
The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2015

Volume 2, Pages 57-62

ICEMST 2015: International Conference on Education in Mathematics, Science & Technology

A COMPUTER SOFTWARE FOR THE EDUCATION OF COMPLEX NETWORK ANALYSIS

İlker TÜRKER

Karabuk University, Department of Computer Engineering, Karabuk-TURKEY

Alper Talha KARADENİZ

Ordu University, Mesudiye Vocational School, Ordu-TURKEY

Serhat Orkun TAN

Karabuk University, Vocational School, Karabuk-TURKEY

ABSTRACT: Complex network analysis is an attractive tool for capturing the self-organizing principles underlying the social, physical or biological communities. Several software are developed for either analyzing or generating complex networks, including the visualization utilities. We developed an open source software in Microsoft .NET platform for generating networks based on the most common models as Barabasi-Albert, Erdos-Renyi, Watts-Strogatz including the analyzing utilities defining the network like average separation, degree distribution, clustering coefficient etc. In contrast with the well-known software, this software aims to contribute the understanding of the underlying mechanisms of complex networks. It also forms a basis to further developments that should provide an extensive view to network construction. As an open source software, the opportunity of editing the core functions about network dynamics offer a pedagogical approach to learn more about self-organizing networks.

Key words: complex networks, software, small world, scale free networks

INTRODUCTION

Complex networks are conceptually used to define the dynamic systems in nature and society. These structures are observed in a variety spanning social, biologic, ecologic, transportation, computer, scientific collaboration or citation networks (Albert and Barabási, 2002). A network can be described by a set of nodes (vertices) and links (edges) which can be displayed by an $N \times N$ matrix where N denotes the number of vertices (Newman, 2001).

Understanding the structure of complex networks is primarily significant for understanding how knowledge, disease, culture, viruses etc. spread in the complex systems (Perc, 2010). The evolution of complex systems was traditionally assumed to be driven by randomly wiring processes which result a so-called random network. But recent studies in the past two decades show that these systems yield some self-organizing principles that are different from the random networks (Albert and Barabási, 2002).

In fact, these organizing principles are the main facts that result the network topology and dynamics, which in turn effects how knowledge or viruses diffuse in that network. Thus, capturing these principles is the main goal of complex network analysis to form a basis of how network modeling should be done programmatically. The wide corpus of scientific papers subjecting complex network analysis by the beginning of this century handles this issue, each stating out the generic organizing principles of specific networks in nature and society (Perc, 2010).

On behalf of the above mentioned part of the network science namely “complex network analysis”, another area of interest grows scoping the modeling counterpart. Employing the output supplied by the first part, modeling networks aims to capture the main mechanisms that affect the evolution of the network, providing a broad range of experiments with several organizing principles along with tunable parameters (Barabasi et.al, 2002).

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: İlker TÜRKER- icemstoffice@gmail.com

Generic principles of complex networks

For modeling a network, the generic principles of real networks should be determined as the ingredients of the algorithm. The first property that a real network should hold is the “small-world” phenomenon. The most popular manifestation of small worlds is the “six degrees of separation” concept, uncovered by the social psychologist Stanley Milgram (1967), who concluded that there was a path of social connections with a typical length of about six between most pairs of people in the United States (Kochen, 1989).

Small-world property is observed in many real networks like www (Albert et.al, 1999), online social networks (Leskovec & Horvitz, 2008), scientific collaboration networks (Barabasi et.al, 2002; Newman, 2001b; Newman, 2001c; Perc, 2010, Cavusoglu & Turker, 2013, 2014), movie actor networks (Amaral et.al, 2000) etc. Barabasi explains being small-world as finding relatively small paths between two randomly chosen nodes, while this phenomenon is valid for most of the node pairs in that network. A characteristic measurement of node distances is “average separation” that stands for the average value of the distances between all node pairs in a network.

Another principal ingredient of real world networks is the scale-free property. A large variety of results of real network analysis demonstrate that many networks are scale free, that is, their degree distribution follows a power law for large k . That means, the probability of having degree k for a network follows the equation $P(k) = k^{-\gamma}$.

This distribution can be validated by drawing the degree distribution in a log-log scale, resulting a straight line having negative slope (Clauset et.al, 2009). The generic mechanism underlying this property is “preferential attachment” that means that a new node connecting the network, connecting to more popular (i.e. having more connections) links displays higher probability than connecting to the less popular ones (Barabasi and Albert, 1999; Albert et.al, 2002).

