Research of Topological Properties of Network Reflections Obtained Using Different Algorithms for Scanning Initial Networks

Dmytro Lande[1,2][0000-0003-3945-1178], Oleh Dmytrenko[1(✉)] [0000-0001-8501-5313]

[1]Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
dwlande@gmail.com, dmytrenko.o@gmail.com

Abstract. Many modern types of researches of networks use mechanisms for their monitoring, after which conclusions about the topology of such networks are drawn. This paper shows that this approach is wrong.
The reflections of the initial networks obtained as a result of monitoring and partially reflect the properties of these initial networks is often significantly different. The properties of these network reflections significantly depend on the algorithms used for scanning the initial network. To demonstrate this statement, the paper researches the properties of network reflections obtained using three scanning algorithms, which implement the following principles: 1) the transition to the node according to the PageRank algorithm; 2) the transition to the node that has the highest PageRank; 3) the transition to the node that has the highest degree. The networks based on the Erdös-Renyi and Barabási-Albert models are considered as basic.
The paper shows that the values of the characteristics of network reflections, which are close to the topology and characteristics of the initial networks, are achieved only during approaching the scanning steps to the number of nodes in these initial networks.
The obtained results are important in the methodical plan and can be considered as a statement of a problem of finding the algorithm of scanning of a complex network using of which gives most adequately parameters of the initial network (first of all, the degree distribution).

Keywords: Monitoring, Network Scanning, Network Topology, Barabási-Albert Model, Erdös-Renyi Model, PageRank algorithm, Network Density, Degree, Degree Distribution, Power Law.

1 Introduction

Recently, more and more scientific papers have appeared on the study of huge dimension information networks. Not only the content of such networks but also the structure (nodes and connections) can be methodically attributed to the category of

Big Data. In practice, it turns out that it is impossible to cover some network in full by common methods. In such cases, special algorithms are used. Due to the past development of so-called peer-to-peer networks, algorithms such as the Breadth-first Search (BFS) method [1], the Random Breadth-first Search (RBFS) method, the Intelligent Search Mechanism (ISM) method, the Depth-first Search (DFS) [2, 3], the Dijkstra's algorithm [4], the Floyd-Worshell algorithm [5, 6], the Bellman-Ford algorithm [7, 8], the finding connection points and bridges in a graph [9], etc [10].

All these algorithms provide the ability for traversing networks and searching tree or graph data structures (primarily, it was assumed that they will be used to find or integrate targeted content). Using these algorithms, some researchers draw conclusions about the topology of networks, in particular co-authorship networks [11]. The use of probabilistic models of the networks [12-15] makes it possible to analyze the structure of dependencies between the corresponding nodes in this network, to find the probabilities of the existence of certain connections and to obtain estimates of various network characteristics. But it was found that mentioned above models do not always display the real network topology. Therefore, this paper considers and researches how the topology of the network reflection, i.e. the network built using a limited number of scanning steps, depends on the initial network and the scanning algorithm.

2 Random network model

For modelling, as an example, three artefact networks namely, Barabási-Albert [16], Erdös-Renyi [17] and Watts-Strogatz [18] networks are investigated. These random networks models can be considered as prototypes of many real networks.

2.1 Barabási-Albert network: model of preferential connection

Most real artefact networks have a power-law distribution. It turned out that this distribution is due to an effect called the cumulative advantage or preferential attachment. The power-law networks include Barabási-Albert networks (Fig. 1) [16].

To build these networks, a special procedure is used, which consists of the fact that new nodes are gradually added to the initially small number of nodes, links from which are more likely to connect to those nodes that have more links. That is, in the process of network growth, new nodes are more likely to form connections with those nodes that are already characterized by a large number of connections.

It has been proven that it is the “rich getting richer” phenomenon that leads to the emergence of power laws in networks). Obviously, when a new node joins the network, only one link is used, i.e. the number of edges in the network is comparable to the number of nodes, and the network is quasi-hierarchical (the hierarchy can be violated only in the initial composition of nodes).

