Modelling and Analysis the Network Model Controlled by Mediation-Driven Attachment Rule

Vadim Shergin[1[0000-0002-4388-8180]], Serhii Udovenko[2[0000-0001-5945-8647]], Larysa Chala[1[0000-0002-9890-4790]]

[1] Kharkov National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine
vadim.shergin@nure.ua, larysa.chala@nure.ua
[2] Simon Kuznets Kharkiv National University of Economics, Nauky Ave. 9-a, Kharkiv, 61166, Ukraine
udovenkosg@gmail.com

Abstract. The problem of modeling scale-free networks is considered. Mediation-driven attachment rule with constant number of incoming links is studied. According to that rule, new nodes do not need access to node statistics of the network, although the probability of joining to existing node is determined by its degree, this information is used implicitly. Such feature looks quite natural and is a key advantage of mediation-driven attachment rule. The dependence of the scaling factor of nodes degree distribution on the control parameter – the number of links associated with incoming nodes – is analyzed both analytically and by numerical simulation.

Keywords: scale-free network; mediation-driven attachment; power law; scaling factor; Yule-Simon distribution

Introduction

The scale-free network's theory is a relatively new science area. The key feature of such networks is that their nodes distribution by degree follows to a power law [1]. Social networks, citations of scientific papers, collaboration networks, neural and protein networks and many other real-world networks [2] are scale-free. On the other hand, power law distribution is a characteristic feature of fractal properties of the system. This is another reason for the interest to the scale-free networks.

At the heart of most of today's scale-free network models is the model proposed by Barabási and Albert in 1999 (BA-model) [3]. It is based on two fundamental concepts: concept of growth and concept of preferential attachment: the probability for a new node to link with some existing one is proportional to its degree. Although at present there exists many specifications and generalizations for the generative rule used to create network models [4-17], their common features ascends to BA-model.

The concept of implicit use of node statistics, implied in the mediation-driven attachment (MDA) model, was proposed by Hassan [13]. Forming and analysis the dynamic model of network, controlled by MDA-rule is the subject of current interest.

1 Scale-Free Network Models Overview

According to the concept of growth, the seed network comprises n_0 nodes and $L_0/2$ edges, and then at each time step, a new node is added to the network. Thus, network size n (i.e. number of nodes) can be used as time measure without lost of generality. In this case, the node number (i) is considered as the time of its birth (i.e. $n-i$ is the node's age).

Aside from the time of birth, each node is characterized by its degree (d_i) and other properties P_i (fitness, etc.). Network growth dynamics is determined by the number of links $m(n)$ that connect the incoming node with existing ones, or by the average degree of nodes $\bar{k}(n)$. The network may also have other control parameters C (such as additional attractiveness, elasticity etc.).

Network properties are determined by the attachment rule, i.e. the probability π_i that a new added node links to existing one i, as a function of the above parameters:

$$\pi_i = f(i, k_i, n_0, L_0, P_i, C) . \quad (1)$$

A key feature of the scale-free networks is the distribution of the nodes by degree $p(k)$ that should follow (at least, asymptotically) to the Yule-Simon law (which is a discrete analogue of the power law):

$$p(k) = C\frac{\Gamma(k)}{\Gamma(k+\gamma)} \propto k^{-\gamma}, \quad k \geq k_0 \neq 0. \quad (2)$$

The parameter $\gamma \geq 2$ is called the scale factor.

The simplest and most known model for scale-free networks is the BA model [3]. It is based on two fundamental concepts: concept of growth and concept of preferential attachment, according to which each incoming node connects to $m = const$ existing ones. The probability π_i that a new edge points to some node i is proportional to its degree k_i:

$$\pi_i = k_i / \sum_i k_i . \quad (3)$$

Thus, BA attachment rule is the simplest case of a general form (1). This model leads to the Yule-Simon distribution for nodes degree (2) with scaling factor $\gamma = 3$.

According to the fitness model [4], each node is associated with its fitness value η_i. Therefore, rule (3) is extended into $\pi_i = f(k_i, \eta_i) = k_i\eta_i / \sum k_i\eta_i$. More advanced variants of weighting also exist [5-7]. This approach allows reproducing the degree correlation among network nodes.