Scale-free property promotes the emergence of a little portion of nodes with high degrees (connections), that can be named as supernodes or hubs. In such a network, if a node gets more popular in the beginning of the network construction, these “first-mover advantage” causes it to have more and more connections later. This fact is known as the Matthew “rich get richer” effect, promoting the occurrence of a small number of popular nodes, while the new connecting nodes or some mid-life nodes of the network have smaller degrees of connections. The above given relation of power-law degree distribution is a result of this mechanism, that can be observed in most of the real networks.

The third important network parameter that measures network clustering and describes symmetry of interaction among trios of actors is the clustering coefficient. It shows the probability of a node’s neighbors to have connections among each other, excluding the links coming from the starting (or center) node. Topologically, it shows the density of the triangles in a network, a triangle being formed when two of one’s neighbors connect with each other (Newman, 2004; Çavuşoğlu & Türker, 2013).

Clustering coefficient gets values in the interval of 0 to 1, where the values close to 1 indicate dense connections between neighbors. Averaging this parameter is averaged over the network, average clustering coefficient can be found to have an idea about the network’s interconnectedness. Real networks display high clustering coefficients compared to random networks, i.e. your followers in a social network follow each other in a high rate, representing a clique of friendships.

Most common network models

Erdos-Renyi (ER) Model

In their classic first article on random graphs, Erdos and Renyi define a random graph with two parameters as N (number of nodes) and p (probability of connecting), as N labeled nodes connected by n edges, which are randomly chosen from the $N(N-1)/2$ possible edges. Programmatically, it can be explained as starting with N nodes, find the number of links by the formula $pN(N-1)/2$ and wire the N nodes with n links calculated by the above formula, as seen in Fig.1 (Erdos & Renyi, 1959; Albert and Barabasi, 2002).

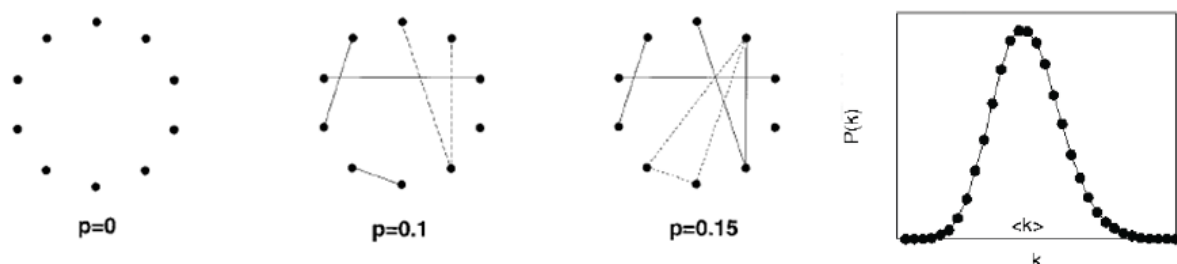


Figure 1. Random graphs generated with different p values. The right side plot is the degree distribution (Albert and Barabasi, 2002)

The expected degree distribution of a random graph is “binomial distribution” which converges to a “Poisson distribution” with large N and small p , demonstrating node degrees having closer values to an average degree and not deviating more than a few percents from that average value. Since the links are generated randomly, the relations between a node’s neighbors are not strong as real networks, resulting a very small average clustering coefficient for the network. on the other hand, long range random links provide short paths between distant nodes, resulting a relatively small average separation (distances) between nodes.

The network parameters mentioned above are not in good consistency with real networks since real networks do not display poisson-like degree distributions and have considerably higher clustering coefficients that random networks have. The only common-point between random networks and real networks is the short average path length between nodes.

Watts-Strogatz (WS) Model

Above mentioned disparity in the topologic properties of random and real networks pioneered the studies of more realistic modeling of real networks. In 1998, Watts and Strogatz proposed a model interpolating between a regular lattice and a random graph (Watts and Strogatz, 1998; Albert and Barabasi, 2002). Their model starts with constructing a regular lattice. Then the only tunable parameter p is used as a probability to decide if an edge (link) will be rewired, preserving the source node and altering the target node with a new one in a random process. If the p parameter converges to 0, the network stays regular, while it gets a completely random one as p converges to 1. For some moderate values of p (for ex. $p=0.01$), Watts and Strogatz showed that there is a regime that the network displays high clustering coefficient and low average distance (separation) as if in the real networks. This is the small-world regime of the network, capturing the similarity with the real network by the means of clustering coefficient and average separation (Fig.2).