The degree distribution of the Barabási-Albert model is scale-free, i.e. it obeys the power law

$$P(k) \sim k^{-3} \quad (1)$$

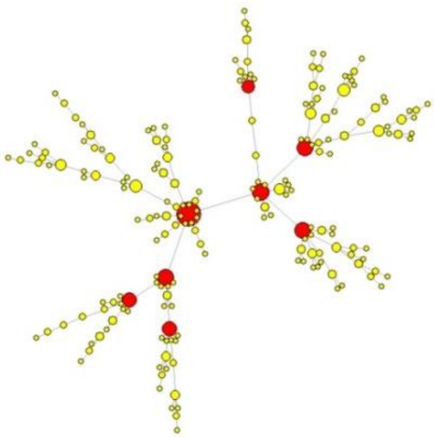

Fig. 1. Barabási-Albert network.

2.2 Erdös-Renyi network

The object that has been intensively reseached in graph theory and is directly related to complex networks is the classical random graph, or Erdős-Renyi random graph. It was proposed and researched in the late 1950s by Paul Erdős and Alfred Rényi [19].

The Erdős-Renyi network can be constructed by randomly distributing *m* connections between *n* nodes. This model is equivalent to a model in which the value of the number of edges *m* is replaced by the corresponding probability *p* of a new edge appearing in the graph (Fig. 2).

It is sometimes called the random graph model or sometimes just the Poisson random graph model because of the Poisson degree distribution for $n \to \infty$ and np=const

$$P(k) \sim {(np)^k e^{-np}}/{k!} \qquad (2)$$

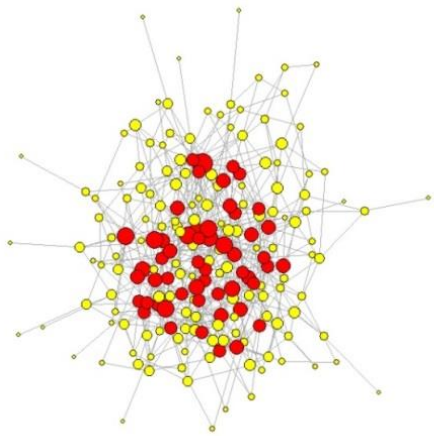

Fig. 2. Erdös-Renyi network.

3 The proposed network scanning algorithms

To build reflections of networks and further research the dependence of the network characteristics of the obtained reflections of networks on the network characteristics of the initial networks besides the different number of scanning steps, three proposed scanning algorithms, which implement the following principles are also used: 1) the transition to the node according to the PageRank algorithm; 2) the transition to the node with the largest value of PageRank; 3) the transition to the node with the largest value of degree

3.1 Building the network reflection using the PageRank algorithm

The PageRank algorithm [20, 21] was proposed in 1996 at Stanford University by Larry Page and Sergey Brin as part of a research project on a new type of information retrieval system. The system uses the PageRank algorithm to measure the importance of website pages and rank them in their search engine results. Taking into account the structure and text of hyperlinks, the PageRank algorithm simulated a random walk of an Internet user starting from a random page. The more random visits to a page, the higher its ranking.

The PageRank for a page A is calculated according to the following rules. Let $T_1,\ldots,T_n$ be the pages that link the page A. The algorithm also uses a damping factor d, the values of which are between 0 and 1, and usually equals 0.85. The function C(T) is equal to the number of links outbound from page T. Then the PageRank of page A, PR(A), is equal to:

$$\mathrm{PR}(A) = (1 - d) \sum_{i=1}^{n} {\mathrm{PR}(T_i)}/{\mathrm{C}(T_i)} \tag{3}$$

In other words, the PageRank are random values, the sum of which for all pages will be equal to 1.

To calculate the PageRank, the Internet space is represented as an oriented graph, the vertices of which correspond to the pages, and the edges correspond to the hyperlink between them. Let n pages be included in the search index. Then a matrix of transitions M of size $n \times n$ is created to model a random walking. The element of this matrix m_{ij}, which is in row i and column j has a value of $1/k$ if the page with number j has k outgoing links, among which there is one that link page with number i. If there is no such outgoing link, the element m_{ij} equals 0.

