

Network Analysis Using Spatio-Temporal Patterns

Gisele H B Miranda¹, Jeaneth Machicao² and Odemir M Bruno²

¹ Institute of Mathematics and Computer Science, University of São Paulo. Av. Trabalhador São Carlense, 400, São Carlos, SP - Brazil

² São Carlos Institute of Physics, University of São Paulo. Av. Trabalhador São Carlense, 400, São Carlos, SP - Brazil

E-mail: gimiranda@usp.br

Abstract. Different network models have been proposed along the last years inspired by real-world topologies. The characterization of these models implies the understanding of the underlying network phenomena, which accounts structural and dynamic properties. Several mathematical tools can be employed to characterize such properties as Cellular Automata (CA), which can be defined as dynamical systems of discrete nature composed by spatially distributed units governed by deterministic rules. In this paper, we proposed a method based on the modeling of one specific CA over distinct network topologies in order to perform the classification of the network model. The proposed methodology consists in the modeling of a binary totalistic CA over a network. The transition function that governs each CA cell is based on the density of living neighbors. Secondly, the distribution of the Shannon entropy is obtained from the evolved spatio-temporal pattern of the referred CA and used as a network descriptor. The experiments were performed using a dataset composed of four different types of networks: random, small-world, scale-free and geographical. We also used cross-validation for training purposes. We evaluated the accuracy of classification as a function of the initial number of living neighbors, and, also, as a function of a threshold parameter related to the density of living neighbors. The results show high accuracy values in distinguishing among the network models which demonstrates the feasibility of the proposed method.

1. Introduction

In the last decade, there has been a growing interest in Complex Networks, which is an area directly related to the study of complex systems [1, 2, 3]. The possibility of modeling different systems through the interactions between their constituent elements motivates the representation of discrete systems by means of a network structure. The focus of analysis becomes the connectivity among the elements of the system and not only the elements themselves. Many real world networks became available and significant studies have been conducted on the analysis of network structures with a large number of vertices. These networks have nontrivial topology and present communities and hierarchies [3, 4]. Consequently, statistical tools combined with graph theory have provided a more robust study of the topology of networks. The importance of network characterization is becoming more evident due to the vast amount of applications that can be handled by this approach. Different areas comprising biology, medicine, epidemic spreading, engineering, transportation, economy, scientific collaboration and social interactions are a few examples that can be benefited from network modeling. Many studies accounted that real networks are scale-free, with degree distributions that follows a power-law. The models that describe these types of networks incorporate important characteristics such as growth and



preferential attachment [5, 6] in contrast to random networks on which all the vertices have equal probability of connecting to each other.

Among the tools that can be used to characterize the properties of networks are Cellular Automata (CA). Originally studied as models of growth and self-reproducing processes, CAs have become a relevant tool for the study of spatio-temporal pattern formation [7]. Besides being used as modeling tools, CAs are also employed in the study of complexity. They can be defined as discrete dynamical systems composed by spatially distributed units over tessellations of the Euclidean space, which are governed by deterministic rules. Regular tessellations are the most common topologies used for modeling CAs, however, real-world topologies are built upon irregular tessellations. Related works have shown that CAs can be combined with networks in order to analyze topological properties of complex systems under the perspective of pattern formation. Previous works on small-world networks were performed by Watts and Tomassini [8, 9]. Both authors discuss the global computing capacities of CAs as the density classification problem in small-world topologies. Marr & Hütt studied the dynamics of evolving networks through the use of CAs [10, 11]. Their results indicate a strong association between entropy measures obtained from the spatio-temporal patterns and the degree distribution of a network.

In this study, we propose a method that performs the classification of theoretical network models based on spatio-temporal patterns evolved by a cellular automaton. The methodology consists in the modeling of a binary totalistic CA over the topology of the network and in the definition of the CA dynamics based on the density of living neighbors (cells that are in the *on* state). Then, quantitative measurements are obtained through the evolved patterns. Specifically, we used the distribution of the Shannon entropy to compose the feature vector of each network. The dataset used in the experiments performed in this paper is composed by four different types of networks: random, small-world, scale-free and geographical. We used two parameters in the evaluation of the classification accuracy: the initial number of living neighbors, σ , and the threshold parameter, μ , related to the density of living neighbors. It was obtained a maximum accuracy of 95.56% which demonstrates the feasibility of the proposed method.

