Universal dependence of distances on nodes degrees in complex networks

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

Faculty of Physics and Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland

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

Keywords: complex networks, distances, scaling, correlations
PACS: 89.75.-k, 02.50.-r, 05.50.+q

The empirical analysis of many real complex networks (for a review see [1, 2, 3, 4]) has revealed the presence of several universal scaling laws. The best known scaling law appears for degree distributions $P(k) \sim k^{-\gamma}$ [5] and it is observed in a number of social, biological and technological systems. May other scaling laws in complex networks have been found, such as a dependence of clustering coefficient on node degree in hierarchical networks $c(k) \sim k^{-\alpha}$ [6], scale-free behavior of the connection weight [7, 8] and load [9] distributions, load dependence on degree [10] and others [11, 12, 13].

In [14] 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 a 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 [16, 17].

Let us notice that following a random direction of a randomly chosen edge one approaches node j with 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

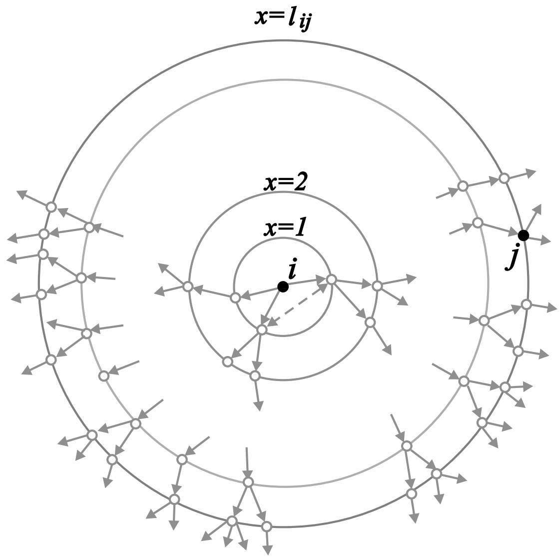

FIGURE 1. Tree formed by a random process, starting from the node i and approaching the node j.

(Fig. 1) that starts at the node i where an average branching factor is κ (all loops are neglected). If a 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

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

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 [18] where the concept of generating functions for random graphs 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 $\langle k^2\rangle/\langle k\rangle - 1$ [3], one can see that the differences between the results (4) and (1) are small, at least for the case when $N \to \infty$ and κ is finite. In general the mean branching factor κ is a mean value over all local branching factors and over all trees in the network. In the first approximation for networks without degree-degree correlations [3] it could be estimated as the mean arithmetic value of a nearest neighbor degree less one (incoming edge): $\kappa = \langle k\rangle_{nn} - 1$. Such a mean value is however not exact because local branching factors in every tree are *multiplied* one by another in (3). The corrected mean value of κ should be taken as an arithmetic mean value over all geometric values from different trees, what is very difficult to perform numerically. To reduce this discrepancy we calculated arithmetic mean branching factor

TABLE 1. Basic properties of examined systems and comparison between experimental and theoretical data. *Astro* and *Cond-mat* are co-authorship networks, *Silwood*, *Yeast* and *Ythan* are biological networks. The number after the Internet Autonomous Systems means the year data were gathered, *Gorzów Wlkp.*, *Łódź* and *Zielona Góra* are public transport networks in corresponding Polish cities. N is the number of nodes, $\langle k \rangle$ - mean degree value. A_e and B_e are mean experimental values (Fig. 2-4) whereas A and B are given by (4).

network	N	$\langle k \rangle$	A_e	A	B_e	B
Erdős-Rényi random graph	1000	8.00	5.43	5.46	1.017	1.143
Erdős-Rényi random graph	10000	8.00	6.77	6.60	1.136	1.143
Barabási-Albert model	1000	8.00	4.54	4.24	0.813	0.830
Barabási-Albert model	10000	8.00	5.17	4.81	0.778	0.777
Astro	13986	25.56	5.24	4.30	0.707	0.595
Cond-mat	17013	9.46	5.90	5.09	0.908	0.786
Silwood	153	4.77	4.22	3.69	0.955	0.941
Yeast	1846	2.39	7.53	6.66	1.406	1.552
Ythan	135	8.83	3.39	3.35	0.649	0.765
Internet Autonomous Systems 1997	3015	3.42	3.99	3.39	0.562	0.596
Internet Autonomous Systems 1998	4180	3.72	4.08	3.41	0.555	0.575
Internet Autonomous Systems 1999	5861	3.86	4.03	3.35	0.532	0.540
Internet Autonomous Systems 2001	10515	4.08	3.96	3.23	0.471	0.481
Gorzów Wlkp.	269	2.48	24.36	16.06	12.270	5.333
Łódź	1023	2.83	24.01	11.67	8.621	3.084
Zielona Góra	312	2.98	10.03	8.96	3.908	2.682

over nearest neighborhood of every node m, i.e. $\kappa^{(m)} = \langle k \rangle_{nn}^{(m)} - 1$, and then averaged it geometrically over all nodes m, i.e. $\kappa = \langle \kappa^{(m)} \rangle_m$.

