The System of Generation of Random Networks and Computation of Their Topological Properties

Svetlana, Avetisyan
Yerevan State University
Yerevan, Armenia
e-mail: avetisian@bk.ru

Ani, Kocharyan
Yerevan State University
Yerevan, Armenia
e-mail: kocharyan_ani@yahoo.com

ABSTRACT

The active study of random networks of various nature highlights the issue of developing a system which is capable of simulating the model of random networks of various types and conducting an effective analysis of their topological and statistical properties. The main properties of a random network are node degree distribution, node-node distance distribution, the distribution of the clustering coefficients of the nodes, cycle length distribution, length distribution of the connected subnetworks, network diameter, the spectrum of adjacency matrix's eigenvalues, the distribution of intervals between eigenvalues. The current work describes a system capable of simulating the network behavior of all main classes, i.e. Erdős-Rényi, small-world, scale-free, Block-Hierarchical networks, and conduct a statistical analysis of all the above mentioned topological properties. With all this, the developed system is oriented towards an effective simulation of the Block-Hierarchical networks which represent quite a new feature in the field [3]. For Block-Hierarchical networks new algorithms of calculations are implemented which allow to significantly increase the size of the networks studied. The system ensures a comparatively quick calculation for Block-Hierarchical networks with the node degree of $\approx 10^6$.

Keywords

Random networks, block-hierarchical networks, topological properties, statistical analysis.

1. Introduction

In recent decades there has emerged a new direction where the real complex systems are seen as network structures. Notably, the elements of the real system are presented by nodes (vertices) of the network, and the interaction (relation) between the elements is revealed by the connections (edges) between them. The existence of nodes and multiple connections studied in the networks is not permitted. A network is called random, if any connection within it appears with some degree of probability. A network is called directed if any connection has direction; if not, the network is called non-directed. In modern literature networks are often used for presenting systems with complex architecture, particularly biopolymers (DNA, RNA, proteins), cell metabolism, information transmission networks and communication systems, neuron networks, as well as various evolutionary, ecological, social and economic networks [1]. The study of random networks suggests a statistical research of various topological properties, such as node degree distribution, clustering coefficient, distribution of cycles of various lengths, node-node distance distribution, etc. Part of this information may be extracted from the spectrum of network adjacency matrix's eigenvalues [2].

Recently there have been a number of vast statistical research on the topological properties of real networks of various nature. It turned out that in the majority of cases real networks differed drastically from totally random ones. For many real networks not the exponential, but a more gradual behavior of distributions was revealed to be typical, owing to what those networks came to be known as scale-free. Another class of networks received the name of small-world because of small (as compared to the size of the network) value of the average distance between the nodes. For a number of systems which include multi-resolution and randomness, a hierarchical order was revealed to be characteristic, and random networks which represent such systems were called Block-Hierarchical random networks [3]. Irrespective of the variety of general (global) properties of random networks which allowed to single out several classes of real networks, they, nonetheless, did not provide information about the local topology of the network, which was also vital for many natural networks. As abovementioned in point [4], local topological properties, such as motives and their distributions, allow to have a more thorough classification of random networks. Block-Hierarchical networks, which turned out to be a convenient model of special structure of complicated biomolecules, such as DNA and proteins [3] [5], have become of particular interest lately.

An active examination of random networks of various nature highlights the issue of developing a system which may simulate the models of random networks of various types and may conduct an effective and versatile analysis of their topological and statistical behavior. The current work describes a system which is capable of simulating the behavior of the networks of all main classes, i.e. Erdős-Rényi, small-world, scale-free, Block-Hierarchical networks, and conduct a statistical analysis of all the above mentioned topological properties. In addition, the developed system is oriented towards an effective simulation of Block-Hierarchical networks which represent quite a new feature in the domain [3] [6]. For Block-Hierarchical networks new algorithms of computation of the network behavior have been implemented (see [6]). The implemented algorithms are effective both in terms of time, and the memory they use. *Table 1* shows comparative estimations of the calculation of various topological properties of the Block-Hierarchical network (with a fixed branching index p) with the growth of the number of nodes of the network N. The estimations are given taking into consideration the constancy of the branching index [6]. Estimations show the system ensures a comparatively quick calculation even for $N \approx 10^6$. The system also comprises implementations of other models: Erdős – Rényi, Watts-Strogatz (generating small-world networks), and Baraba´si-Albert (generating scale-free networks). In these cases the system uses standard algorithms, and it is distinguished only by the form it uses to store the network graph. Naturally, all these algorithms demand huge computing resources and are effective for the analysis of networks with 100 edges. To increase the effectiveness, in those cases the system uses distributed computation.

