Petnica Science Center, Serbia
Marina Radulaški

Global and Local Properties of Scale-free Networks

mentor: Dr Aleksandar Belić
Scientific Computing Laboratory
Institute of Physics, Belgrade

Abstract

This paper investigates global and local properties of scale-free networks generated by the Barabasi-Albert model. Small world coefficient and clustering coefficient dependences on network size have been expanded to dependences on parameter m. Dependence of size three network substructure occurrences on network size and parameter m has been discovered. Counting of size three substructures occurrences for random and preferentially selected subset of nodes was investigated as a heuristic for grading same values on the level of the whole network. Random selection is found to be simpler, but less precise method. Formulas and coefficients for preferential selection are introduced and can be used for estimation of the number of size three network substructures for $1 \leq m \leq 8$.

Introduction

Research of complex networks started in 1950's with the development of graph theory. A simple realization, random network – Erdos-Renu (ER) model was used troughout the second half of 20th century. Random network represents a set of nodes connected to each other with a certain probability.

Given a number of links per node (node degree) and number of nodes with *k* degree, degree distribution can be observed. Degree distribution *P*(*k*) gives the probability that a selected node has exactly *k* links. ER model has Poissonian degree distribution.

Since 1999, degree distribution of real networks, such as the World Wide Web, the Internet, metabolic networks etc. [1] have been found to follow power-law,

$$P(k) \sim k^{-\gamma}.$$

These networks have been named scale-free (SF) networks. Their characteristic is a big number of low degree nodes and small number of high degree nodes. Scale-free networks have shown success in modeling social, biological, infrastructure and many other networks.

This paper investigates global and local properties of undirected scale-free networks.

Scale-free network generation and its parameters

Barabasi-Albert (BA) model [1] is used for generation of undirected scale-free networks. This model starts from a small random network and imports new nodes into the network using preferential attachment.

Random network is generated by connection of n_0 nodes to each other with probability *p*. Probability for each pair of nodes' link is equal to *p*. In every time-step a new node is introduced to the network, so after *t* time-steps the network size (number of nodes) is $n_0 + t$. Each new node is linked to *m* nodes from the existing network, thus the number of links is increased by *m* after each time-step. Preferential attachment is used, which means new node probability for linking with existing node *k* is proportional to its degree s_k and is equal to $\frac{s_k}{\sum_i s_i}$, with $\sum_i s_i$ denoting the sum of degrees of all nodes in the existing network.

Preferential attachment results in power-law degree distribution $P(k) = ck^{-\gamma}$ (Figure 1). Values of γ coefficient obtained belong to the interval (2.50, 2.85). Most of real scale-free networks have $\gamma \in (2,3]$.

Distance between two nodes represents the least number of links needed to make a path between the two nodes. Neighbor node has the distance one.

Mean distance between all pairs of nodes in the network is called network radius or small-world coefficient. Radius dependence on network size is logarithmical [1] and can be fitted to the function $y = a_r + b_r \ln x$ (Figure 2).

Highly connected networks have small values of small-world coefficient. An example is the network of acquaintances with nodes representing people, linked if they have met each other. The radius of this network is (roughly) six.

Another important global network parameter is clustering coefficient. Clustering coefficient represents the mean ratio between existing and possible links in the set of a node's neighbors and can be calculated as $\frac{1}{n}\frac{t_i}{s_i\left(s_i-1\right)/2}$ where *n* stands for network size, s_i for *i*-th node degree, t_i for number of links between *i*-th node neighbors. Clustering coefficient has power-law dependence on number of nodes in the network [1] and can be fitted to the function $y=a_c x^{-b_c}$ (Figure 3).

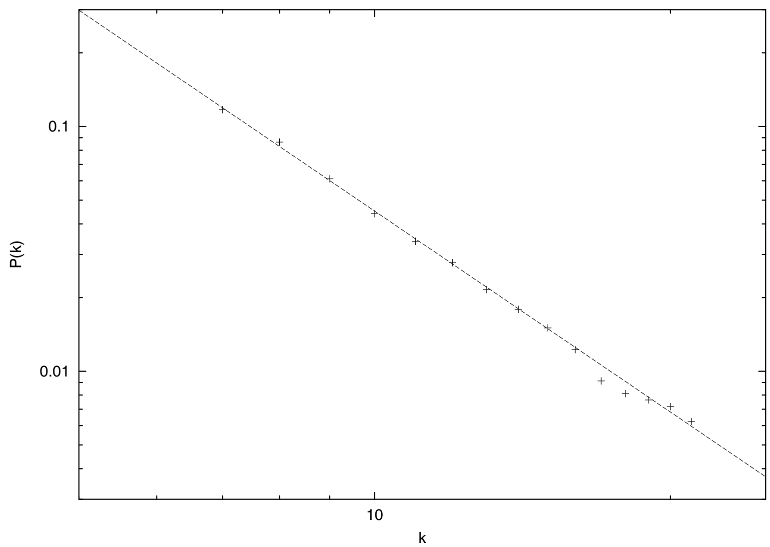

Figure 1: Scale-free degree distribution *P*(*k*) fitted to the function $y=cx^{-\gamma}$ is shown on log-log plot, *c* = 24,6(6), γ = 2,74(2). The network size is *n* = 5.010, other parameters are n_0 = 10, *p* = 0.5, *m* = 5.