Another extensions of BA model are based on aging [6], i.e. using the attachment rule (1) in form $\pi_i(n) = f(i, k_i(n))$.

In the Price model [2] the network is supplied by global additional attractiveness parameter a, and attachment rule has the form $\pi_i = (k_i + a)/\sum(k_i + a)$. This approach was used for modelling citation networks, which are considered as directed. Therefore, in the attachment rule k_i denotes only in-links. Clearly, if $a = 0$ (i.e. if the BA rule (3) is used), the incoming paper having any reference yet, cannot be cited.

At now, there exists many other specifications and generalizations for the generative mechanism (1) used to create network models (based on nonlinear preferential attachment [8-9], rewiring [10], second level of neighborhood data [11] etc.).

Not all of the above models produce scale-free networks. In fact [9, 12], the only class of attachment rules (1) that produce a scale free networks is determined by the fact that attachment probability is asymptotically linear under k_i, i.e. $\pi_i \sim a_\infty k_i$.

However, the BA-model just as all above inherited ones, has bears some original features that limit their application area. One of them is explicit using of nodes degree information. In real-world networks, the incoming node cannot assess the information for the whole of existing nodes due to the large size of the network, confidentiality reasons etc. Really, for a first choose a site or scientific paper, we are not interested how many references to other sites or papers it has. Similarly, a novice businessman cannot obtain the information about who of his potential trade partners has how many trade links. However, after visiting the initial website or reading an introductory scientific paper, the reader may want to follow some available links from this site / paper to other sources. Thus, this original site / paper act as a mediator.

The mediation-driven attachment (MDA) model was proposed by Hassan et al. [13] to overcome the above restriction. It can be regarded as a version of the earlier copy-model [15-16]. MDA model forms a subject of interest of current paper.

2 The Mediation-Driven Attachment Rule Model

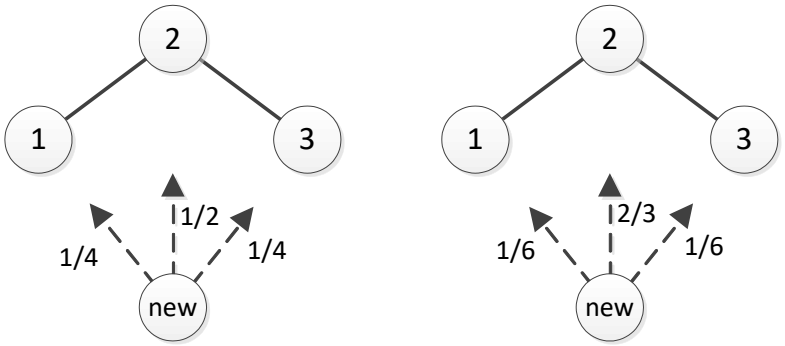
According to MDA rule, each incoming node chooses some existing one uniformly at random. This node is called mediator. Then incoming node links at random to m neighbors of this mediator. The probability π_i for node i to link with incoming one can be found by the following way: let the neighbors of node i are enumerated $1, 2, ..., k_i$. The probabilities for each of them to be chosen as mediator are equal to $1/n$. The probability for node i to be reached through mediator j is $1/k_j$. Thus,

$$\pi_i = \frac{1}{n} \sum_{\substack{j=1, \\ j \square i}}^{k_i} \frac{1}{k_j}. \qquad (4)$$

As per (4), the probability for incoming node to connect with some existing ones also depends only on their degrees but in contrary to BA models this dependence is implicit.

The comparison of BA and MDA rules is illustrated on Fig. 1. According to (3), in the BA model the incoming node "knows" that $k_1 = k_3 = 1$, and $k_2 = 2$. Thus,

$\pi_i = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. As for the MDA rule, the incoming node is linked with node 1 only if node 2 is selected as an mediator, which occurs with probability 1/3, and then if node 1 is choosen from the neighbours of node 2 (the probability is 1/2). So, $\pi_1 = \frac{1}{6}$, and similarly, $\pi_3 = \frac{1}{6}$. Node 2 will be linked with the incoming one if node 1 or node 3 is selected as an mediator, therefore $\pi_2 = \frac{2}{3}$.