2. Proposed Methodology

A cellular automaton can be described by the quintuple $\mathcal{C} = \langle \mathcal{T}, \mathbb{S}, s_0, \mathcal{N}, \phi \rangle$, where, \mathcal{T} is the tessellation of the automaton, \mathbb{S} is the set of states, s_0 is the initial configuration of the states in $t = 0$, \mathcal{N} is the neighborhood function, and, finally, ϕ is the transition function that governs the dynamics of the automaton. The first step of the proposed method consists in the modeling of the CA over the network topology. Therefore, the network becomes the tessellation \mathcal{T} of the automaton and each network node corresponds to one CA cell. We assumed a binary state space \mathbb{S} and s_0 is controlled by the parameter σ , which is a threshold that defines the percentage of initial alive nodes.

Since the network is an irregular tessellation, the number of neighbors of each CA cell is restricted to the neighborhood of each network node. Therefore, function \mathcal{N} maps each node to its degree, which is defined as: $k_i = \sum_{j=1}^N A_{ij}$, where, N is the total number of nodes and A is the adjacency matrix. $A_{ij} = 1$, if i is connected to j and $A_{ij} = 0$, otherwise. Finally, the transition function of the referred network CA is given as a function of the neighborhood density, ρ_i , defined as $\rho_i = \frac{1}{k_i} \sum_{j=1}^N A_{ij} s_j(t)$, where $s_j(t)$ is the state of node j in time t . Once we have a binary state space, $s_j(t)$ can assume 0 or 1. Therefore, ρ_i represents the rate of alive neighbors of node i . The transition function ϕ incorporates the dynamic of the nodes according to the following equation [10]

$$x_i(t+1) = \begin{cases} x_i(t), & \rho_i \leq \mu \\ 1 - x_i(t), & \rho_i > \mu, \end{cases} \quad (1)$$

where μ is a threshold for ρ_i under which the node maintains its state in the next time step, otherwise, the node inverts its state. Fig. 1-a) illustrates the evolution of a given network with $\mu = 0.5$ using Eq. 1. The first network ($t = 0$) represents the initial configuration of states for each node where black cells represent the “dead” nodes, while the white cells, the “alive” nodes, *i.e.*, $s_i = 0$ and $s_i = 1$, respectively. Moreover, the subsequent figures show the configuration of the state for each node during three time steps.

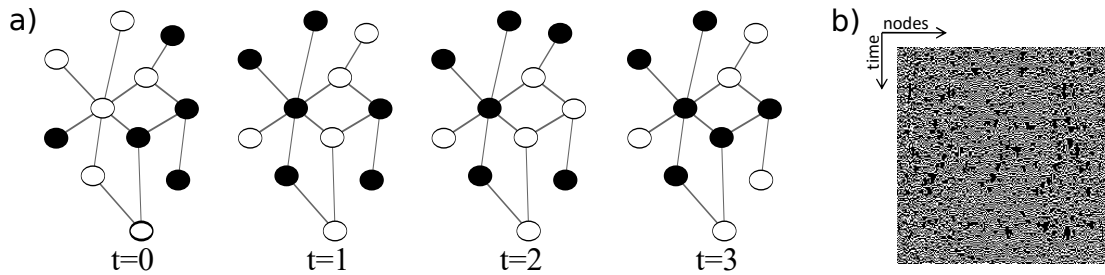


Figure 1. Evolution dynamics. a) Illustration of the configurations of the states of each node for a sample network using Eq. 1 as transition function. b) Space-time diagram for a small-world network with 200 nodes. The CA was evolved by 200 time steps.

The evolved pattern obtained by applying the transition function can be represented as a spatio-temporal diagram on which each column represents a network node and each row represents the states of all nodes at each time step. Fig. 1-b) illustrates an example for a small-world network. The nodes were ordered by k_i , therefore, the left-most nodes are the ones with the smallest values of k_i , and, the right-most, corresponds to the ones with the highest values of k_i . Further, we evaluated the evolution of each node as a time series containing zeros and ones. Then, the Shannon entropy was calculated for each node according to the following equation: $S_i = -(p_i^0 \log_2 p_i^0 + p_i^1 \log_2 p_i^1)$, where p_i^0 is the probability of finding zeros in the time series and p_i^1 is the probability of ones. The Shannon entropy quantifies how heterogeneous is the series and is normalized between $[0, 1]$. Finally, given the values of S_i for each node, we calculated the distribution of the Shannon entropy, Φ_S , for the whole network. Therefore, the vector Φ_S represents the network descriptor that characterize a network model.

3. Results and Discussion

In previous works, Marr&Hütt [12, 10] have shown that for different intervals of μ parameter, different spatio-temporal patterns can be obtained for a specific network topology. For low values of μ , oscillating patterns tend to prevail since the majority of the nodes will change their states at each time step. In contrast, for higher values of μ , the majority of the nodes will not change their states, and, consequently, stable patterns will prevail. Finally, for intermediary values of μ , chaotic and complex behavior tend to emerge. In this paper, we analyzed four different theoretical network models. Fig. 2-a) shows the variation of the average Shannon entropy (calculated for each network model by iterating over the Shannon entropy of each network node), for different values of μ . We can observe that the transition from oscillating patterns (high Shannon entropy) to stable patterns (low Shannon entropy) occurs given distinct values of μ for each network model.