Figs. (2-4) present mean distances $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$ in selected complex networks. We include data for Erdős-Rényi random graphs, Barabási-Albert evolving networks, biological networks [19, 20, 21] (*Silwood*, *Ythan*, *Yeast*), social networks [22, 23] (co-authorship groups *Astro* and *Cond-mat*), Internet Autonomous Systems [24] and selected networks for public transport in Polish cities [25, 26] (Gorzów Wlkp., Łódź, Zielona Góra) (see Table 1. for characteristic parameters of these networks) . One can observe, that the relation (2) is very well fulfilled over several decades for all our data. Let us stress that the networks mentioned above display a wide variety of basic characteristics. Among them there are both scale-free and single scale networks, with either negligible or very high clustering coefficient, assortative [27], disassortative or uncorrelated. The only apparent common feature of all above systems is the small-world effect. We have checked however that for the small-world Watts-Strogatz model [28], the scaling (2) is nearly absent and it is visible only for large rewiring probability, and only for highly connected nodes.

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. 5 presents the dependence of average path length l_{ij} 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 (see Table 2) (short-range attraction), pairs of nodes with similar degrees are in average further away than different

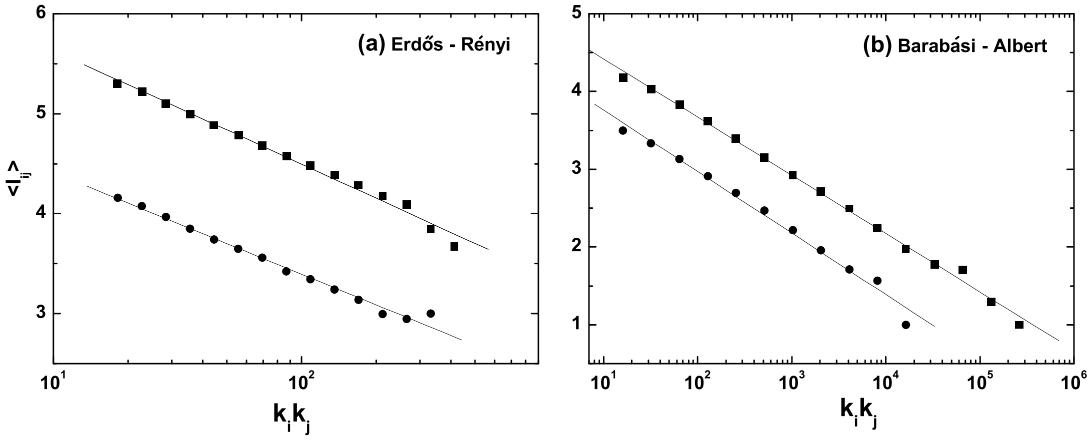

FIGURE 2. Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. **(a)** *Erdős-Rényi* random graphs: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares), **(b)***Barabási-Albert* networks: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares). Data are logarithmically binned with the power of 2.