BA attachment rule MDA attachment rule

Fig. 1. Comparison among the BA-rule and MDA-rule in the case of m = 1

The evolution of the MDA network in the case $m = 1$ is shown on Fig. 2. The nodes of the networks are labelled by the probabilities of joining a new node. The probabilities of implementing the corresponding network structures are shown in the headers.

It can be seen that in contrast to the BA-model, the probability of forming a star-shaped structure (hyperhab) is very high, so, if m is small, the MDA model is prone to the "winner takes all" paradigm.

Indeed, if the network with n nodes is star-shaped, then the new node will be linked with the central one if any non-central node is choosen as the mediator. Thus, the probability of persisting the star-like structure on n-th step is equal to

$$\pi_*^{(n)} = (n-1)/n\,. \tag{5}$$

Therefore, the unconditional probability of forming a star-like structure under MDA-rule with $m = 1$ is very high:

$$P_*^{(n)} = \frac{2}{n-1}\,. \tag{6}$$

In comparison, the probability of forming a star-like network under the BA model is only $P_*^{BA} = 2^{3-n}$.

The first two step of evolution of the MDA network under the case of $m = 2$ are shown on Fig. 3.

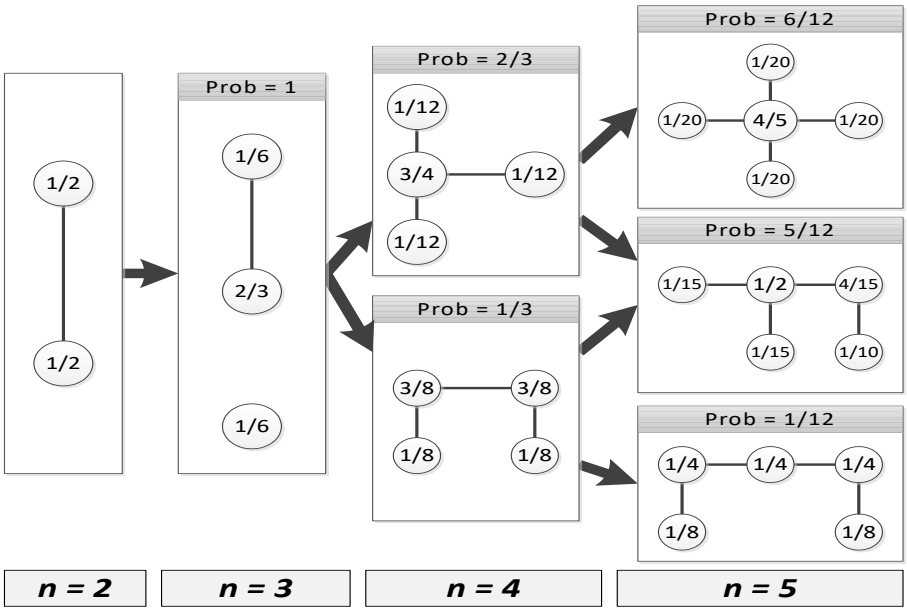


Fig. 2. Evolution of the MDA network in the case of m = 1

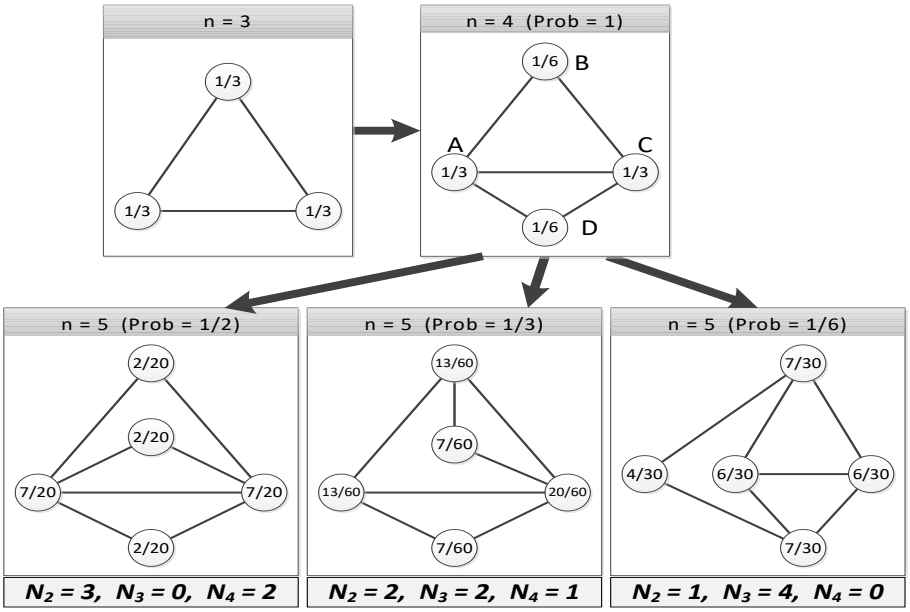


Fig. 3. Evolution of the MDA network in the case of m = 2

Comparing Fig. 2 and Fig. 3, it can be concluded that the probability of a high concentration of links (i.e. the presence of "super hubs") under $m = 2$ is slightly lower than in the case of a tree network with $m = 1$, but still is much greater than for the BA network.

According to results of numerical simulation from the original paper [13], the distribution for nodes degree asymptotically also follows to a power law (2) with scaling factor

$$\gamma \approx 3(1 - 0.43531 \cdot e^{-0.11m}) . \tag{7}$$

However, this dependence is purely empirical. It does not reflect the nature of the impact of the control parameter (m) on the scaling factor. Thus, it is necessary to consider a dynamic model of network, controlled by MDA-rule.

