Complexity and Structural Properties in Scale-free Networks

Yesid Madrid[1,2], Carlos Gershenson[3,4] and Nelson Fernández[1]

[1] Laboratorio de Investigaciones en Hidroinformática, Universidad de Pamplona, 543050 Pamplona, Colombia.
[2] Grupo de Investigación en Ciencias Computacionales-CICOM, Universidad de Pamplona, 543050 Pamplona, Colombia.
[3] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México
[4] Massachusets Institute of Technology, Cambridge, MA 02139, USA
nfernandez@unipamplona.edu.co

Abstract

We apply formal information measures of emergence, self-organization and complexity to scale-free random networks, to explore their association with structural indicators of network topology. Results show that the cumulative number of nodes and edges coincides with an increment of the self-organization and relative complexity, and a loss of the emergence and complexity. Our approach shows a complementary way of studying networks in terms of information.

Introduction

Among representative structural properties of networks we can list: the degree of nodes and their distribution, the clustering coefficient, and the average path length. The degree is informative of how many nodes are connected to each other. The clustering coefficient is a measure of the number of triangles in a graph. The average path length is the average number of steps along the shortest paths between all possible pairs of network nodes (Newman et al., 2006). In spite of the value of these measures to characterize some complex networks, measuring complexity in networks is desirable. Recently, measures of emergence, self-organization, complexity, and relative complexity based on information theory have been developed and their usefulness can be evaluated (Fernández et al., 2014).

In this paper we analyze the association of topological structural indicators like the number of nodes, clustering coefficient and average path length with formal measures of emergence, self-organization, complexity, relative complexity to scale-free random networks.

In the next section we present the methods for generating networks and the formalism to measure complexity. In section 3, we briefly present and discuss our results obtained from the applications of multivariated machine learning unsupervised techniques. Section 4 presents conclusions and future work.

Methods

Using the Barabasi-Albert model implemented in SocNetV software (Kalamaras D., 2015), ten random networks of the following number of nodes were generated: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000 and 3000. In total, 310 scale-free networks were created. First, structural properties of clustering coefficient and average path length were calculated. Then, information measures of emergence (E), self-organization (S), complexity (C), and relative complexity (R) were applied to the vector obtained from the horizontal sum of the grade of each node, on the adjacency matrix. Summarizing, E is equivalent to Shannon information (Shannon, 1948), depending of the probabilities p_i for all i symbols in a finite alphabet: $I = -\sum_{i=1}^{n} p_i \log p_i$. In this work, we use $\log_{10}$. Based on this equation we define that $E = I$, $S = 1 - E$, and $C = 4 \times E \times S$. Relative Complexity $R = CN_i/CN_T$ where CN_i was the complexity of the network i and CN_T was the average complexity of the other networks. Since $E, S, C \in [0, 1]$ a numerical, color and category scale has been defined for a better interpretation. The ranges are $[0.8, 1]$, $[0.6, 0.8)$, $[0.4, 0.6)$, $[0.2, 0.4)$,$[0, 0.2)$. The corresponding colors are: blue, green, yellow, orange, and red. The matching categories are: very high, high, fair, low, and very low. R just has two colors of codification: blue if$R > 1$ and red if $R < 1$.

To facilitate the visualization of the relationship of complexity properties and network structure descriptors, according to the increment of nodes and edges, multivariated techniques was carried out. Consequently, a principal component analysis (*PCA*) was integrated with a hierarchical cluster analysis (*HCPC*). *PCA* was used to summarize and to visualize the information contained in structural attributes and complexity properties. *HCPC* was used for identifying clusters of networks with similar characteristics. Also, statistical indicators such as *v-test* (a criterion of normal distribution), mean in cluster, overal average, and p-values were estimated to associated clusters and properties (Le and Worch, 2015).

Results and Discussion

The PCA depicts the relationship, variation, and patterns among structural and complexity properties (fig. 1). A high percentage of the explained variance of data was captured in the first two axes (91.46%). In consideration of the po-

DOI: http://dx.doi.org/10.7551/ 978-0-262-33936-0-ch116

sition of variables in the multidimensional space, it is easy to see that self-organization and relative complexity are positively correlated with the increasing of nodes and edges. Meanwhile, when nodes, edges, self-organization, and relative complexity increase, the complexity and emergence of the network decreases. These two groups, in spite of the fact that they have an opposite behavior, are very close to the first component (Dim 1) and explain the variance of scale-free networks in 74.24%. As a relevant fact, it is possible to see that complexity is more related to the change (emergence) in the system than its regularity (self-organization). This suggest that some adaptability of scale-free networks could be related to high variability with a minor proportion of uniformity in the degree distribution of the nodes.