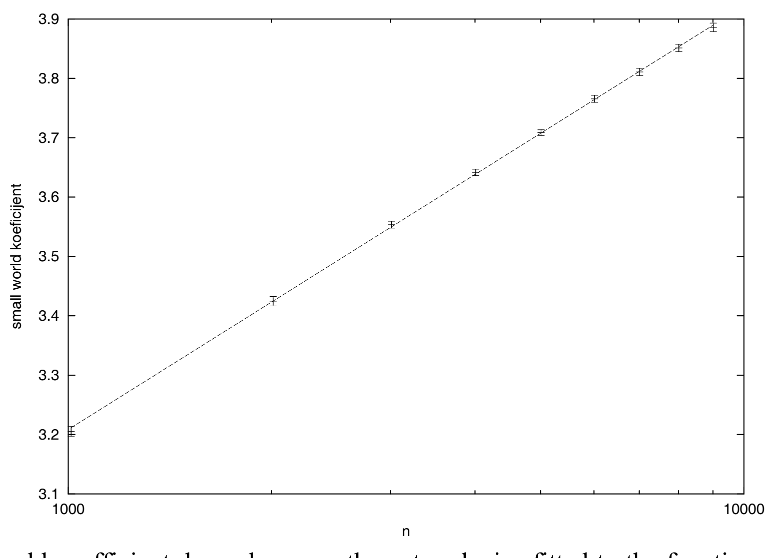

Figure 2: Small world coefficient dependence on the network size fitted to the function $y=a_r+b_r\ln x$, a_r = 1.07(2), b_r = 0.310(2). Network parameters are n_0 = 10, *p* = 0.5, *m* = 4. Data on the plot are mean values obtained on four different networks.

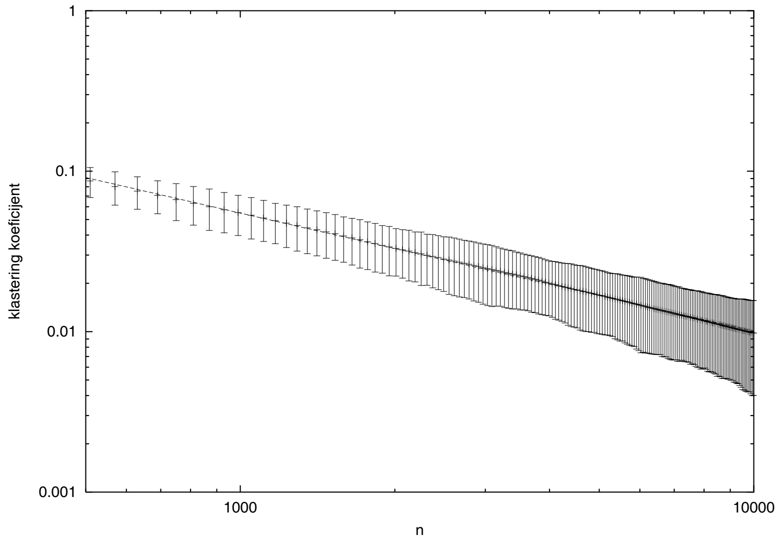

Figure 3: Clustering coefficient dependence on the network size fitted to the function $y = a_c x^{-b_c}$ is shown on log-log plot, a_c = 8.87(2), b_c = 0.7362(2). Network parameters are n_0 = 10, *p* = 0.5, *m* = 8. Data on the plot are mean values obtained on ten different networks.

Value of *m*-parameter (number of links assigned to a new node) influences network connectivity and thus the values of small-world and clustering coefficients. Therefore, a_r, b_r, a_c, b_c coefficients from previously mentioned functions are *m*-dependent.

For *m*≥2 new node attachment can result in creation of new contours, whereas for *m*=1 contours can't be made. In further *m*-dependence investigation, case *m* = 1 is excluded, because of its specific influence on radius and clustering coefficient values.

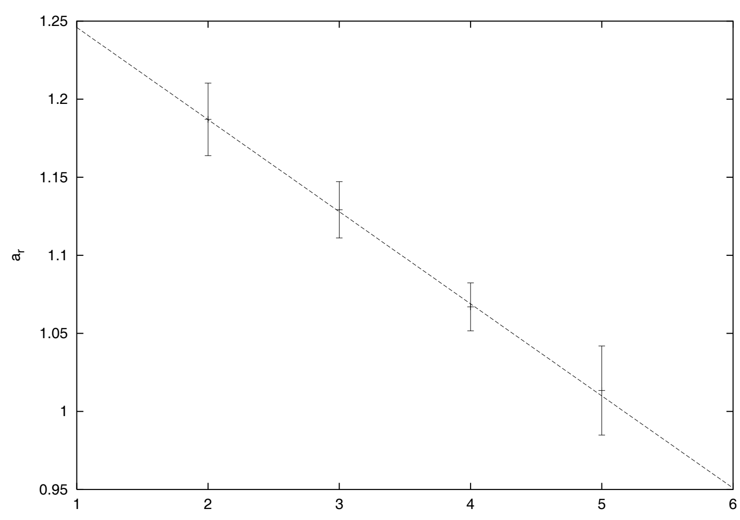

Figure 4: *m*-dependence of a_r coefficient fitted to the function $y = a_{r1} - a_{r2}x$, a_{r1} = 1.305(5), a_{r2} = 0.059(2).