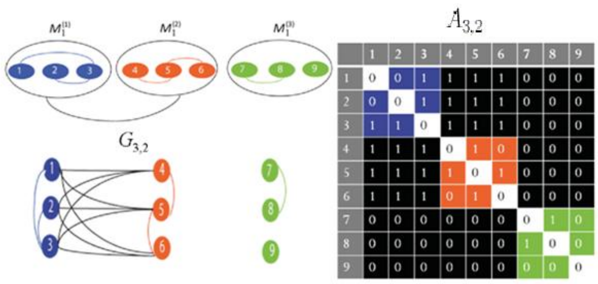
Property	Estimation	Standard Estimation
Node degree	$O(log_p N)$	$O(N)$
Node-node distance	$O(log_p N)$	$O(N^2 * log(N))$
Number of connections	$O(N)$	$O(N^2)$
Number of cycles of length 3	$O(N)$	$O(N^3)$
Number of cycles of length 4	$O(N)$	$O(N^4)$
Number of cycles of length 3, which contain the node x	$O(log_p N)$	$O(N^2)$

Table 1. Estimations of the calculation of topological properties.

2. Random Regularly Branching Block-Hierarchical Network

Block-Hierarchical model. Let p and Γ be natural numbers, $p > 1$. For the given p and Γ a class of regularly branching Block-Hierarchical networks $\mathfrak{R}_{p,\Gamma}$ is defined. The number of the network nodes $G_{p,\Gamma} \in \mathfrak{R}_{p,\Gamma}$ equals p^{Γ}. The network is constructed by levels. On every new level $\gamma,\ 0 \le \gamma \le \Gamma$ new clusters (subnets) are formed via merging the clusters formed on the previous level, and by introducing new connections between them via joining some of them [6]. Let us define a set of clusters of γ level by M_γ, and the i-th cluster of γ level by $M_\gamma^{(i)}$, where $1 \le i \le n_\gamma,\ n_\gamma = p^{\Gamma-\gamma}$,

$$M_\gamma = \{M_\gamma^{(1)},\ M_\gamma^{(2)},\ \ldots, M_\gamma^{(n_\gamma)}\}.$$



Fugure 1. Block-Hierarchichal network $G_{3,2}$ and its adjacency matrix $A_{3,2}$. The cluster $M_1^{(1)}$ is joined with the cluster $M_1^{(2)}$. The network $G_{3,2}$ is presented by cluster $M_2^{(1)}$ which comprises the inner clusters $M_1^{(1)},\ M_1^{(2)},\ M_1^{(3)}$.

To store such networks a marked tree (connectivity tree) is used, the leaves of which are the nodes of the network, and the subtrees correspond to clusters (groups of nodes) and represent the connections between the nodes of the cluster. All the algorithms of calculation of statistical properties of the Block-Hierarchical network model use the connectivity tree.

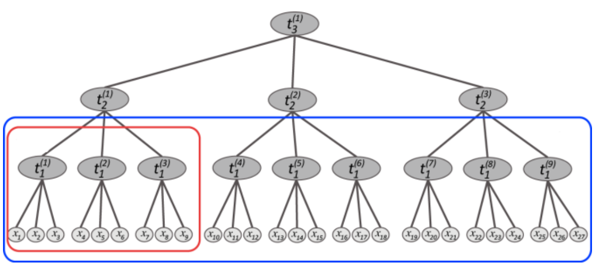


Figure 2. $G_{3,3}$ network's connectivity tree.

Thus, to determine the connections in the network it is sufficient to break the network into clusters hierarchically inserted one into another, by using values p and Γ. The model of random Block-Hierarchical network is defined by the probability of the occurrence of the connections which varies from level to level and is given in the following way: $q_\gamma = p^{-\mu\gamma}$, where μ is the parameter of the model (network density) [3],and γ is the cluster level for which the connections are formed. By introducing a hierarchical structure of connections to the network and a set of probabilities q_γ, network models with the given hierarchical structure may be built, and their topological properties may be examined.

The statistical analysis of topological properties of the networks studied is conducted on an ensemble of networks with fixed parameter values p, Γ, μ.

3. The System of Generation of Random Networks and Calculation of Their Topological Properties

The developed system is a universal instrument for generating, analyzing random networks of various types and conducting statistical research on various topological properties of the network. It possesses sufficient flexibility, independence of functioning from specific models and the possibility to include new models.

The system allows:

1. To generate an ensemble of networks (realizations) of a specific model. The system comprises the models Erdős – Rényi (random graphs), Watts-Strogatz (small-world networks), Baraba´si-Albert (scale-free networks), Block-Hierarchical;
2. To calculate topological properties for each realization in the ensemble;
3. To store results in the database;
4. To conduct a statistical analysis of the networks.