Considering the structural indicators of clustering coefficient and average path length, we can observe that they are opposite as it has been noticed in the literature. They are associated with the second component and represent a minor variance explained of the dataset (12.22%). As the form of clustering coefficient and average path length are positioned, we cannot establish any relation between them and emergence, self-organization, complexity, and relative complexity.

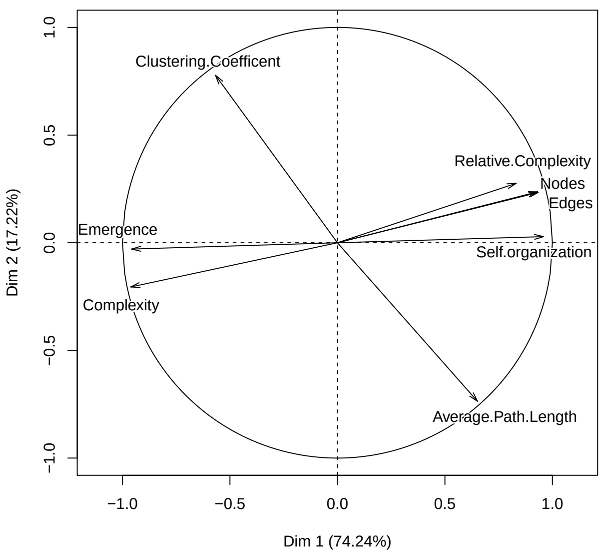

Figure 1: Principal component analysis of complexity and structural properties in 310 simulations of random scale-free networks obtained using BA model

Regarding the incremental number of nodes, *HCPC* analysis allows the statistical creation and characterization of five clusters (table 1). The first cluster included networks with just five nodes. These systems are related statistically with the clustering coefficient, because of a relatively high density of ties in small networks. Cluster two groups networks with 10 to 50 nodes, indicating that they are the most emergent and complex of all. Indeed, in cluster two, E reached the fair category (yellow), in comparison with the low category of overall (in orange). C was classified as very high (blue), meanwhile overall was categorized as high (green). Besides, we can note that these small networks have a negative relationship with the average path length which is less than overall. Networks between 60-175 nodes were grouped in cluster 3. They also have a very high complexity. Cluster four included networks with 100-600 nodes and are the most self-organized due to having a very high S. Finally, networks with the highest number of nodes (700-3000) gain some relative complexity. That means the increase of the number of nodes and edges resulting in networks moderately more complex.

Table 1: Statistical description for Clusters and Structural Properties in Scale-free Networks.

Cluster	Number of Nodes in Network Grouped	Property Associated	V.Test	Property μ in cluster	Property Overall μ	p-value/Significance
1	5	C.C.	4.808	0.213	0.019	1.521^{06}***
2	10,15,20,25,30 35,40,45,50	E	4.316	0.545	0.304	$1.586e^{05}$***
		C	3.177	0.949	0.676	1.586^{-05}***
		Av.P.Length.	-3.165	2.283	2.619	1.553^{-03}
3	60,70,80,90,100 125,150,175	C	2.116	0.873	0.676	0.034*
4	200,225,250,300,350 400,450,500,600	S	2.928	0.872	0.045	0.003**
5	700,800,900 1000,2000,3000	R	4.358	1.007	1.003	1.308e-05***
		S	3.082	0.945	0.696	2.049^{03}***

Final Remarks

Our first results are encouraging. It was interesting to find that growth in random scale-free graphs implies more self-organization and relative complexity. The relative complexity could be useful to analyze cases when two or more networks interact. Thus, the gain of self-organization and relative complexity could in time be a useful characteristic to regulate feedback and guide the management of complex networks.

Further work is required. We are planning to broaden our explorations and perform further analysis to understand and clarify the relationship between complexity properties and structural indicators in random networks, and other complex topologies.

References

Fernández, N., Maldonado, C., and Gershenson, C. (2014). Information measures of complexity, emergence, self-organization, homeostasis, and autopoiesis. In Prokopenko, M., editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 19–51. Springer, Berlin Heidelberg.

Kalamaras D. (2015). Social Network Visualizer (SocNetV). Social network analysis and visualization software.

Le, S. and Worch, T. (2015). *Analyzing Sensory Data with R*. Chapman and Hall/CRC, Boca Raton, FL 33487-2742.

Newman, M., Barabasi, A.-L., and Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3 and 4):379–423 and 623–656.

DOI: http://dx.doi.org/10.7551/ 978-0-262-33936-0-ch116