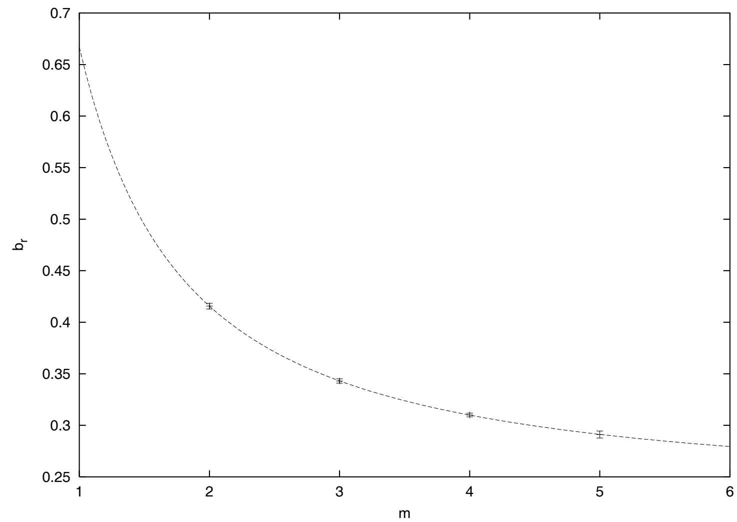

Figure 5: *m*-dependence of b_r coefficient fitted to the function $y = b_{r1} + b_{r2} x^{-b_{r3}}$, b_{r1} = 0.233(2), b_{r2} = 0.434(3), b_{r3} = 1.25(2).

For same network size, connectivity increases with *m*-value growth which causes radius fall. This can also be observed within monotonic decrement of *r*(*m*) function coefficients, where *r* stands for radius. Dependence $a_r(m)$ fits best on linear function $y = a_{r1} - a_{r2} x$ (Figure 4), while $b_r(m)$ fits best on power-law function $y = b_{r1} + b_{r2} x^{-b_{r3}}$ (Figure 5).

For same network size, clustering coefficient increases with *m*-value growth. Still, the inclination of *c*(*m*) dependence (where *c* represents clustering coefficient) on semi-log plot is same for all *m* values. Therefore b_c value can be approximated as constant (Figure 7) while $a_c(m)$ can be fitted to exponential function $y = a_{c1} e^{a_{c2} x}$ (Figure 6).

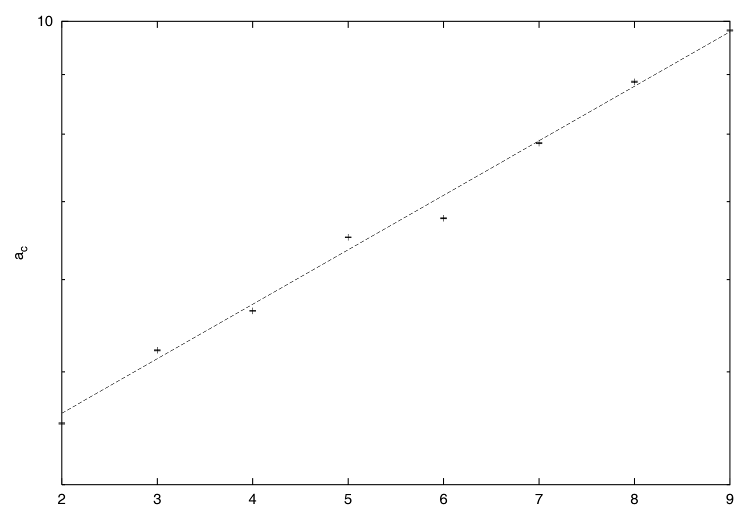

Figure 6: *m*-dependence of a_c coefficient fitted to the function $y = a_{c1} e^{a_{c2} x}$ shown on semi-log plot, a_{c1} = 3.7(1), a_{c2} = 0.108(5).

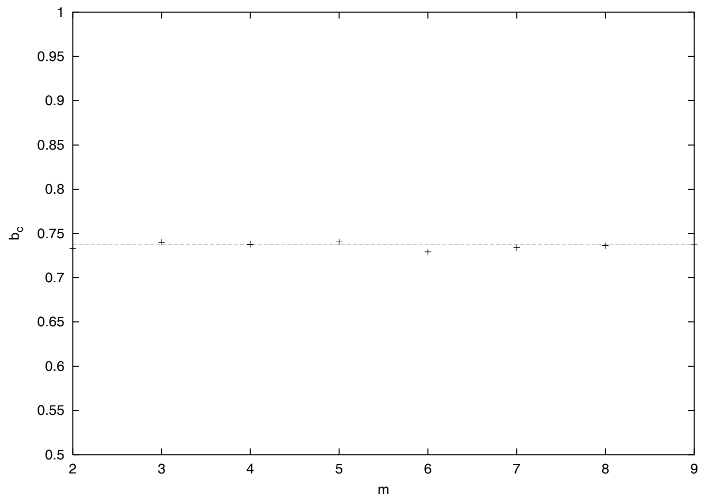

Figure 7: *m*-dependence of b_c fitted to the constant function – obtained value is b_c = 0.737(2).

Scale-free network substructures

While last section investigates global scale-free network characteristics, this section looks at structural properties.

Connected subset of nodes is called a substructure and it characterizes network on local level. Structural properties investigation in this paper is limited to size three substructures. Since all nodes in a substructure have to be connected (directly or indirectly), there are two types of size three substructures, triangle and arc (Figure 8).

