Universal scaling of inter-node distances in complex networks

Janusz A. Hołyst*, Julian Sienkiewicz, Agata Fronczak, Piotr Fronczak, Krzysztof Suchecki and Piotr Wójcicki

August 31, 2005

Abstract

We have studied dependence of distances between nodes in various networks and degrees of such vertices. We have observed that the mean distance between two nodes of degrees k_i and k_j equals to $\langle l_{ij}\rangle = A - B\log(k_i k_j)$. The relation holds for the following systems: Erdős-Rényi random graphs, scale-free Barabási-Albert models, science collaboration networks, biological networks, Internet Autonomous Systems and public transport networks. A simple heuristic theory for this scaling law is presented. Corrections due to the network clustering coefficient and node degree-degree correlations are taken into account.

Keywords: complex networks, distances, scaling, correlations

In [1, 2] an analytical model for average path lengths in random uncorrelated networks was considered and it was shown that the shortest path length between nodes i and j possessing degrees k_i and k_j can be described as

$$l_{ij}(k_i,k_j) = \frac{-\ln k_i k_j + \ln\left(\langle k^2\rangle - \langle k\rangle\right) + \ln N - \gamma}{\ln\left(\langle k^2\rangle/\langle k\rangle - 1\right)} + \frac{1}{2}, \tag{1}$$

where $\gamma = 0.5772$ is the Euler constant, whereas $\langle k\rangle$ and $\langle k^2\rangle$ correspond to the first and the second moments of node degree distribution $P(k)$. It follows that the mean distance between two nodes is lineary dependent on the logarithm of their degree product

$$\langle l_{ij}\rangle = A - B\log(k_i k_j). \tag{2}$$

Below we show that the relation (2) can also be obtained from a simple model of branching trees exploring the space of a random network [3, 4] (see Fig. 1). Let us consider a path from a randomly chosen node i to a randomly chosen node j in such a network. Following a random direction of a randomly chosen edge one approaches a node j with the probability $p_j = k_j/(2E)$, where $2E = N\langle k\rangle$ is a double number of links. It means that in average one needs $M_j = 1/p_j = 2E/k_j$ of random trials to arrive at the node j. Now let us consider a branching process represented by the tree T_i (see Fig. 1) that starts at the node i where an average branching factor is κ (all loops are neglected). If the distance between the node i and the surface of the tree equals to x then in average there are $N_i = k_i\kappa^{x-1}$ nodes at such a surface and there is the same number of links ending at these nodes. It follows that in average the tree T_i touches the node j when $N_i = M_j$ i.e. when

$$k_i k_j \kappa^{x-1} = N\langle k\rangle. \tag{3}$$

*Corresponding author: Faculty of Physics, Warsaw University of Technology, ul. Koszykowa 75 , 00-662 Warszawa, *Tel.:* +48 22 6607133; *fax:* +48 22 6282171; *email address:* jholyst@if.pw.edu.pl