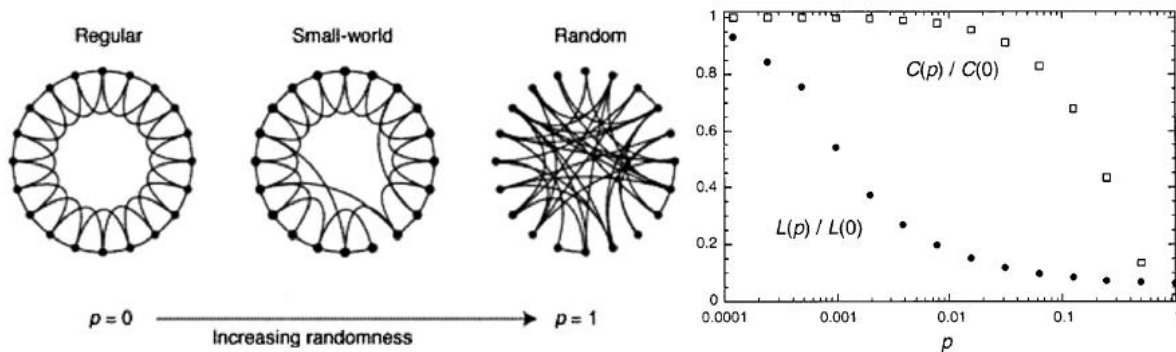


Figure 2. Watts-Strogatz (1998) network structure with varying p values. The right side plot demonstrates the deviation of clustering coefficient and average separation with increasing p values.

The limitation of the WS network is the lack of capturing the degree distribution of a real network. It produces a network having a similar degree distribution like Erdos-Renyi type network, having the advantage of adding small-world property to the structure.

Barabasi-Albert (BA) Model

The model proposed by Barabasi and Albert (1999) was the first in capturing the power-law degree distribution observed in most of the real networks. They suggested that the organizing principles of real networks should be imitated to maintain the generic scale-free property. Thus, a network grows continuously by the addition of new nodes, and the new nodes likely prefer to connect to the nodes of higher degrees (i.e. popular nodes) rather than the ones with lower degrees. The former property is known as growing, while the latter property is known as preferential attachment.

They modeled growing property by starting with a small number (m_0) of nodes and add a new node with $m(\leq m_0)$ edges (links) that connect the new node to m different nodes already present in the system, at each time step.

To define the connectivity function including preferential attachment, Barabasi and Albert used Eq.1 as the probability for a new node to connect to node i . As seen in Eq.1, the node having higher degree (connections) has a higher attractiveness to have connection with a new node.

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (1)$$

Successfully capturing the organizing principles of real networks, BA model provides a perfect power-law degree distribution together with small-world properties as if in the WS network as in Fig.3 (Albert and Barabasi, 1999). In this perspective, it forms a basis for realistic modeling of networks with the opportunities of adding some variations for capturing the alternations from perfect power-law observed in real networks.

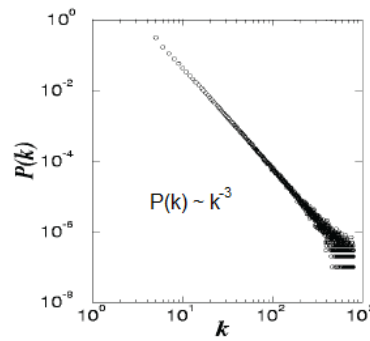


Figure 3. Power-law degree distribution of a Barabasi Albert network model (Barabasi and Albert, 1999).

METHODS

We developed a software that generates networks of ER, WS and BA models (Fig.4). The inspiration of the development depend on both supplying a pedagogical view on the understanding of complex networks in the post-graduate education, and also to form a basis for the further studies in network modeling giving the opportunity of editing the core functions about network dynamics.

Network parameters are the output of the organizing principles that take part in network construction. The tendencies of node selection of the current nodes for making new connections are the main fact that drives the resulting parameters or distributions of that network. By this view, tuning the input parameters or the opportunity of editing the core functions of node selection takes a significant part in the understanding of how networks grow and organize.

Another pedagogical output of this software is the visualization of the network constructed, whereas the main output parameters like degree distribution, average separation, clustering coefficient etc. are also supplied.