The probability distribution of finding a random traveler can be described by a column vector whose row j will be equal to the probability of being on the page j [22]. This vector corresponds to the simplest and idealized variant of the PageRank.

To build a reflection of the initial network according to the described above PageRank algorithm, you need to make a traversing along the path, which is determined by a random walk across the initial network. The number of network scanning steps is limited. During the scan, all nodes and connections are fixed. The set of these nodes and connections and determines the reflection of the initial network.

3.2 Building the network reflection using the algorithm of maximum PageRank

At the initial stage, the PageRank characteristic is assigned for each node of the initial network. Next, the network is scanned as follows: from the current node via the outgoing link the transition to the node that has the highest PageRank is made. The network scanning process is continued during a limited number of steps. If during traversing the network there is a return to the node from which the transition has already been made, then the next transition via the not yet passed outgoing link that leads to the node with the highest PageRank is made. If there is no such outgoing link, the scanning process is continued from any random node of the initial network.

3.3 Building the network reflection using the algorithm of maximum degree

Node degree is a characteristic of the total number of both incoming and outgoing links [10, 23, 24].

Building the network reflection using the algorithm of maximum degree is made on the same principle as in the previous case, however, the defining characteristic of the node is not PageRank, but its other characteristic the degree of the node is used.

Since all the proposed algorithms to some extent have a probabilistic nature, then for a particular initial network many different reflections of the network can be obtained. Therefore, the resulting values of the characteristics of the network reflections are averaged over many implementations.

4 Results

The network obtained as a result of using the Barabási-Albert model was used for the research. The graph corresponding to the generated network consists of 200 vertices randomly connected by 398 edges (each new node attached to the network was connected to the existing one only using the 1 edge). The density of the obtained network is 0.01, and the average degree is 1.99

The network scanning was performed using the three algorithms proposed above. The maximum number of scanning steps was 50, 100, 150, and 200.

Fig. 3-5 presents different reflections of the initial Barabási-Albert network for respectively different network scanning algorithms.

Fig. 3 shows the reflection (highlighted in bold) of the initial Barabási-Albert network, which was obtained using 50 scanning steps, which in turn were made on the principle of random walking, i.e. on the principle that uses the common PageRank algorithm. The obtained reflection of the initial network consists of 18 nodes and 33 edges (Fig. 3). The average node degree is 1.833, and the network density is 0.108.

The degree distribution of the nodes of the obtained reflection, as well as the degree distribution of the initial Barabási-Albert network, at least asymptotically follows a power law.