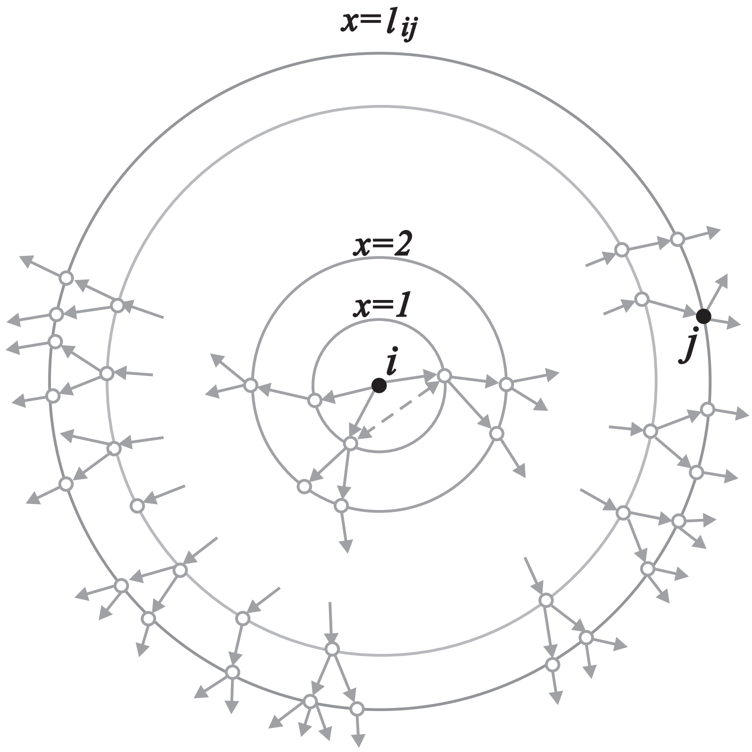

Figure 1: Tree formed by a random process, starting from the node i and approaching the node j.

Since the mean distance from the node i to the node j is $\langle l_{ij} \rangle = x$ thus we get the scaling relation (2) with

$$A = 1 + \frac{\log(N\langle k \rangle)}{\log \kappa} \quad \text{and} \quad B = \frac{1}{\log \kappa}. \tag{4}$$

The result (4) is in agreement with the paper [5] where the concept of generating functions for random graphs with arbitrary degree distributions has been used.

One has to take into account that in the above considerations we have assumed that there are no degree-degree correlations, we have neglected all loops and we have treated the branching level x as a continuum variable to fulfill the relation (3). Assuming that the branching factor κ can be expressed as $\langle k^2 \rangle / \langle k \rangle - 1$ [6], one can see that the differences between the results (1) and (4)are small, at least for the case when $N \to \infty$ and κ is finite.

Fig. 2 present mean distances $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$ for the following systems: Erdős-Rényi random graphs, Barabási-Albert evolving networks, biological networks [7, 8, 9], social networks [10, 11], Internet Autonomous Systems [12] and selected networks for public transport in Polish cities [14, 15]. The relation (2) is very well observed over several decades of data points, although among the systems mentioned there are both scale-free and single scale networks, with either negligible or very high clustering coefficient, assortative [16], disassortative or uncorrelated.

Although the scaling (2) works well for distances averaged over all pairs of nodes specified by a given product $k_i k_j$, there can be large differences if one changes k_i while keeping $k_i k_j$ constant. The Fig. 3 presents the dependence of average path length $\langle l_{ij} \rangle$ on k_i, for a fixed product $k_i k_j$ in the case of several networks from different classes. One can see that although the *Astro* network is assortative (short-range