3.1. Generation of an Ensemble of Random Networks

The ensemble of realizations is defined by the number of network nodes and the probability of connections between the nodes. All the realizations within the ensemble are not interdependent. Let us consider the process of creation of an ensemble of realizations for each model.

The model Erdős – Rényi is defined by the number of N nodes and the probability of existence of a connection between the nodes q. The networks realized with the same parameters N, q form the given ensemble.

Watts-Strogatz model is defined by the number of N nodes, number of edges $N * k/2$ (k is an even natural number) and with the probability of randomization (rewiring) of q edges. The network generation starts with the ring lattice with N nodes, each of which is connected with its k neighbors ($k/2$ connections from each side). At every given moment of time the connections randomize with the probability q with self-connection and link duplication not being allowed when rewiring [1]. The generation of the ensemble realization starts with the same network which is identically defined with the parameters N and k. For each realization s steps of connection randomization are carried out with the probability of q. The networks realized with the same parameters N, k, q, s form the given ensemble.

Baraba´si-Albert model is defined by two mechanisms, growth and preferential attachment [1]. The generation of the ensemble realization starts with the same random network with m_0 vertices and connection probability q. For each realization s steps are carried out. On every step a new vertex is added which is joined with the already existing network vertices with m edges. The probability q_i that the new vertex will turn out to be connected with the vertex i via an edge depends on its level k_i : $q_i = k_i / \sum_j k_j$ where the summation is carried out at all the network vertices (preferential attachment). As a result, all the realizations in the ensemble have an identical number of vertices $N = m_0 + s$, and the connection probability is defined according to the rules of preferential attachment. The networks realized by the same parameters m_0, m, p, s form the given ensemble.

The Block-Hierarchical model is defined by the branching index p, the level of network Γ and the network density parameter μ. For the given model the number of edges is $N = p^{\Gamma}$, and the probability of connection is defined by the equation $q_\gamma = p^{-\mu\gamma}$ where γ is the cluster level for which connections are formed. The networks realized by the same parameters p, Γ, μ form the given ensemble.

3.2. Calculation of the Topological Properties of the Networks

The analysis of the topological properties is conducted on an ensemble of realizations, that is why for each ensemble of realizations network properties defined for the model are calculated. The main properties of the networks are node degree distribution, node-node distance distribution, the distribution of clustering coefficient of the nodes, cycle length distribution, length distribution of the connected subnetwork, the spectrum of eigenvalues, distance distribution between eigenvalues, network diameter.

When a vast quantity of calculations is required, distributed calculations are used. The realizations of the ensemble are distributed by local networks via computers, and the results of computations are stored in the database. For Block-Hierarchical models such a mode allows to create huge network ensembles with the node number of $\approx 10^6$. For the models Erdős – Rényi, Watts-Strogatz, Baraba´si-Albert standard algorithms of calculations are used which allow to conduct an analysis just for the network ensembles which comprise a few hundred nodes. In that case parallel computations on Linux clusters are provided; a fact which has significantly increased the size of the networks studied.

3.3. The Storage of Results in the Database

All the results of the calculations of topological properties of an ensemble of realizations are stored in the database. The database is implemented in MS SQL, and, in order to process data, stored procedures and base functions are implemented. Statistical analysis is conducted on the information stored in the database. The majority of the results of the statistical analysis are calculated via averaging the results of the ensemble. The averaged values are calculated beforehand and are stored in the database, which significantly decreases the amount of time needed for getting statistical properties. The system has a possibility of storing results in an XML file with a subsequent possibility of their statistical analysis for local work with small networks.

3.4. The Analysis of Topological and Statistical Properties of the Networks

The analysis of topological properties of the network is conducted on the results of calculations on the ensemble of realizations. The possibilities of the analysis of topological properties of a specific realization, of an ensemble of realization, and of all the ensembles with the same parameters of generation are included. When calculating network properties we average the values of the selected realizations. Statistical research includes the analysis of average values which describe the network globally, the analysis of probability distributions (with the possibility of anti-aliasing and approximation), calculations of mathematical expectations and distribution variances. The results of the statistical analysis take the form of graphs and tables of values.

To take an example, for the Block-Hierarchical model degree distributions (*Fig. 3*) and distribution of clustering coefficient (*Fig. 4*) are given. They are the results for an ensemble of 100 realizations. *Table 2* depicts the results of global properties for the Block-Hierarchical model with different parameter μ.