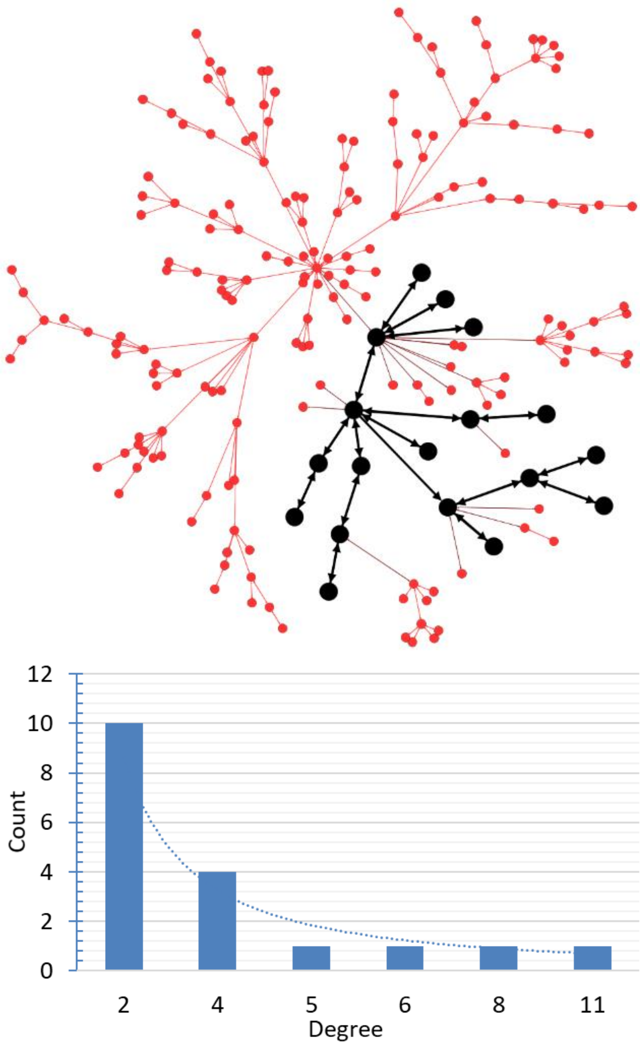

Fig. 3. The reflection of the initial network built using the PageRank algorithm, and the degree distribution of the obtained network reflection.

The research was also done using the algorithm that implements the principle of transition to the node that has the highest degree. After 100 scanning steps of the initial Barabási-Albert network using the above mentioned algorithm of maximum degree, the reflection of the network, which consist of 37 nodes and 69 edges, was obtained (Fig. 4). The average node degree is 1.865, and the network density is 0.052.

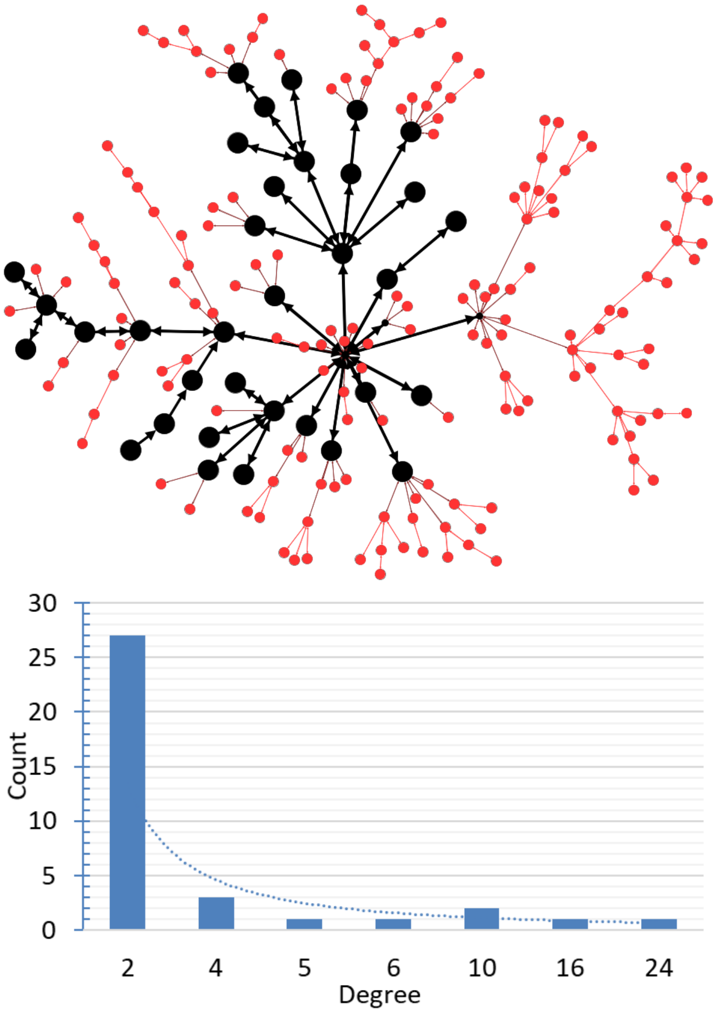

Fig. 4. The reflection of the initial network built using the algorithm that implements the principle of transition to the node that has the highest degree, and the degree distribution of the obtained network reflection.

After 200 steps scanning of the initial Barabási-Albert network using the algorithm of transition to the node that has the highest PageRank, the reflection of the network that consist of 65 nodes and 126 edges was obtained (Fig. 5). The average node degree is 1.938, and the network density is 0.03.