Number of triangles occurrences dependence on network size *n* is fitted to the function $y = a_t (\ln x)^{b_t}$ (Figure 9). Number of arcs occurrences dependence on network size *n* is fitted to the function $y = a_k x \ln x$ (Figure 10). Intuition for the types of functions was gained due to paper [3] which deals with substructures within directed networks.

New node attachment for m = 1 can't produce new triangles, therefore in this case number of triangles T is constant for all network sizes and equal to number of triangles in starting random network. Thus, for m = 1, b_t value is equal to zero and there is no point in evaluation of a_t, because its value originates from random properties of starting network and has no relation with preferential attachment nor scale-free concept.

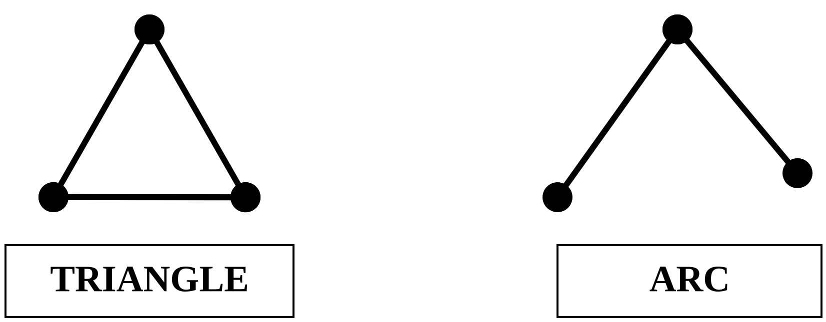

Figure 8: Types of size three substructures – triangle and arc.

Number of links in a network increases with *m*-value growth, therefore number of triangles and arcs grow, so there is an *m*-dependence of these two values. Dependence $a_t(m)$ can be fitted to the power-law function $y = a_{t1} + a_{t2}x^{a_{t3}}$ (Figure 11). Dependence $b_t(m)$ is convergent and can be fitted to the function $y = b_{t1}\left(1 - x^{-b_{t2}}\right)^{b_{t3}}$ (Figure 12). Dependence $a_k(m)$ can be fitted to the power-law function $y = a_{k1} + a_{k2}x^{a_{k3}}$ (Figure 13).

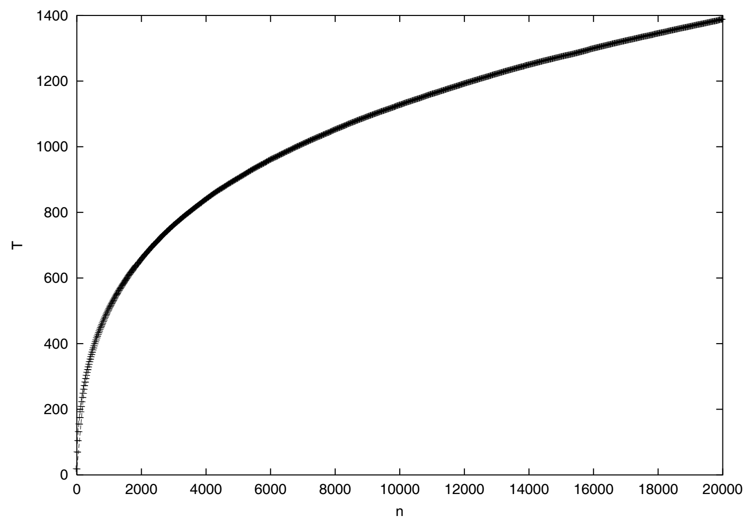

Figure 9: Number of triangles dependence on the network size fitted to the function $y = a_t(\ln x)^{b_t}$, a_t = 2.1788(6), b_t = 2.8152(2). Network parameters are n_0 = 10, p = 0.5, m = 4. Data on the plot are mean values obtained on 100 different networks.

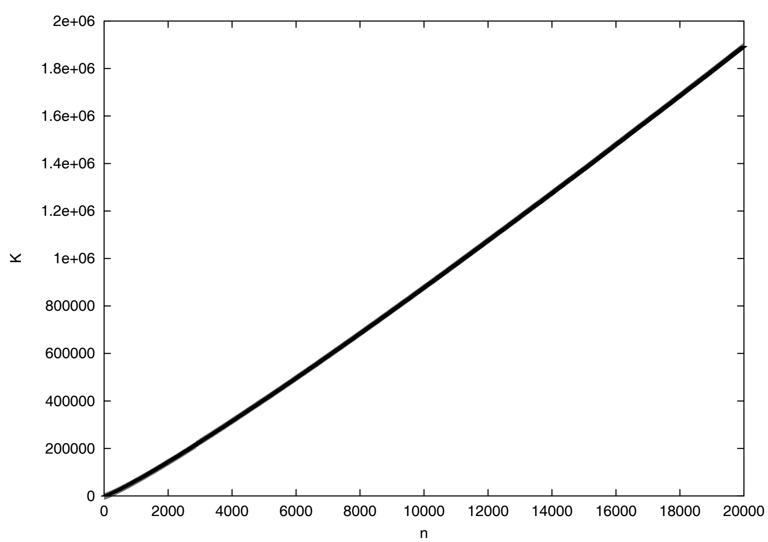

Figure 10: Number of arcs dependence on the network size fitted to the function $y = a_k x \ln x$, a_k = 9.5289(3). Network parameters are n_0 = 10, p = 0.5, m = 4. Data on the plot are mean values obtained on 100 different networks.