3 Dynamic Properties of the MDA Network

Lets consider the general case of MDA-network dynamics in the case of an arbitrary (but constant) value of the control parameter $m = const$.

We assume that the initial network is a complete graph of $n_0 = m + 1$ nodes. Thus, all nodes have degree m. A complete graph of $m+1$ nodes is the minimum number of nodes among all graphs for which the degree of nodes is not less than m. Initially network contains $L_0 = m \cdot n_0 = m(m+1)$ links (or $e_0 = m(m+1)/2$ edges).

Since each step a single node is added to the network, it is natural to use the network size as the step number (i.e. as the time measure). Each new node adds m edges, or $2m$ links, so the total number of links and the average node degree can be expressed as

$$L(n) = m \cdot (2n - n_0) , \tag{8}$$

$$\bar{k}(n) = m(2 - \tfrac{n_0}{n}) , \qquad \lim_{n \to \infty} \bar{k}(n) = 2m . \tag{9}$$

The probability for an existing node i to get a new link is $m \cdot \pi_i(n)$, so the expected value of i-th node degree increased just on this value:

$$k_i(n+1) - k_i(n) = m \cdot \pi_i(n) . \tag{10}$$

The expression (4) for $\pi_i(n)$ can be transformed into

$$\pi_i(n) = \frac{1}{n} \sum_{\substack{j=1, \\ j \neq i}}^{k_i(n)} \frac{1}{k_j(n)} = \frac{k_i(n)}{m \cdot n} \left(\frac{m}{k_i(n)} \sum_{\substack{j=1, \\ j \neq i}}^{k_i(n)} \frac{1}{k_j(n)} \right) . \tag{11}$$

We may denote the expression in parentheses of (11) as $b_i(n)$:

$$b_i(n) = \frac{m}{k_i(n)} \sum_{\substack{j=1, \\ j \neq i}}^{k_i(n)} \frac{1}{k_j(n)} . \tag{12}$$

Next, we assume that for a large network ($n \to \infty$), parameter (12) tends to some constant value:

$$\lim_{n \to \infty} b_i(n) = \beta . \tag{13}$$

Then, substituting (13) (instead of (12)) to (11), and then to (10), we obtain the dynamic equation for evolution the *i*-th node degree:

$$k_i(n+1) - k_i(n) = \beta \cdot \frac{k_i(n)}{n} . \tag{14}$$

This difference equation has a general solution:

$$k_i(n) = C_i \cdot \frac{\Gamma(n+\beta)}{\Gamma(n)} , \tag{15}$$

where $\Gamma(x)$ is the Euler's gamma function.