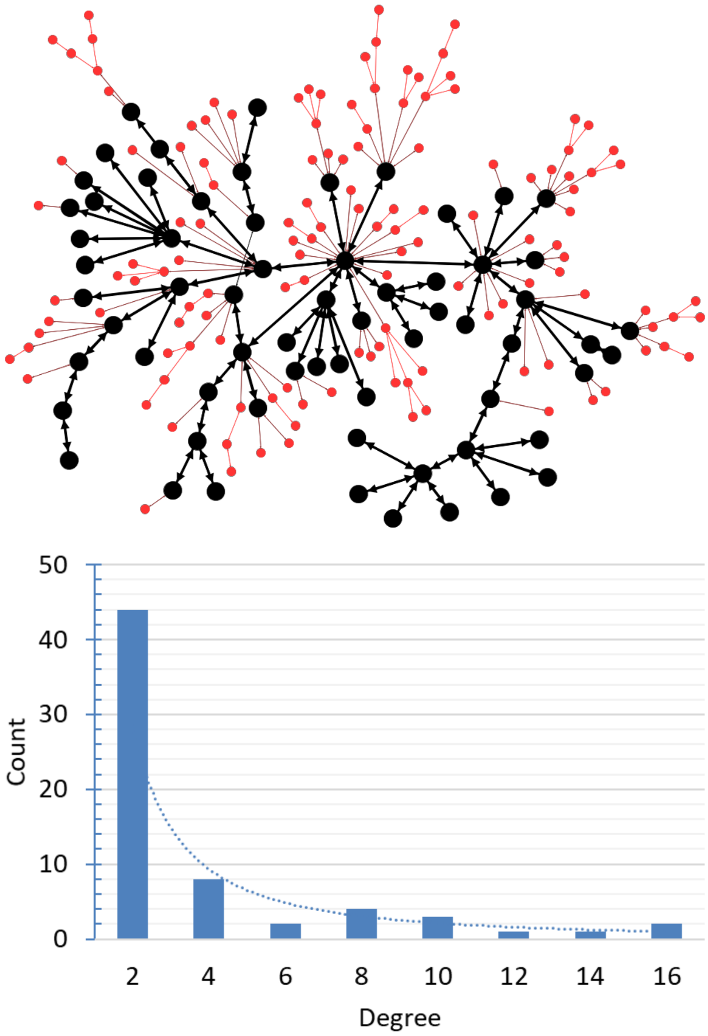

Fig. 5. The reflection of the initial network built using the algorithm that implements the principle of transition to the node that has the highest PageRank, and the degree distribution of the obtained network reflection.

Analyzing the results presented in Fig. 3-5, it can be seen that networks that are the reflections of the initial Barabási-Albert network and built using a limited number of scanning steps, have the same power-law degree distribution as the initial network.

In general, it can be observed that during approaching the number of scanning steps to the number of nodes in the initial network for each of the considered algorithms for scanning this network, the degree distribution of the obtained network reflections approximates a power-law degree distribution, by which the Barabási-Albert network is characterized a priori.

Also, to research the dependence of the network characteristics of the reflections on the number of scanning steps as a initial network was used a network built using the Erdös-Renyi model. The graph corresponding to the generated network consists of 200 vertices randomly connected by 428 edges (the probability for edge creation is 0.01). The average node degree the generated network is 2.14, and the network density is 0.011.

To build the reflection of the initial network (Fig. 4), 200 scanning steps using an algorithm that implements the principle of transition to the node that has the highest degree is made. As result, the reflection of the initial Erdös-Renyi network, which consists of 65 nodes and 143 edges, was obtained. The average node degree is 2.2, and the network density is 0.034.

