colizza, V., Vespignani, A.: Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic. *PloS One* 6(1) (2011)
16. Monasson, R.: Diffusion, localization and dispersion relations on small-world lattices. *The European Physical Journal B-Condensed Matter and Complex Systems* 12(4), 555–567 (1999)
17. Jeong, H., Néda, Z., Barabási, A.-L.: Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)* 61(4), 567 (2003)
18. Kingman, J.F.C.: Poisson processes, vol. 3. Oxford University Press (1992)
19. Wen, S., Haghighi, M.S., Chen, C., Xiang, Y., Zhou, W., Jia, W.: A sword with two edges: Propagation studies on both positive and negative information in online social networks. *IEEE Transactions on Computers* (2013)
20. Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M.: Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431(7006), 308–312 (2004)
21. Yegnanarayana, B.: Artificial neural networks. PHI Learning Pvt. Ltd. (2004)

22. Park, D.C., El-Sharkawi, M.A., Marks, R.J., Atlas, L.E., Damborg, M.J., et al.: Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems* 6(2), 442–449 (1991)
23. Agatonovic-Kustrin, S., Beresford, R.: Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis* 22(5), 717–727 (2000)
24. Alon, U.: *An introduction to systems biology: Design principles of biological circuits*, vol. 10. CRC Press (2007)
25. Ai, B.-Q., Wang, X.-J., Liu, G.-T., Liu, L.-G.: Correlated noise in a logistic growth model. *Physical Review* (2003)
26. Marx, M.L., Larsen, R.J.: *Introduction to mathematical statistics and its applications*. Pearson/Prentice Hall (2006)
27. Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., et al.: Regulondb v8. 0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research* 41(D1), D203–D213 (2013)
28. Oliveira, D., Carvalho, M.: A comparison of community identification algorithms for regulatory network motifs. In: *IEEE International Conference on BioInformatics and BioEngineering* (2013)
29. Erdos, P., Rényi, A.: On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–61 (1960)
30. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. *Nature* 393(6684), 440–442 (1998)