Without lost of generality, we can assume, that the nodes are sorted by age, so, at Then at step $i \geq n_0$ the network consists of i nodes. The new node will receive the number $i+1$, and, in accordance with the MDA rule, it will establish exactly m connections with existing nodes. So, $k_i(i) = m$, and, therefore,

$$C_i = m \frac{\Gamma(i)}{\Gamma(i+\beta)} , \qquad i \geq m . \tag{16}$$

For nodes that are the "founding parents" of the network, i.e. for the first $n_0 = m+1$ nodes, the initial conditions are $k_i(n_0) = m$, so

$$C_i = m \cdot \frac{\Gamma(n_0)}{\Gamma(n_0+\beta)} , \qquad 1 \leq i \leq n_0 . \tag{17}$$

Taking into account (16)-(17), the expected value of *i*-th node degree (15) takes its final form:

$$k_i(n) = m \cdot \frac{\Gamma(i^*)}{\Gamma(i^*+\beta)} \cdot \frac{\Gamma(n+\beta)}{\Gamma(n)} , \qquad i^* = \max(i, n_0) . \tag{18}$$

The dependence (18) is a Yule-Simon distribution (2). Asymptotically, as $i \gg 1$ and $n \gg 1$ it goes into the power law with scaling factor β :

$$k_i(n) \propto m \cdot (n/i)^{\beta} . \tag{19}$$

From (18)-(19) it is easy to see that the dependence of the node degree on its number is monotonically decreasing, so the node number (i.e. age) can be an expectation of its rank.

Considering the age distribution (18), or its approximation (19), as a nodes rank distribution, the power rank distribution with the scaling factor β corresponds to the power law of frequency distribution with the scaling factor (exponent)

$$\gamma = 1 + \frac{1}{\beta}. \tag{20}$$

Thus, we can summarize that applying the MDA rule as well as in the case of using the explicit preferential linking (BA-model) should lead to a scale-free network. The only difference is the value of the scaling factor: $\gamma = 3$ for the BA model, but has the form (20) for the MDA one.

However, it should be noted that the above conclusion is based on assumption (13), i.e. that the scaling factor β is a constant. Estimation of this parameter poses the following problem to solve.

4 Scaling Factor Estimation for MDA Model

The assumption of independence $b_i(n)$ on i (as well as on k_i) means that asymptotically the average values of inverse degrees of nodes connected to the node i is the same as the average of the values inverse to the degree nodes throughout the whole network:

$$b_i(n) = \frac{m}{k_i(n)} \sum_{\substack{j=1,\\ j\sim i}}^{k_i(n)} \frac{1}{k_j(n)} \approx \frac{m}{n} \sum_{i=1}^{n} \frac{1}{k_i(n)}. \tag{21}$$

The sum in the approximate expression (21) can be found by summing $1/k_i(n)$ from (18):

$$\sum_{i=1}^{n} \frac{1}{k_i(n)} = \frac{n+\beta}{m(\beta+1)} + \frac{\Gamma(n)}{\Gamma(n+\beta)} \cdot \frac{\Gamma(n_0+\beta)}{\Gamma(n_0)} \cdot \frac{\beta}{\beta+1}. \tag{22}$$

As $n \to \infty$, the second component in (22) decreases to zero as $O(n^{-\beta})$, so the estimate of parameter $b_i(n)$ (21) has the form

$$\lim_{n\to\infty} b_i(n) = \lim_{n\to\infty} \left(\frac{m}{n} \cdot \frac{n+\beta}{m(\beta+1)} \right) = \frac{1}{\beta+1}. \tag{23}$$

On the other hand, according to (13), just value $b_i(n)$ was denoted as β, so the value of β can be found by equating (13) to (23):

$$\beta = 1/(\beta + 1)\,. \tag{24}$$

Equation (24) is well-known and its solution is the golden ratio:

$$\lim_{n\to\infty} b_i(n) \approx \beta = \varphi\,, \qquad \varphi = \frac{\sqrt{5}-1}{2} \approx 0.618\,. \tag{25}$$

From (20) and (25) it follows that the scaling parameter of the nodes degree frequency distribution for the MDA network should be asymptotically equal to

$$\lim_{n\to\infty} \gamma = 2 + \varphi \approx 2.618\,. \tag{26}$$

It should be noted that obtained dependence is in contradiction with the results of the original research [13], according to which scaling factor γ is not a constant under m, but monotonously increasing function (7) up to bound $\gamma \propto 3$.