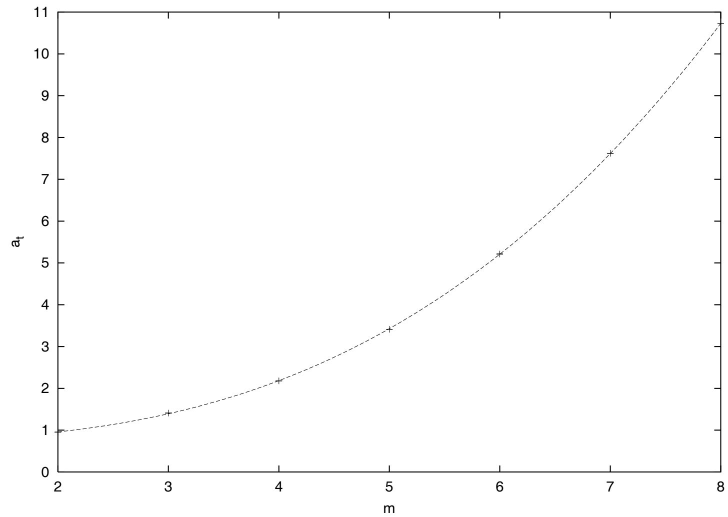

Figure 11: *m*-dependence of a_t coefficient fitted to the function $y = a_{t1} + a_{t2}x^{a_{t3}}$, a_{t1} = 0.75(1), a_{t2} = 0.0295(8), a_{t3} = 2.80(2).

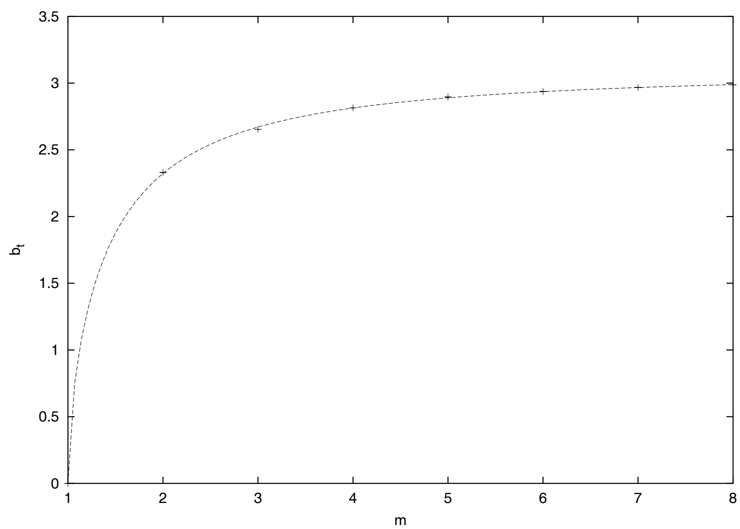

Figure 12: *m*-dependence of b_t coefficient fitted to the function $y = b_{t1}\left(1 - x^{-b_{t2}}\right)^{b_{t3}}$, b_{t1} = 3,10(2), b_{t2} = 1.4(1), b_{t3} = 0.59(5).

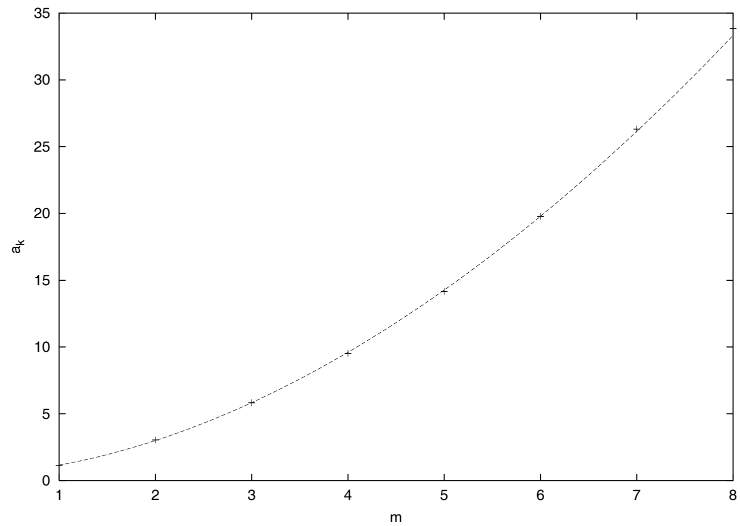

Figure 13: *m*-dependence of a_k coefficient fitted to the function $y = a_{k1} + a_{k2}x^{a_{k3}}$, a_{k1} = 0.44(5), a_{k2} = 0.71(2), a_{k3} = 1.84(2).

Network substructures heuristic estimation

Last section deals with exact evaluation of size three substructure occurrences in scale-free network for variable network size and number of links attached to a new node. Exact evaluation requires counting of all triangles t_i and arcs k_i where *i*-th node is included, for each *i*. Number of all triangles *T* and arcs *K* in the network is evaluated as

$$T = \sum_i \frac{t_i}{3},\ K = \sum_i \frac{k_i}{3}.$$

The question is: Can the number of triangles and arcs in the network be estimated by counting t_i and k_i for only some nodes in the network? This would speed up the evaluation algorithm, especially in the case of large networks such as Internet and World Wide Web.