Based on this result, we evaluated the accuracy of network model classification as a function of the threshold parameter μ as well as the distribution of the alive and dead populations given by σ . We created a dataset composed of synthetic networks generated according to the following theoretical models: 1) random, with connection probability of $p = \langle k \rangle / n$; 2) small-world, with rewiring probability of $p = 0.1$; 3) scale-free, with power exponent $\gamma = 1$, and, 4) geographical,

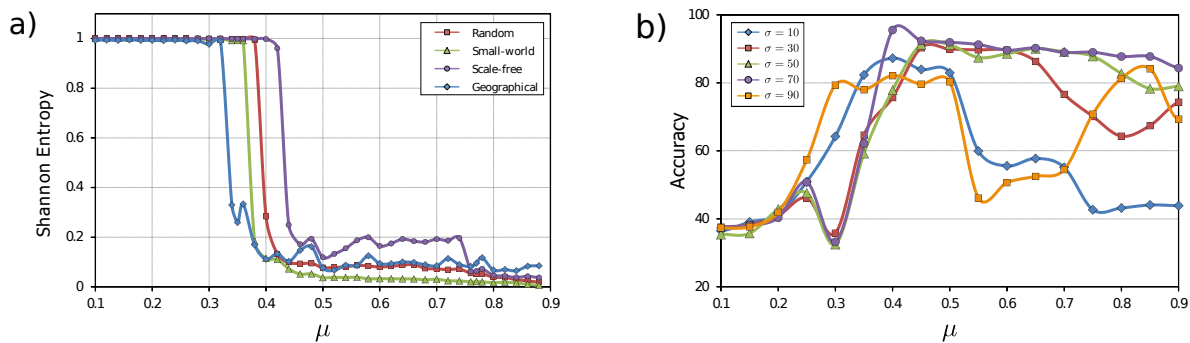


Figure 2. a) Shannon Entropy as a function of μ parameter for the four network models. b) Accuracy values (%) as a function of μ parameter for distinct initial conditions (σ parameter).

on which the Cartesian coordinates of the nodes are predefined. This dataset contains 100 networks of each model with mean degree $\langle k \rangle = 8$. Each network model represents a class in this experiment. By applying the proposed methodology, we obtained the Shannon entropy distribution, for each network, which were used as the network descriptor Φ_S . We have employed 10-fold cross-validation scheme and SVM (*Support Vector Machines*) classifier. In this context, accuracy values represent the number of correct classified instances. Fig. 2-b) presents the results for five different values of σ : 10, 30, 50, 70 and 90. Each curve of this figure represents a σ value. The x -axis corresponds to μ parameter and the accuracy obtained is shown on the y -axis.

We can see from Fig. 2-b) that the highest values of accuracy were obtained for intermediary values of μ and the influence of σ on accuracy becomes more apparent from $\mu = 0.4$. The highest accuracy obtained in the classifications of the correct network models was 95.56% for $\mu = 0.4$.

4. Conclusion

We have presented a method that performs network model classification based on the spatio-temporal pattern of a one parameter totalistic cellular automaton. This method achieved very good accuracy results and can be an alternative for network characterization.

Acknowledgements

G.H.B. Miranda and O.M. Bruno are grateful for the support from São Paulo Research Foundation (FAPESP), with grant 2015/05899-7. G.H.B. Miranda and J.Machicao are grateful for the support of Coordination for the Improvement of Higher Education Personnel (CAPES).

References

- [1] Strogatz S H 2001 *Nature* **410** 268–276
- [2] Albert R and Barabási A L 2002 *Reviews of modern physics* **74** 47
- [3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 *Physics reports* **424** 175–308
- [4] Fortunato S 2010 *Physics Reports* **486** 75–174
- [5] Barabási A L and Albert R 1999 *science* **286** 509–512
- [6] Dorogovtsev S N and Mendes J F 2002 *Advances in physics* **51** 1079–1187
- [7] Wolfram S 1984 *Nature* **311** 419–424
- [8] Watts D J 1999 *Small worlds: the dynamics of networks between order and randomness* (Princeton university press)
- [9] Tomassini M, Giacobini M and Darabos C 2005 *Complex Systems* **15** 261–284
- [10] Marr C and Hütt M T 2009 *Physics Letters A* **373** 546–549
- [11] Marr C and Hütt M T 2012 *Entropy* **14** 993–1010
- [12] Marr C and Hütt M T 2005 *Physica A: Statistical Mechanics and its Applications* **354** 641–662