Thus, numerical experiment for estimating the scaling factor, just as determining the threshold at which the superpreferential attachment changes to the usual preferential one, is of practical value.

5 Numerical Simulation of MDA Network

Numerical simulations were made. Initially, $M = 100$ networks each of 4096 nodes was generated to find whether they could be considered as scale-free (i.e. whether the nodes degree is distributed by a power law). Rank distributions of nodes degree for cases $m = 1$ and $m = 5$ are shown on Fig. 4 and Fig. 5. The distribution of nodes degree by their age for the case $m = 100$ is shown on Fig. 6.

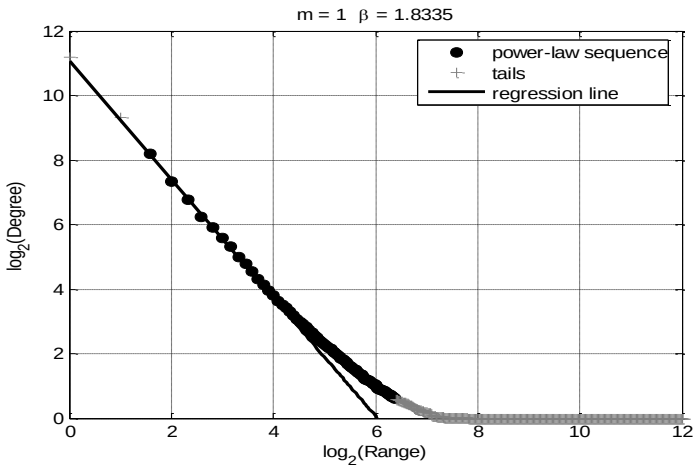


Fig. 4. Rank distribution of nodes degree for the case $m = 1$

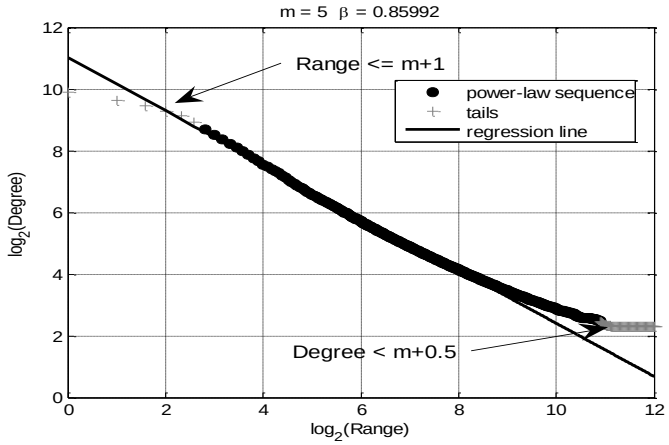


Fig. 5. Rank distribution of nodes degree for the case $m = 5$

According to Fig.4-Fig.6, the graphs can be segmented into three sections: initial (left tail), middle, and final (right tail). It is easy to see that rank distributions of nodes degree of the middle section follow to the power law with great accuracy. Regression lines are also shown on Fig. 4 - Fig. 6, and the corresponding estimates of scaling factor are shown in the graphs headings.

As one can see, the first $n_0 = m + 1$ nodes significantly deviate from the power law sequence, i.e. from regression lines, shortage the links. This phenomenon is quite natural, because the distribution law (18) is a Yule-Simon distribution, which coincides with the power law only asymptotically. In addition, according to (18) there is an operation of rounding small values of n up to n_0. That is why the regression lines were built without the initial nodes.