In a subset *S* of *n*-size network nodes, *q* is defined as $\frac{|S|}{n} = q$. With calculation of t_i and k_i values for each *i*-th node from *S*, following values can be evaluated:

$$T_q = \sum_{i \in S} \frac{t_i}{3},\ K_q = \sum_{i \in S} \frac{k_i}{3}.$$

By definition $T = T_{100\%}$, $K = K_{100\%}$. t_q and k_q quantities are defined as $t_q = \frac{T_q}{T}$, $k_q = \frac{K_q}{K}$. Values of t_q and k_q for various scale-free networks and differently chosen subsets *S* are examined in this section.

In the case of random selection of subset *S* (each node has the same probability 1/*n* to be included into *S*) intuitively $t_q = k_q = q$, which is confirmed with following results. Values T_q / q and K_q / q can be used as an estimation of values *T* and arcs *K*, which is shown on figures 14 and 15. Even though in most of the cases number of triangles/arcs varies from exact value for only a few percent, there are cases with greater distinction. Table 1 shows the maximal difference between the values T_q / q and *T*, as well as K_q / q and

K, for variable values of q. Estimations are made for 10^4-size network with parameters n_0 = 10, p = 0.5, m = 1, 2,..., 8. Values represent mean value of 25 different results, cited with standard deviation.

q	20%	40%	60%	80%
Triangle number distinction	15%	10%	5%	4%
Arc number distinction	10%	6%	5%	4%

Table 1: Maximal error for numbers of triangles and arcs using random method estimation for different percentages q.

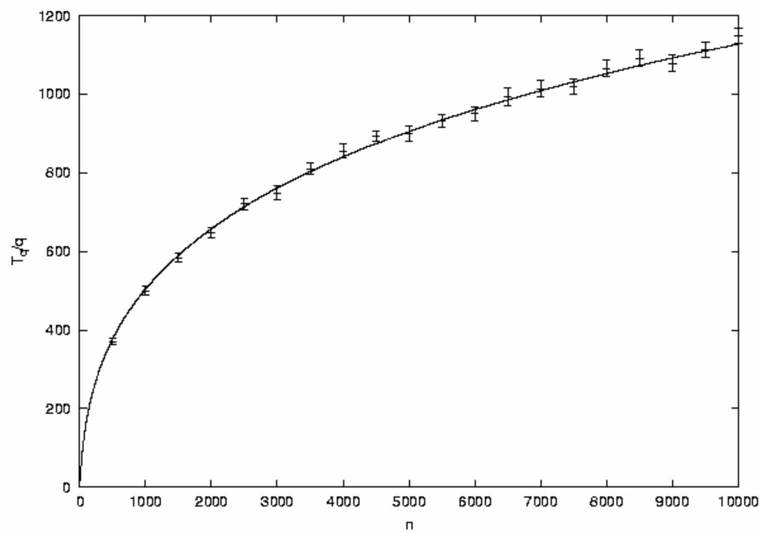

Figure 14: T_q / q dependence on network size, for q = 80% randomly chosen nodes, is shown on the plot, as well as exact dependence $T(n)$. Network parameters are n_0 = 10, p = 0.5, m = 4. T_q values are mean values obtained on 25 different networks.

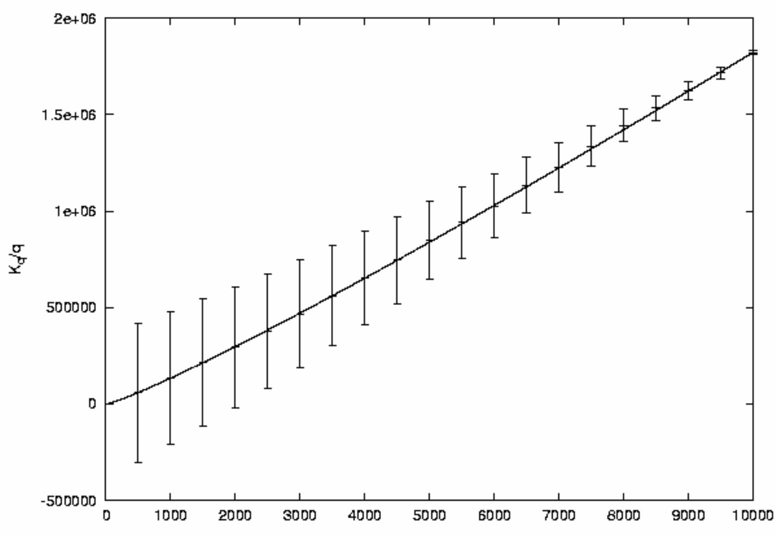

Figure 15: K_q / q dependence on network size, for q = 60% randomly chosen nodes, is shown on the plot, as well as exact dependence $K(n)$. Network parameters are n_0 = 10, p = 0.5, m = 6. K_q values are mean values obtained on 25 different networks.

A better approach can be made by taking into account scale-free network degree distribution. Subset *S* is chosen preferentially (*i*-th node has probability proportional to its degree s_i to be included into *S*). There is no intuitive approach to estimation of values t_q and k_q, thus they were investigated.