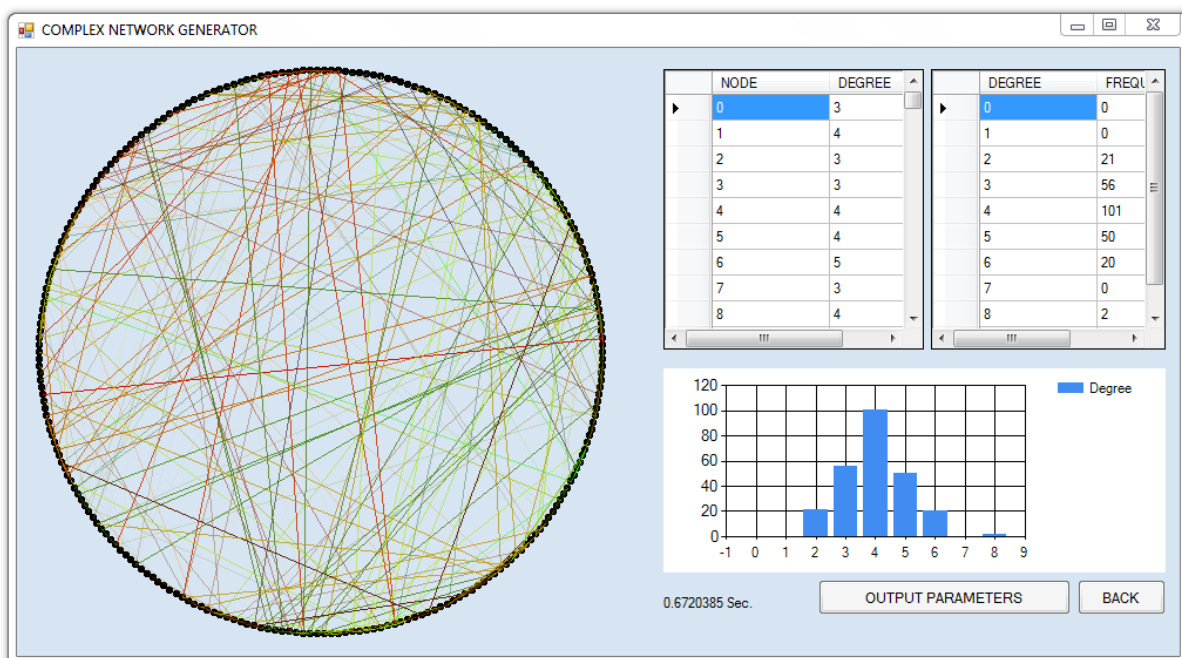


Figure 4. The user interface of the software.

The software is developed in the Microsoft .NET platform using C# language and standard form controls. While executing the network generation or calculating the output parameters, the most readable code and algorithms were used in order to enhance the understanding of complex networks. That is, fast execution is sacrificed in some functions in order to increase readability.

To enable further plotting opportunities in Matlab, R, etc., the node degrees are displayed in a datagrid control together with the node frequencies that are used to plot degree distribution. This is a necessity not only for generating graphics in semi-log or log-log plots, but also curve fitting to test if a degree distribution follows exponential, power-law, binomial, Poisson or some mixed variations of these functions.

The software is ready-to-use for the systems having .NET Framework 4.0 or later, and can be downloaded via <http://www.ilkerturker.com/cn/nwmodel/>.

RESULTS AND FINDINGS

The three network models mentioned above have different topologies that can be observed in the output parameters. Degree distribution plots as seen in Fig.5 are as consistent with the theoretic expectations. Since the power-law consistency of BA network in linear plot is not obvious, we exported the degree distribution data to Matlab and showed the power-law fitting in log-log scale with the exponent -3.