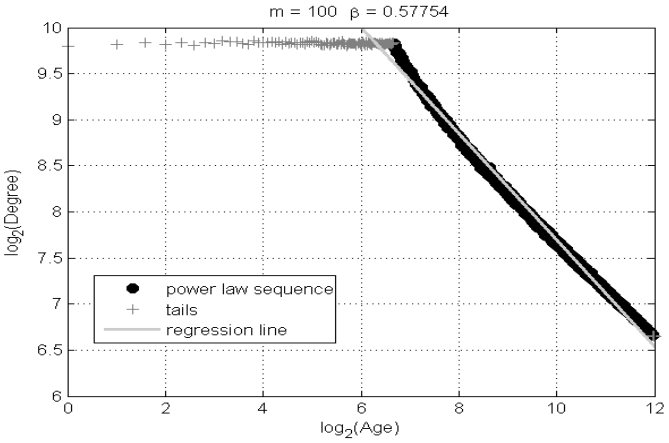


Fig. 6. Distribution of nodes degree by age for the case $m = 100$

On the other hand, the end nodes are also not subject to a power law. The existence of this tail has no theoretical justification, but it can be explained empirically: in the super-preferential network, the nodes that appeared later do not have time to collect links, additional to those they have at birth. During the numerical simulation, the threshold was set manually as $m+0.5$. According to the simulation results (Fig. 4 - Fig. 6), such threshold value allows to separate quite clearly the "young" nodes (i.e. the right tail) from the "middle-aged" nodes, which follows to a power law. It can be noted that the fraction of the tail section decreases rapidly with increasing the control parameter m: the tail contains almost 98% of the total number of nodes under $m=1$, but about half under $m=5$, and less than 10% under $m=100$.

Studying the empirical dependence of the scaling factor under control parameter was the next stage of current research.

According to the results of the initial research (Fig. 4 - Fig. 6), the distribution of nodes by degree can be considered as power law, but it's scaling factor differs from the theoretically predicted golden ratio (25), and secondly, significantly depends on the control parameter m: it is equal to $1.8335, 1.2861, 0.8599, 0.6731, 0.5775$ under $m=1,2,5,10,100$ respectively. Thereby, the dependence of the scaling index (β) on the control parameter m (the number of links for newly added nodes) were studied.

For each m from a given set of values $1\le m\le 100$, $M=100$ networks each of $N=4096$ nodes were generated. For each network, the scaling factor was estimated separately. The results are shown in Fig. 7. Single dots correspond to the individual networks. Since obtained estimates has a very high variability, especially for small values of m, the median averaging is more appropriate instead of mean one. The solid line connects the medians of the obtained estimates. For comparison, in the same figure, the empirical dependence (7) is shown as dotted line with circle markers.

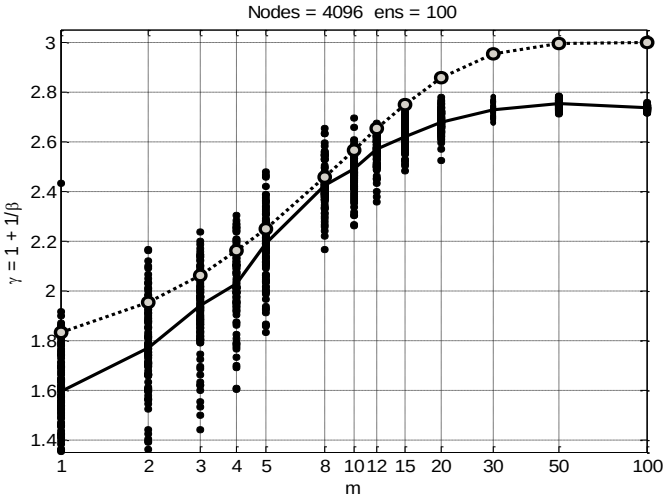


Fig. 7. Dependence of the scaling factor of the frequency distribution of the nodes degree on the control parameter *m* for MDA-ruled networks

Analyzing the results, shown on Fig. 7, we can draw some conclusions:

- for small values of the control parameter m the variability of exponents is so high that we cannot talk about their more or less reliable value;
- moreover, under $m < 5$ the scaling factor of frequency distribution of the nodes degree is $\gamma < 2$, so it does not even have an expected value. This means that for small m the network is super-preferential;
- with increasing the control parameter, the fraction of nodes, which degree follows to a power law, increases (Fig. 4 - Fig. 6), and the variability of the scaling factor decreases;
- the dependence of scaling factor $\gamma(m)$ is not monotonous; it reaches a maximum under $m \approx 50$, however, this effect may be caused by the bounded networks size, so it needs additional research;
- the obtained empirical dependence $\gamma(m)$ does not correspond to either the theoretical model developed in Section 4 (according to which $\gamma = 2.618$) or the model (7) with $\gamma \propto 3$.

Given the above, the MDA model requires a more detailed analytical study of its properties.