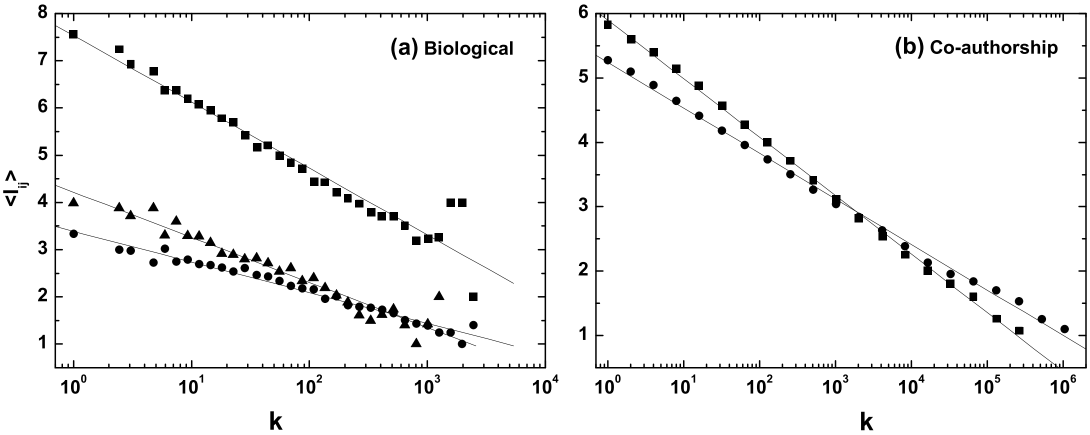

FIGURE 3. Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. **(a)** Biological networks: *Silwood* (triangles), *Yeast* (squares), *Ythan* (circles), **(b)** Co-authorship networks: *Astro* (circles), *Cond-mat* (squares). In (a) Data are logarithmically binned with the power of 1.25 and in (b) with the power of 2.

degree pairs (long-range repulsion). For the disassortative network AS [27] the behavior is opposite. For uncorrelated networks (Erdős-Rényi, Barabási-Albert), the average path length is constant if the product $k_i k_j$ is fixed [16].

Table 2 shows results for parameters A and B from real networks and from numerical simulations as compared expressions (4). One can see that 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. In fact, the relations (4) can be improved by taking into account effects of loops and

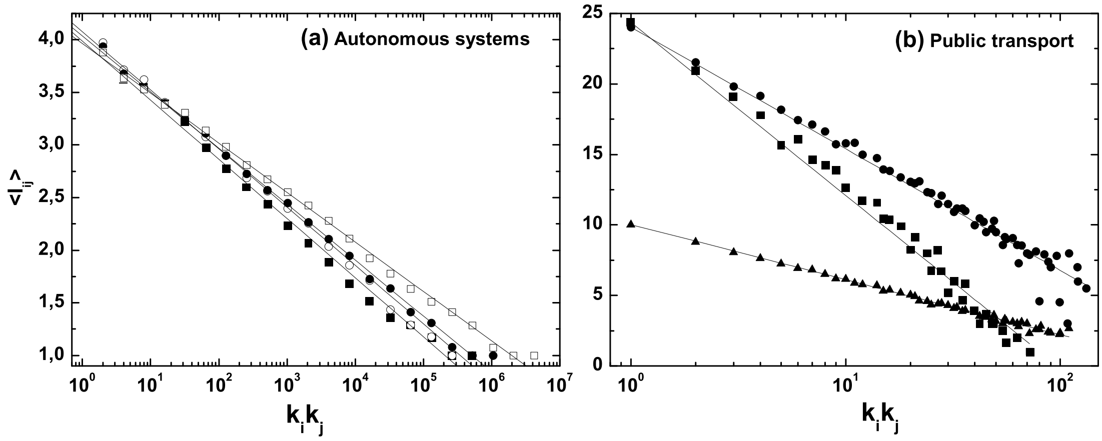

FIGURE 4. Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. **(a)** Internet Autonomous Systems: *Year 1997* (filled squares), *Year 1998* (empty circles) *Year 1999* (full circles), *Year 2001* (empty squares), **(b)** Public transport networks in Polish cities: *Gorzów Wlkp.* (squares), *Łódź* (circles), *Zielona Góra* (triangles) In (a) data are logarithmically binned with the power of 2. In (b) data are not binned.

TABLE 2. Comparison between experimental and theoretical data. ER stands for Erdős-Rényi random graph network, BA for Barabási-Albert model and AS is an acronym for Internet Autonomous Systems, c is network's clustering coefficient, r - assortativity coefficient and ϕ is the scaling exponent for degree correlations. A_e and B_e are mean experimental values (Fig. 2-4) whereas A' and B' are given by (6), while A_ϕ and B_ϕ follow (9).

network	c	r	ϕ	A_e	A'	A_ϕ	B_e	B'	B_ϕ
ER $N = 10^3$	0.007	0	-	5.43	5.48	-	1.017	1.147	-
ER $N = 10^4$	0.001	0	-	6.77	6.61	-	1.136	1.143	-
BA $N = 10^3$	0.038	0	-	4.54	4.27	-	0.813	0.842	-
BA $N = 10^4$	0.007	0	-	5.17	4.81	-	0.778	0.779	-
Astro	0.609	0.055	1.23	5.24	4.98	4.41	0.707	0.786	0.732
Cond-mat	0.604	0.053	1.19	5.90	6.38	5.05	0.908	1.150	0.935
Silwood	0.142	-0.316	0.71	4.22	3.78	3.19	0.955	1.004	0.668
Yeast	0.068	-0.158	0.59	7.53	6.87	5.71	1.406	1.629	0.916
Ythan	0.216	-0.254	0.61	3.39	3.45	2.81	0.649	0.832	0.466
AS 1997	0.182	-0.229	0.46	3.99	3.42	2.58	0.562	0.629	0.274
AS 1998	0.250	-0.200	0.48	4.08	3.45	2.65	0.555	0.620	0.276
AS 1999	0.250	-0.183	0.49	4.03	3.38	2.55	0.532	0.579	0.265
AS 2001	0.289	-0.185	0.45	3.96	3.25	2.50	0.471	0.518	0.217
Gorzów Wlkp.	0.082	0.385	1.44	24.36	19.76	16.67	12.270	6.651	7.679
Łódź	0.065	0.070	1.19	24.01	12.70	11.89	8.621	3.389	3.670
Zielona Góra	0.067	0.238	1.41	10.03	9.63	9.62	3.908	2.917	3.781