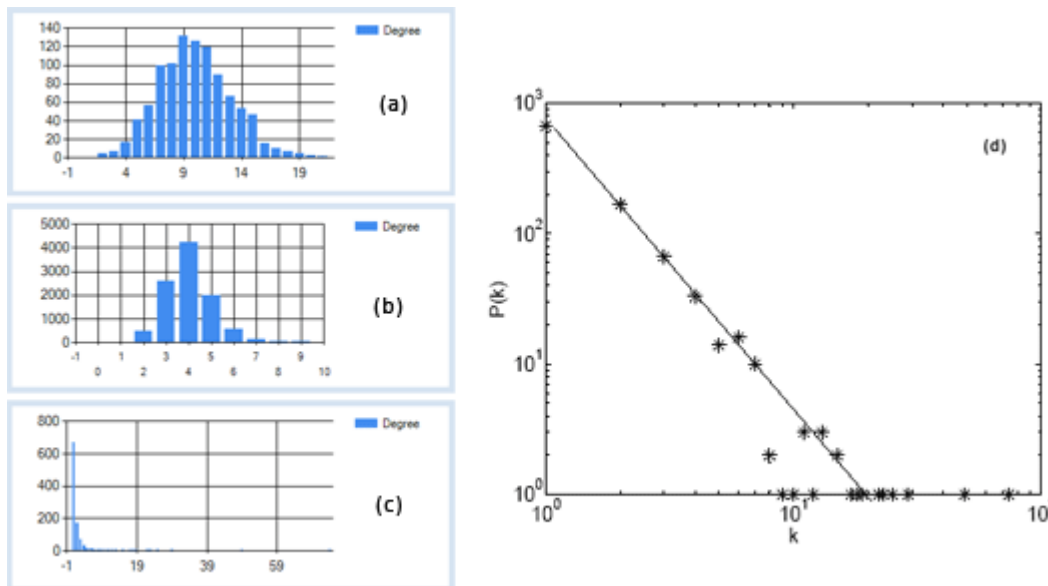


Figure 5. Degree Distribution Plots Of The Three Network Models. (A) ER Model, (B) WS Model, (C) BA Model, (D) BA Model Data Plotted In Log-Log Scale In Matlab To Show Power-Law Consistency.

Similar with the degree distributions, the output parameters (average clustering coefficient, average separation and average degrees) are consistent with the theoretic expectations as well.

CONCLUSION

Both analysis and modeling of complex networks aim to uncover the underlying mechanisms in the self organization processes of complex systems. Getting the analytical feed from the analysis section, the modeling section consists of simulations in generating networks with variable principles and parameters. By this point of view, our software employing basic and robust network models can be an initial point for the researchers who want to make further modeling simulations. The basic output measurements supplied in the software will also provide a rapid start to modeling projects, especially in post graduate studies.

RECOMMENDATIONS

The development process of the software will move along by adding new enhancements in the future, and will be shared in the same URI supplied above. Researchers who want to construct networks of different algorithms can feel free to modify and share the source code.

Especially a challenging area in network modeling is “spreading”. In real world, the network structure plays a significant role on spreading of information, epidemics, opinions and has various impacts on the evolution of science, sociology, health etc. Introducing a realistic spreading model to our software should provide a broad range of experiments on spreading. Also the impact of breaking some kinds of links on spreading is another novel subject to investigate. These opportunities are the forward stage of this software open for all researchers.

REFERENCES

Albert, R., Jeong, H., & Barabasi, A.-L. (1999). The diameter of the world-wide web. *Nature*, 401, 130.

- Albert, R., & Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47.
- Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11149.
- Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509.
- Barabasi, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311, 590.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661.
- Çavuşoğlu, A., Türker, İ., (2013). Scientific collaboration network of Turkey. *Chaos, Solitons & Fractals* 57, 9-18.
- Çavuşoğlu, A., Türker, İ., (2014). Patterns of collaboration in four scientific disciplines of the Turkish collaboration network. *Physica A* 413, 220–229.
- Erdos, P., & Renyi, A., (1959). *Publ. Math. (Debrecen)* 6, 290.
- Kochen, M., 1989, Ed., *The Small World* (Ablex, Norwood, NJ).
- Leskovec, J., & Horvitz, E. (2008). *Proceeding of the 17th international conference on World Wide Web ACM*, New York, (p. 915).
- Milgram, S., 1967, *Psychol. Today* 1, 60.
- Newman, M. E. J. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E*, 64, 025102(R).
- Newman, M. E. J. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64, 016131.
- Newman, M. E. J. (2001c). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64, 016132.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5200.
- Perc, M., (2010). Growth and structure of slovenia's scientific collaboration network. *Journal of Informetrics* 4: 475–482.
- Petersen, A.M., Jung, W-S., Yang, J-S., Stanley, H.E., (2011). Quantitative and empirical demonstration of the Matthew effect in a study of career longevity, *Proc.Natl. Acad. Sci. USA* 108, 18–23.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440.