Conclusion

The problem of modeling scale-free networks is considered. The properties of generated network are strongly depends on the attachment rule. An actual models and corresponding attachment rules were analyzed.

The mediation-driven attachment rule has an undoubted advantage due to indirect using of nodes statistics, which is fully consistent with the properties of the real-world networks. The properties of networks controlled by the above rule were analyzed. These properties are mainly determined by the control parameter – number of incoming links brought by each incoming node.

The scaling properties of the network, formed according to the studied rule, were analyzed in detail, both analytically and numerically. The threshold boundaries of the scale-free segment of network were found as well as its dependence on the control parameter. The dependence of the scaling factor of the frequency distribution of the nodes degree on the control parameter for networks under mediation-driven attachment rule were obtained numerically and analyzed in detail. An estimate of the threshold boundary of the control parameter at which the network is super-preferential is obtained.

A detailed study of the properties of studied model, as well as analysis of its assortativity properties, forms the subject of interest for future research.

References

1. Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 2005, 46(5). p.323-351.
2. Dorogovtsev S. N., Mendes J. F. F. (2003). Evolution of Networks: From Biological Networks to the Internet and WWW – Oxford, USA: Oxford University Press, 2003. – 280 p.
3. Albert, R., Barabasi A.-L (2002). Statistical mechanics of complex networks: Rev. Mod. Phys. - V. 74. - p. 42-97.
4. Bianconi, Ginestra; Barabási, Albert-László (2001). "Competition and multiscaling in evolving networks". Europhysics Letters. 54 (4): 436–442. doi:10.1209/epl/i2001-00260-6
5. Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2000), Structure of growing networks with preferential linking, Phys. Rev. Lett., 85, 4633.
6. Caldarelli, Guido; Catanzaro, Michele (2012). Networks: A Very Short Introduction. Oxford University Press. p. 78
7. Choromański, K.; Matuszak, M.; MięKisz, J. (2013). Scale-Free Graph with Preferential Attachment and Evolving Internal Vertex Structure. Journal of Statistical Physics. 151 (6): 1175–1183.
8. Kunegis, Jérôme & Blattner, Marcel & Moser, Christine. (2013). Preferential Attachment in Online Networks: Measurement and Explanations. WebSci13 ACM Web Science Conference 2013. 13. doi: 10.1145/2464464.2464514.
9. Krapivsky, P. L.; S. Redner; F. Leyvraz (2000). "Connectivity of Growing Random Networks". Phys. Rev. Lett. 85: 4629–4632. doi: 10.1103/physrevlett.85.4629
10. R. Noldus, P. Van Mieghem, “Assortativity in complex networks”, J. Complex Networks, vol. 3, pp. 507-542, 2015.
11. Dangalchev Ch., Generation models for scale-free networks, Physica A 338, 659 (2004).
12. Broido, Anna; Aaron Clauset: "Scale-free networks are rare". Nature Communications. 10 (1): 1017. (2019) doi:10.1038/s41467-019-08746-5.
13. Hassan, M. K.; Islam, Liana; Arefinul Haque, Syed (2017). "Degree distribution, rank-size distribution, and leadership persistence in mediation-driven attachment networks". Physica A. 469: 23–30. doi:10.1016/j.physa.2016.11.001.
14. Vadim Shergin, Serhii Udovenko, Larysa Chala (2021). Assortativity Properties of Barabási-Albert Networks. In Data-Centric Business and Application (vol.5). Springer. (2021). https://doi.org/10.1007/978-3-030-43070-2_4.
15. Kumar, Ravi; Raghavan, Prabhakar (2000). Stochastic Models for the Web Graph (PDF). Foundations of Computer Science, 41st Annual Symposium on. pp. 57–65. doi:10.1109/SFCS.2000.892065
16. Callaway, Duncan S.; Newman, M. E. J.; Strogatz, S. H.; Watts, D. J. (2000). "Network Robustness and Fragility: Percolation on Random Graphs". Physical Review Letters. 85 (25): 5468–71. doi:10.1103/PhysRevLett.85.5468.
17. Shergin V. L., Chala L. E., Udovenko S. G. (2018). Fractal dimension of infinitely growing discrete sets. Advanced Trends in Radioelecrtronics Telecommunications and Computer Engineering (TCSET), no. 348.