degree-degree correlations.

The influence of loops of the length three can be estimated as follows. 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). In fact such a situation

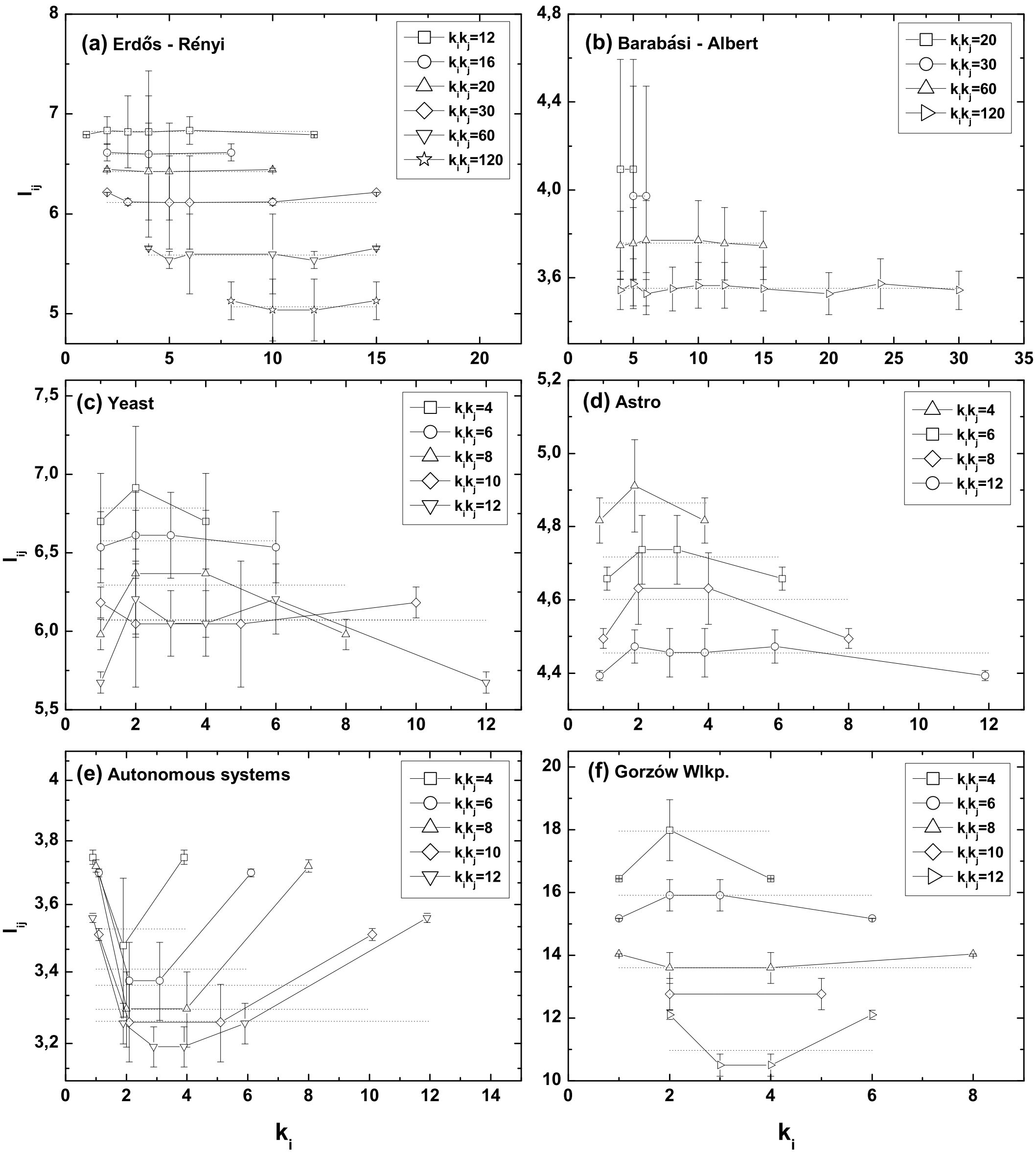

FIGURE 5. Dependence of average path length on k_i, for fixed k_ik_j product. The lines connecting the symbols are there for clarity. The bars show point weight, meaning relative numbers of pairs ij. The horizontal lines are weighted averages over k_i, and are just average path lengths for given k_ik_j. Note: The very small shifts on k_i axis between data for different k_ik_j are artificially introduced to make the weight bars not overlap.