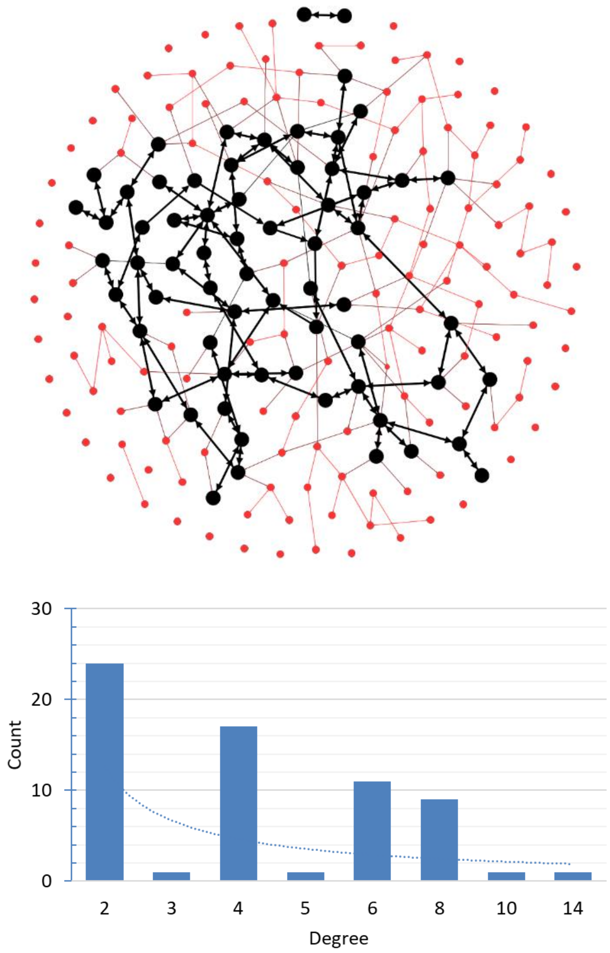

Fig. 6. The reflection of the initial network built using the algorithm that implements the principle of transition to the node that has the highest degree, and the degree distribution for the obtained network reflection.

Fig. 6 also presents the degree distribution of the nodes of the obtained network reflection. As can be seen, using a limited number of scanning steps, the reflection of the initial Erdös-Renyi network, which has distinct from a Poisson degree distribution was obtained. While the initial Erdös-Renyi network has a Poisson degree distribution.

After more detailed research on the example of networks built using the Erdös-Renyi network model, and applying different scanning algorithms, the reflection of the original networks, which have close to a power-law degree distributions were obtained. For example, the Erdös-Renyi network, which has 1000 nodes and 10030 connections (the probability for edge creation is 0.01) and which has a Poisson degree distribution (Fig. 7) was researched. 500 scanning steps using the algorithms that implement the principle of transition to the node that has the highest PageRank and degree, respectively. As a result, the close to a power-law degree distribution of the corresponding network reflections were obtained (Fig. 8 and Fig. 9).

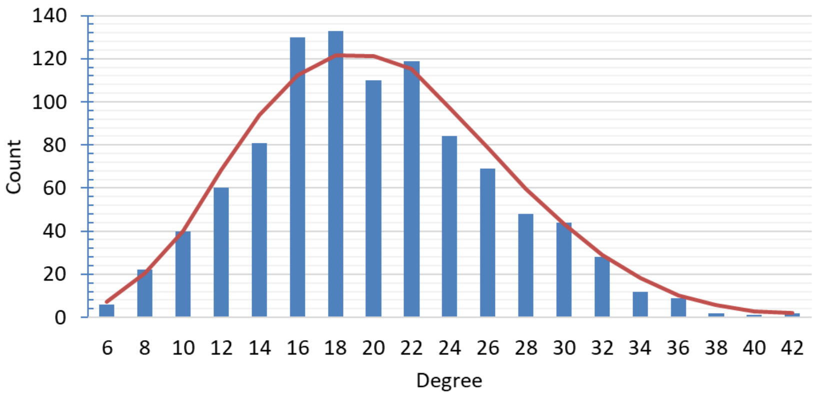

Fig. 7. The Poisson degree distribution of the initial Erdös-Renyi network.