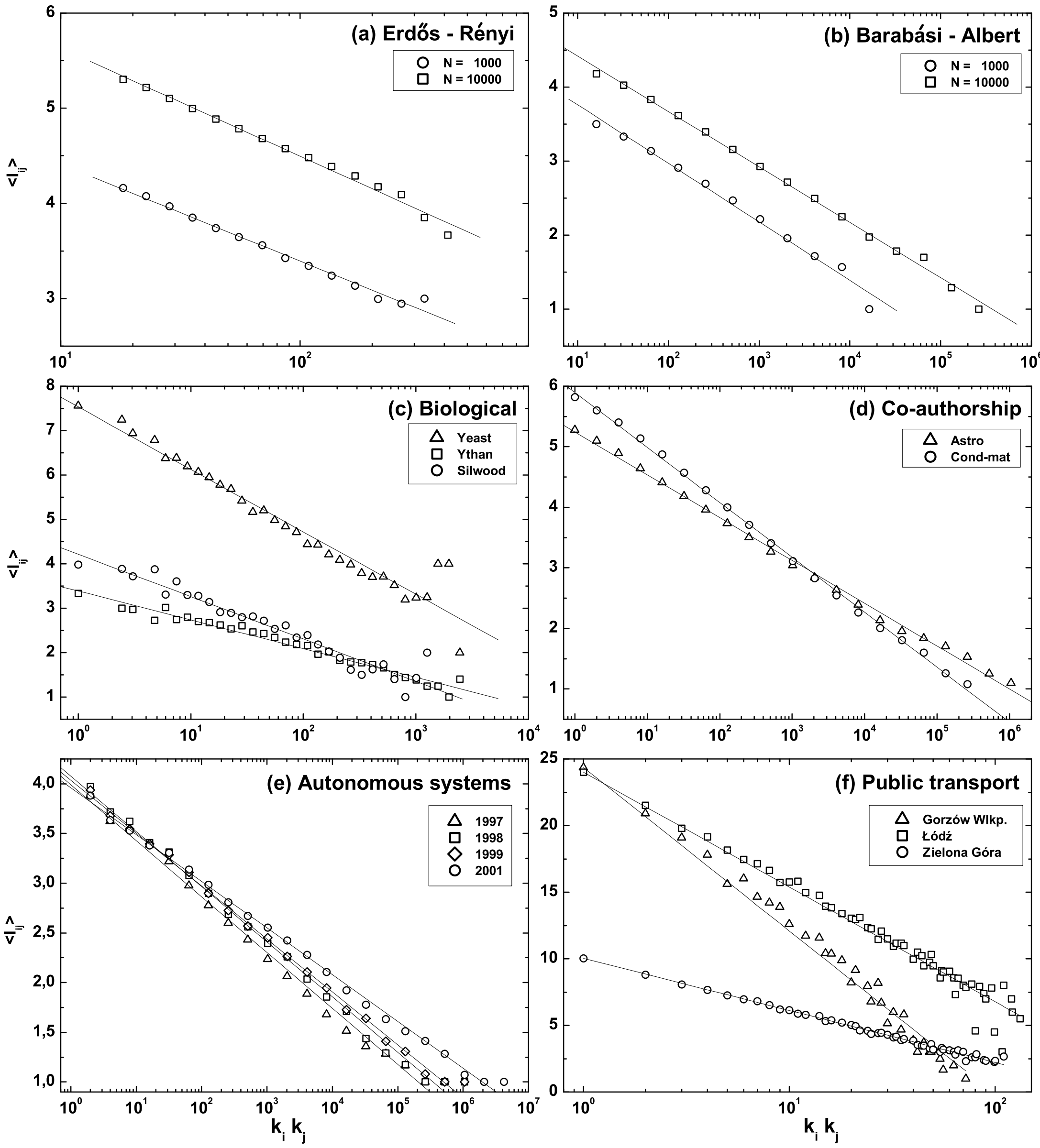

Figure 2: Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. **(a)** *Erdős-Rényi* random graphs: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares), **(b)***Barabási-Albert* networks: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares), **(c)** Biological networks: *Silwood* (circles), *Yeast* (triangles), *Ythan* (squares), **(d)** Co-authorship networks: *Astro* (triangles), *circles* (circles), **(e)** Internet Autonomous Systems: *Year 1997* (triangles), *Year 1998* (squares) *Year 1999* (diamonds), *Year 2001* (circles), **(f)** Public transport networks in Polish cities: *Gorzów Wlkp.* (triangles), *Łódź* (squares), *Zielona Góra* (circles) In **(a)**, **(b)**, **(d)** and **(e)** data are logarithmically binned with the power of 2, in case of **(c)** with the power of 1.25 and in case of **(f)** data are not binned.

attraction), pairs of nodes with similar degrees are in average further away than different degree pairs (long-range repulsion). For the disassortative network AS [16] the behavior is opposite. For uncorrelated networks (Erdős-Rényi, Barabási-Albert), the average path length is constant given the product $k_i k_j$ fixed [4].

In [3] we compare (4) to results from real networks and numerical simulations. In fact our approximate approach (4) fits very well to random Erdős–Rényi graphs and BA models but the corresponding coefficients A and B for real networks are different from results of our simple theory. The formulas (4) can be improved by taking into account effects of loops and node degree-degree correlations.