Results from last section show that *n*-dependence of number of triangles/arcs, *T*(*n*) and *K*(*n*), can be fitted to the functions $y = a_t (\ln x)^{b_t}$ and $y = a_k x \ln x$. Further on, obtained values of coefficients a_t, b_t i a_k for various network parameters are considered determined. In order to obtain t_q and k_q values for preferential selection of subset *S*, dependences $T_q(n)$ and $K_q(n)$ are fitted to the functions $y = t_q a_t (\ln x)^{b_t}$ and $y = k_q a_k x \ln x$, where the values of coefficients a_t, b_t i a_k are taken from previous section.

To illustrate superiority of estimation with preferential selection, figures 16 and 17 show *n*-dependences of $\frac{T_q}{t_q}$ and $\frac{K_q}{k_q}$ for the same parameters as figures 14 and 15, which show estimations for random selection.

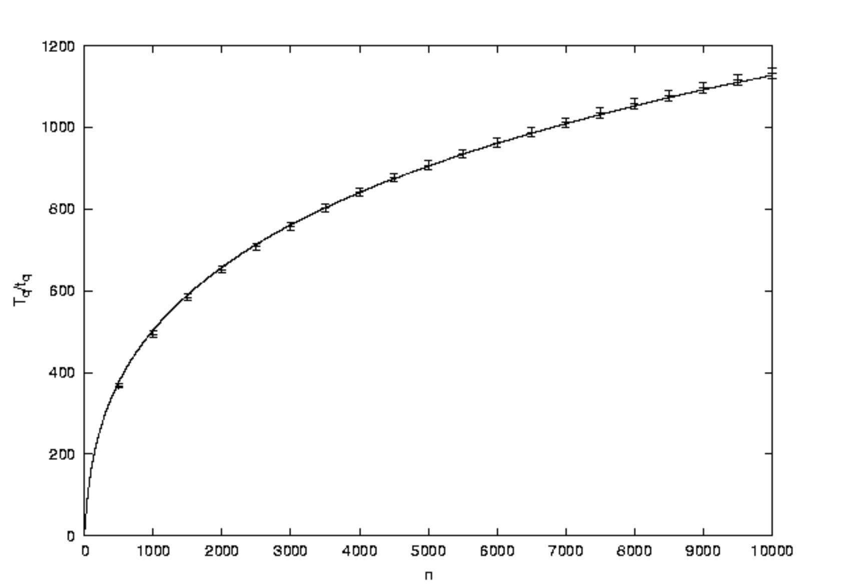

Figure 16: $\frac{T_q}{t_q}$ dependence on network size, for q = 80% randomly chosen nodes, is shown on the plot, as well as exact dependence *T*(*n*), t_q = 0.958(2). Network parameters are n_0 = 10, p = 0.5, m = 4. T_q values are mean values obtained on 25 different networks.

Within a network with fixed parameters *q*-dependences of t_q and k_q are linear on log-log plot and thus can be fitted to the functions $y = \alpha_t x^{\beta_t}$ and $y = \alpha_k x^{\beta_k}$ (Figures 18 and 19). Values of $\alpha_t, \beta_t, \alpha_k, \beta_k$ coefficients for different values of *m* are given within table 2.

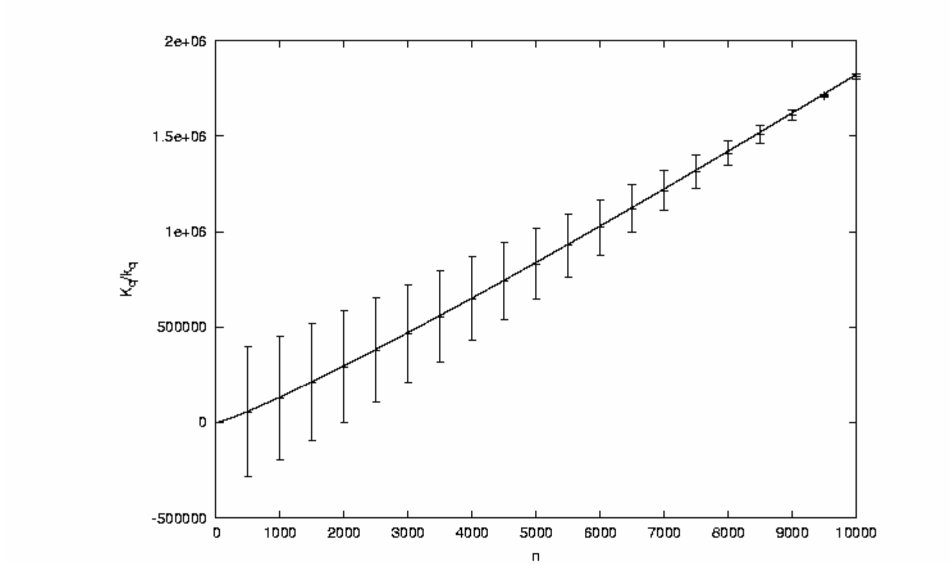

Figure 17: $\frac{K_q}{k_q}$ dependence on network size, for *q* = 60% randomly chosen nodes, is shown on the plot, as well as exact dependence *K*(*n*), k_q = 0.7855(1). Network parameters are n_0 = 10, *p* = 0.5, *m* = 6. K_q values are mean values obtained on 25 different networks.