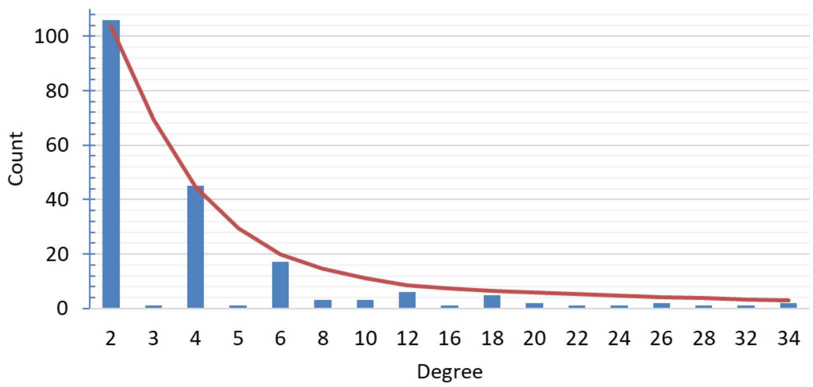

Fig. 8. The degree distribution of the Erdös-Renyi network reflection that built using the algorithm of maximum PageRank.

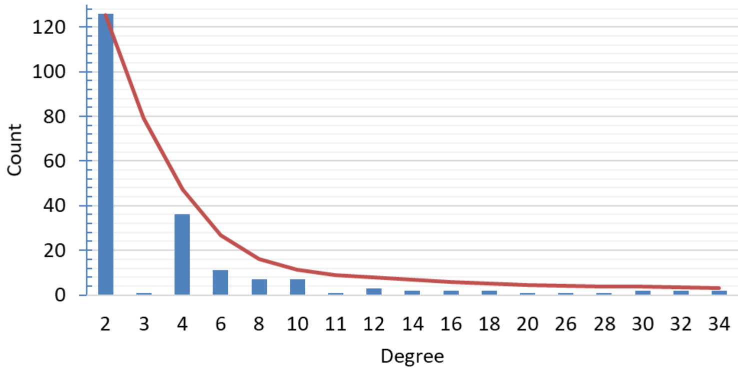

Fig. 9. The degree distribution of the Erdös-Renyi network reflection that built using the algorithm of maximum degree.

The experimental data presented in Table 1 illustrate the dependence of the characteristics (the average node degree and the network density) of the networks reflections on the algorithms used for scanning the initial networks (Barabási-Albert and Erdös-Renyi networks) and on the number of scanning steps.

In general, for a different limited number of scanning steps, depending on the chosen scanning algorithm, the different characteristics of the reflections of the Barabási-Albert and Erdös-Renyi networks were obtained.

It was found that increasing scanning steps leads to an approximation of the average degree and density of the obtained reflections of the Barabási-Albert and Erdös-Renyi networks to the average degree and density of these initial networks. That is, when approaching the scanning steps to the number of nodes in the initial networks for each of the proposed scanning algorithms, it is possible to achieve such characteristics of the networks reflections that are close to the characteristics of the initial networks. Also based on the data obtained by computational experiments, it is shown that almost all reflections of the initial networks that built using the model Barabási-Albert and Erdös-Renyi, and also using a limited number of scanning steps and the different scanning algorithms have a power-law degree distribution.

5 Conclusion

The paper shows that the characteristics of network reflections, which are close to the topology and characteristics of the initial networks, are achieved only during approaching the scanning steps to the number of nodes in these initial networks.

For the first time, the experimentally obtained characteristics of the reflections of networks depending on the characteristics of the initial networks, the scanning algorithms and the number of scanning steps are presented.

Also, based on the data obtained by computational experiments, it was shown that almost all reflections of the initial networks that built using the model Barabási-Albert and Erdös-Renyi, and also using a limited number of scanning steps and the different scanning algorithms have a power-law degree distribution.

The obtained results are important in the methodical plan and can be considered as a statement of a problem of finding the algorithm of scanning of a complex network using of which gives most adequately parameters of the initial network (first of all, the degree distribution)

Table 1. Dependence of the network characteristics of the reflections on the number of scanning steps and the chosen scanning algorithm.