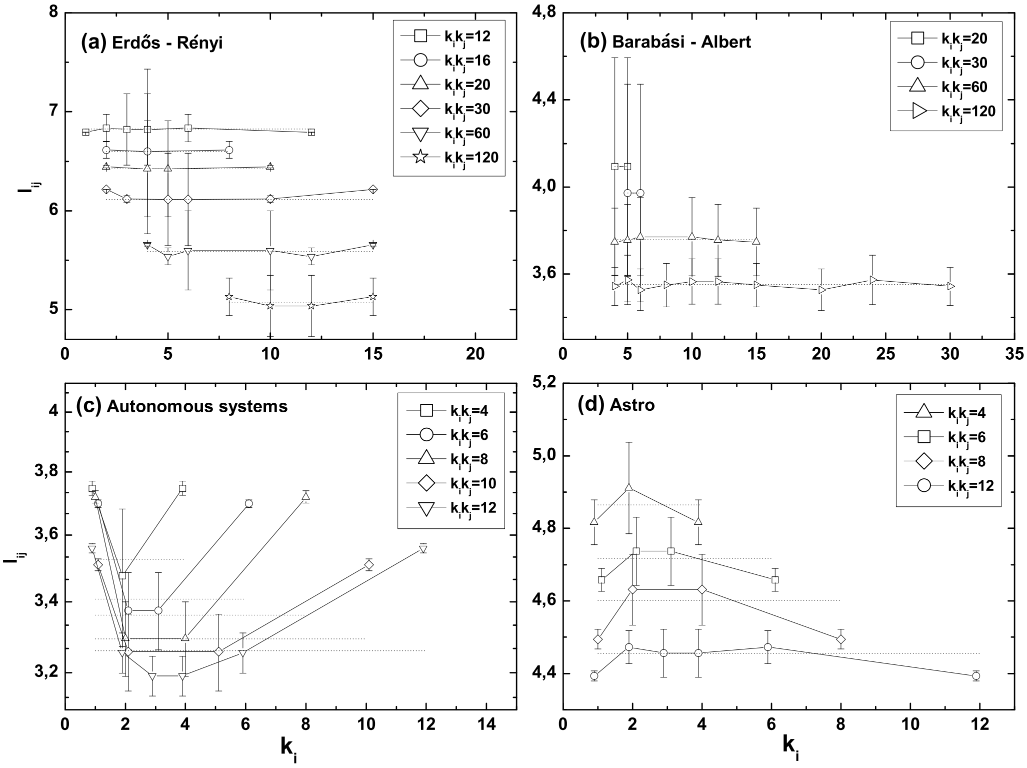

Figure 3: Dependence of average path length on k_i, for fixed $k_i k_j$ product. The lines connecting the symbols are there for clarity. The bars show point weight, meaning relative numbers of pairs ij. The horizontal lines are weighted averages over k_i and represent average path lengths for the given product $k_i k_j$. Note: the very small shifts on k_i axis between data for different $k_i k_j$ are artificially introduced to make the weight bars not overlap.

The influence of loops of the length three can be estimated as follows [3]. Let us assume that in the branching process forming the tree T_i two nodes from the nearest neighborhood of the node i are *directly* linked (the dashed line at Fig.1). Such a situation can occur at any point of the branching tree T_i and corresponding links are useless for further network exploration by the tree T_i. It follows the *effective* contribution from both connected nodes to the mean branching factor of the tree T_i is decreased. Assuming that clustering coefficients of every node are the same, the corrected factor for the branching process equals to $\kappa_c = \kappa - c\kappa$ where c is the network clustering coefficient. This equation is not valid for the branching process around the node i where $\kappa'_i = \kappa - c(k_i - 1)$. A similar situation arises around the node j. Replacing k_i and k_j with $\langle k \rangle$ in κ'_i and κ'_j one gets

$$k_i k_j [\kappa(1-c')]^2 [\kappa(1-c)]^{x-3} = N\langle k\rangle, \tag{5}$$

where $c' = c(\langle k\rangle - 1)/\kappa$. It follows that instead of (4) we have

$$A' = 3 + \frac{\log(N\langle k\rangle) - 2\log[\kappa(1-c')]}{\log[\kappa(1-c)]}, \quad B' = \frac{1}{\log[\kappa(1-c)]}. \tag{6}$$