μ	**Number of cycles of order 3**	**Number of cycles of order 4**	**Average clustering coefficient**	**Average node degree**
0.3	$\approx 79 * 10^{12}$	$\approx 8 * 10^{15}$	0.4332	5573
0.5	$\approx 9 * 10^{9}$	$\approx 6 * 10^{15}$	0.4336	1229
0.8	$\approx 15 * 10^{6}$	$\approx 5 * 10^{12}$	0.2574	21

Table 2. Average values for topological properties for the Block-Hierarchical model. For 100 realizations: $p = 3, \Gamma = 11$ ($N = 177\,147$) *are given the results* $\mu = 0.3$, $\mu = 0.5$, $\mu = 0.8$.

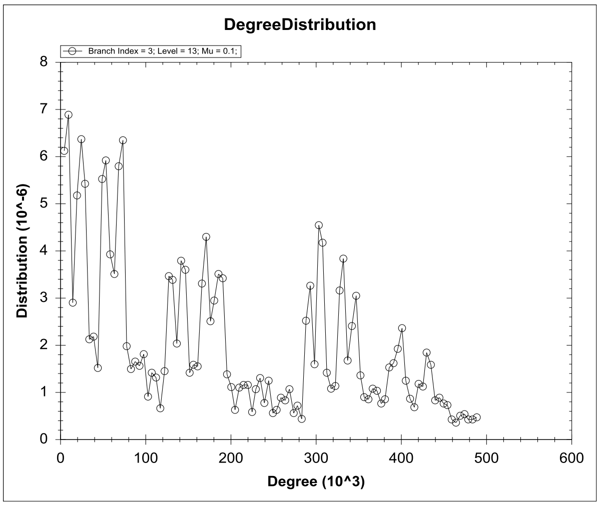


Figure 3. Degree distribution for the Block-Hierarchical model $p = 3, \Gamma = 13, \mu = 0.1$ *(*$N = 1\,594\,323$*). Results of an ensemble of 100 realizations.*

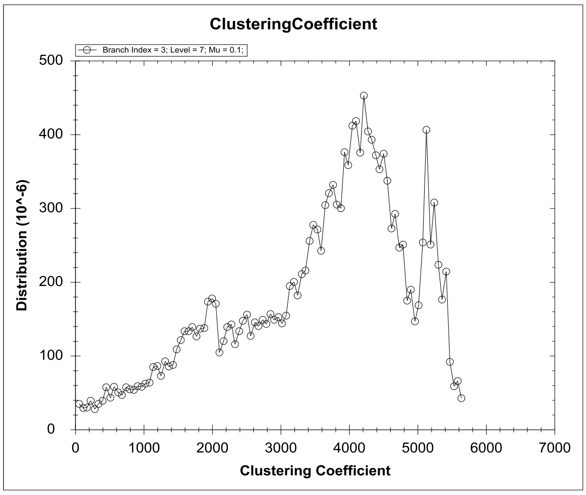


Figure 4. The distribution of clustering coefficient for the Block-Hierarchical model $p = 3, \Gamma = 7, \mu = 0.1$ *(*$N = 2\,187$*). Results of an ensemble of 100 realizations.*

4. ACKNOWLEDGEMENT

We would like to express our gratitude to Dr V.Avetisov for constructive discussions and helpful suggestions.

REFERENCES

[1] *Albert R, Barab´asi A-L* Statistical mechanics of complex networks. Rev. Mod. Phys. 74: 47–97. *(2002)*.

[2] *V.A. Avetisov, A.V. Chertovich, S.K. Nechaev, O.A. Vasilyev*. On spectra of random hierarchical networks. Stat. Mech:Theory and Exper. 07 07008 (2009).

[3] *V.A. Avetisov, A.Kh. Bikulov, O.A. Vasilyev, S.K. Nechaev, A.V. Chertovich*, Some Physical Applications of Random Hierarchical Matrices., JETP, 109(3) 485 (2009).

[4] *R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Science 298 (2002) 824. 38; R. Milo, Sh. Itzkovitz, N. Kashtan, R. Levitt, Sh. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon*, Science 303 (2004) 1538.

[5] *Mirny LA*. The fractal globule as a model of chromatin architecture in the cell. Chromosome Res. 2011 Jan; 19(1):37-51.

[6] *S. Avetisyan, A. Harutyunyan, D. Aslanyan, M. Karapetyan, A. Kocharyan*. Algorithms of computation of the statistical properties of regular block-hierarchical networks. Sixth Annual Scientific conference (5-9 December, 2011). Collected articles. Yerevan. RAU publishing house, 2012 - 312 p (in Russian).