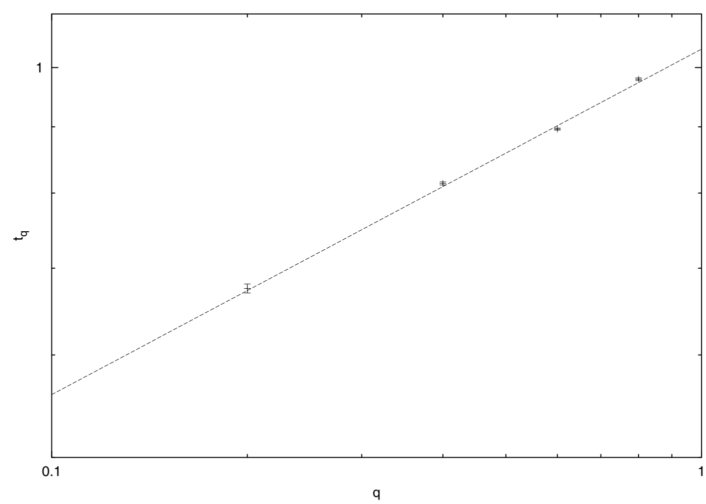

Figure 18: *q*-dependence of t_q coefficient fitted to the function $y = \alpha_t x^{\beta_t}$, α_t = 1.03(1), β_t = 0.27(2). Network parameters are n_0 = 10, *p* = 0.5, *m* = 3.

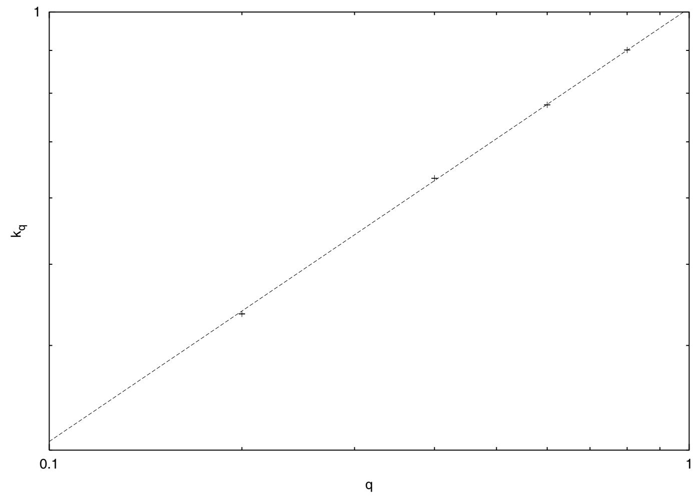

Figure 19: q-dependence of k_q coefficient fitted to the function $y = \alpha_k x^{\beta_k}$, α_k = 1.010(3), β_k = 0.517(7). Network parameters are n_0 = 10, p = 0.5, m = 7.

m	αt	βt	αk	βk
1	-	-	1.01(2)	0.47(2)
2	1.004(6)	0.224(7)	0.992(3)	0.469(3)
3	1.03(1)	0.27(3)	1.012(2)	0.489(6)
4	1.017(7)	0.28(2)	0.999(5)	0.49(1)
5	1.025(6)	0.27(2)	1.015(2)	0.506(3)
6	1.022(6)	0.28(2)	1.012(2)	0.514(2)
7	1.010(6)	0.28(2)	1.010(3)	0.517(7)
8	1.007(6)	0.27(2)	1.001(3)	0.49(2)

Table 2: $\alpha_t, \beta_t, \alpha_k, \beta_k$ values for different values of m parameter.

Summary

Global and local properties of scale-free networks are investigated within this paper. Barabasi-Albert algorithm, with m-degree new nodes, is used for generation of networks with power-law (scale-free) degree distribution.

Small-world and clustering coefficient dependences on network size are expanded to dependences on m-parameter.

Dependence of number of size three substructures (triangles and arcs) occurrences in the network on network size and m-parameter are found.

Counting of triangles and arcs in randomly and preferentially selected subset are examined as heuristics for estimation of number of triangles and arcs in the whole network. Random selection is found as simpler, but less precise method. Formulas and coefficients for the method of preferential selection are found for m = 1,2,...,8 and an arbitrary size of subset S.

These heuristics are especially useful for local characterization of large networks, because of the time reduction.

Acknowledgments

I would like to express my gratitude to the Petnica Science Center, and especially to Ms. Jelena Grujić, head of physics seminar – as a part of which I worked on this project. I am very grateful to my mentor Dr Aleksandar Belić for his useful advice, and to the Scientific Computing Laboratory for making the most sophisticated computer equipment available to me for the needs of my research. I would also like to thank Miloš Srećković and Miloš Savić for their useful comments and the discussions we had regarding scale-free networks.

References

1. Reka Albert[1], Albert-Laszlo Barabasi[2], 2000, *Statistical Mechanics of Complex Networks*, [1]School of Mathematics,127 Vincent Hall, University of Minnesota, Minneapolis, Minnesota 55455, [2]Department of Physics, College of Science, 225 Nieuwland Science Hall, University of Notre-Dame, Notre-Dame, Indiana 46556

2. S. S. Manna, 2003, *Diffusion limited friendship network: A model for six degrees of separation*, Satyendra Nath Bose National Centre for Basic Sciences Block-JD, Sector-III, Salt Lake, Kolkata-700098, India

3. Danilo Sergi, 2005, *Random graph model with power-law distributed triangle subgraphs*, Departement de Physique Theorique, Universite de Geneve, 1211 Geneve 4, Switzerland