Now, let us consider the presence of degree correlations [4]. Such correlations mean that average degrees $k_i^{(nn)}$ of nodes in the neighborhood of a node i depend on the degree k_i. Let us assume that this relation can be written as

$$\kappa_i \equiv k_i^{(nn)} - 1 = D k_i^{\phi-1} \tag{7}$$

If ϕ is larger than one then the network is assortative, i.e. high degree nodes are mostly connected to other high degree nodes and similarly low degree nodes are connected to other low degree nodes. Such a situation occurs for example in networks describing scientific collaboration [16]. If ϕ is smaller than one, then the network is disassortative and high degree nodes are mostly connected with low degree nodes what is typical for the Internet Autonomous Systems [16]. If we neglect higher order correlations then Eq.3 should be replaced by

$$k_i k_j \kappa_i \kappa_j \kappa^{x-3} = N\langle k\rangle \tag{8}$$

Taking into account Eq. 7 we can replace parameters A and B given by the Eq. 4 with

$$A_\phi = A + 2 - 2B\log D \quad \text{and} \quad B_\phi = \phi B \tag{9}$$

In conclusions we have observed a universal path length scaling for different classes of real networks and models. The mean distance between nodes of degrees k_i and k_j is a linear function of $\log(k_i k_j)$. The scaling holds over many decades and does not depend on network degree distributions, clustering coefficients or degree-degree correlations. We have showed that a simple model of random tree exploring the network explains such a scaling behavior. In an extended version of the model clustering effects and first order degree-degree correlations have been introduced to improve theoretical predictions for real world systems. In our opinion a better agreement between theoretical results and experimental data may be obtained taking into account higher order correlations.

Authors acknowledge the State Committee for Scientific Research in Poland for financial support under grant No. 134/E-365/6.PRUE/DIE239/2005-2007. This work has also been partially supported by European Commission Projects: CREEN FP6-2003-NEST-Path-012864 and MMCOMNET FP6-2003-NEST-Path-012999.

References

[1] Fronczak A, Fronczak P, Hołyst JA: How to calculate the main characteristics of random uncorrelated networks. E-print 2005: cond-mat/0502663.

[2] Fronczak A, Fronczak P, Hołyst JA: Average path length in random networks. Phys. Rev. E 2004; 70, 056110.

[3] Hołyst JA, Sienkiewicz J, Fronczak A, Fronczak P, Suchecki K: Universal scaling of distances in complex networks. Phys. Rev. E 2005; 72, 026108.

[4] Hołyst JA, Sienkiewicz J, Fronczak A, Fronczak P, Suchecki K: Scaling of distances in correlated complex networks. Physica A 2005; 351, 167.

[5] Motter AE, Nishikawa T, Lai YC: Range-based attack on links in scale-free networks: Are long-range links responsible for the small-wrold phenomenon? Phys. Rev. E 2002; 66, 065103(R).

[6] Newman MEJ, Strogatz SH, Watts DJ: Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E 2001; 64, 026118.

[7] Data for Foodwebs (*Silwood* and *Ythan*) [8] have been taken from the website http://www.cosin.org /extra/data/foodwebs/ whereas data for protein interaction network (*Yeast*) [9] has been taken from http://www.nd.edu/ networks/database/index.html.

[8] Garlaschelli D, Caldarelli G, Pietronero L: Universal scaling relations in food webs. Nature 2003; 423, 165.

[9] Jeong H, Mason SP, Barabási A-L, Oltavi ZN: Lethality and centrality in protein networks. Nature 2001; 411, 41.

[10] Scientific collaboration network [11] data have been collected from two publicly available databases of papers http://arxiv.org/archive/astro-ph (*Astro*) and arxiv.org/archive/cond-mat (*Cond-mat*) for the period 1995-2001.

[11] Newman MEJ: Scientific collaboration networks. I. Network construction and fundamental results. Phys. Rev. E 2001; 64, 016131.

[12] Data for the Internet [13] has been taken from www.cosin.org/extra/data/internet/.

[13] Pastor-Satorras R, Vespignani A: Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge Univ. Press 2004.

[14] In those networks vertices are bus- and tramstops while an edge exists if at least one public transport line crosses two stops. [15].

[15] Sienkiewicz J, Hołyst JA: Statistical anlysis of 22 public transport networks in Poland. E-print 2005: physics/0506074.

[16] Newman MEJ: Assortative Mixing in Networks. Phys. Rev. Lett. 2002; 89, 208701.

[17] Watts DJ, Strogatz SH: Collective dynamics of 'small-world' networks. Nature 1998; 393, 440.