Initial Network	Scanning Algorithm	Number of Scanning Steps	Number of Nodes	Number of Edges	Average Degree	Density
Barabási-Albert network (Number of node is 200, Number of edges to attach from a new node to existing nodes is 1)	PageRank	50	18	33	1.833	0.108
		100	33	61	1.848	0.058
		150	43	78	1.814	0.043
		200	60	115	1.917	0.032
	Max PageRank	50	25	44	1.76	0.073
		100	36	70	1.944	0.056
		150	50	95	1.9	0.039
		200	65	126	1.938	0.03
	Max Degree	50	25	44	1.76	0.073
		100	37	69	1.865	0.052
		150	53	102	1.925	0.037
		200	63	122	1.937	0.031
Erdös-Renyi network (Number of node is 200, Probability for edge creation is 0.01)	PageRank	50	20	34	1.7	0.089
		100	42	70	1.667	0.041
		150	58	103	1.776	0.031
		200	80	132	1.65	0.021
	Max PageRank	50	23	41	1.783	0.081
		100	38	79	2.079	0.056
		150	56	108	1.929	0.035
		200	73	147	2.014	0.028
	Max Degree	50	21	38	1.81	0.09
		100	48	95	1.979	0.042
		150	57	118	2.07	0.037
		200	65	143	2.2	0.034

References

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C.: Introduction to algorithms. MIT press (2009).
2. Tarjan, R.: Depth-first search and linear graph algorithms. SIAM journal on computing, 1(2), 146-160 (1972).
3. Korf, R. E.: Depth-first iterative-deepening: An optimal admissible tree search. Artificial intelligence, 27(1), 97-109 (1985).
4. Dijkstra, E. W. A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271 (1959).
5. Floyd, R. W.: Algorithm 97: shortest path. Communications of the ACM, 5(6), 345 (1962).
6. Warshall, S.: A theorem on Boolean matrices. Journal of the ACM (JACM), 9(1), 11-12 (January 1962).
7. Bellman, R.: On a routing problem. Quarterly of applied mathematics, 16(1), 87-90 (1958).
8. Ford, Jr, L. R., Fulkerson, D. R.: Flows in networks. Princeton university press. (2015).
9. Tarjan, R. E., RE, T.: A Note on Finding the Bridges of a Graph. Information Processing Letters, 2(6), 160-161 (1974).
10. Snarskii A. A., Lande D. V.: Modeling of complex networks: tutorial. K.: Engineering (2015). ISBN 978-966-2344-44-8
11. Newman, M. E.: Finding community structure in networks using the eigenvectors of matrices. Physical review E, 74(3), 036104 (2006).
12. Jensen D., Neville J.: Data Mining in Social Networks. In: Proceedings of the National Academyof Sciences Symposium on Dynamic Social Network Analysis, pp. 289–302 (2002).
13. Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A.: Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 1-37 (2011).
14. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 177-187 (2005).
15. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 462-470 (2008).
16. Barabási, A. L., & Albert, R.: Emergence of scaling in random networks. Science, 286(5439), 509-512 (1999).
17. Erdős, P., Rényi, A.: On Random Graphs. I. Publicationes Mathematicae, 6, 290–297 (1959).
18. Watts, D. J., & Strogatz, S. H.: Collective dynamics of 'small-world'networks. nature, 393(6684), 440-442 (1998).
19. Erdős, P., & Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 5(1), 17-60 (1960).
20. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Stanford InfoLab (1999).
21. Langville, A. N., Meyer, C. D.: Google's PageRank and beyond: The science of search engine rankings. Princeton University Press (2011).
22. Rajaraman, A., Ullman, J. D. Mining of massive datasets. Cambridge University Press. (2011).

23. Albert, R., Barabási, A. L.: Statistical mechanics of complex networks. Reviews of modern physics, 74(1), 47 (2002).
24. Newman, M. E. The structure and function of complex networks. SIAM review, 45(2), 167-256 (2003).