can occur at any point of the branching tree T_i. Since such links are useless for further network exploration by the tree T_i thus an *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

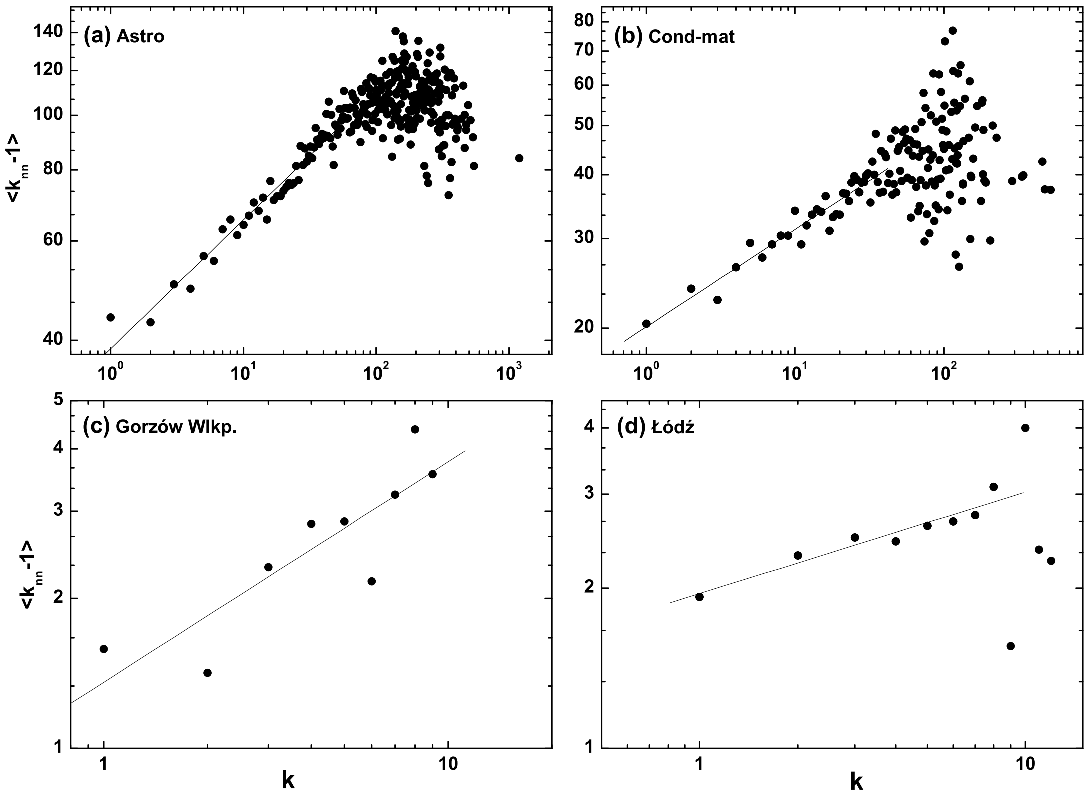

FIGURE 6. Estimation ϕ value for assortative networks. (a) *Astro*: the slope corresponds to exponent $\phi-1=0.23$, (b) *Cond-mat*: the slope corresponds to exponent $\phi-1=0.19$, (c) *Gorzów Wlkp.*: the slope corresponds to exponent $\phi-1=0.44$,(d) *Łódź*: the slope corresponds to exponent $\phi-1=0.19$.

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

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

The results (6) are presented in corresponding columns of the Table 2. One can observe a fairly good agreement between experimental data and (6) for co-authorship and biological networks as well as for the Internet Autonomous System and public transport network in Zielona Góra while for two other transport systems it leads to larger errors. Corrections due to clustering effects give a better fit for the coefficient A', while for some networks the coefficient B is closer to experimental value B_e than B'.

Now, let us consider the presence of degree correlations. Such correlations mean that average degrees $k_i^{(nn)}$ of nodes in the neighborhood of a node i depend on the degree k_i.

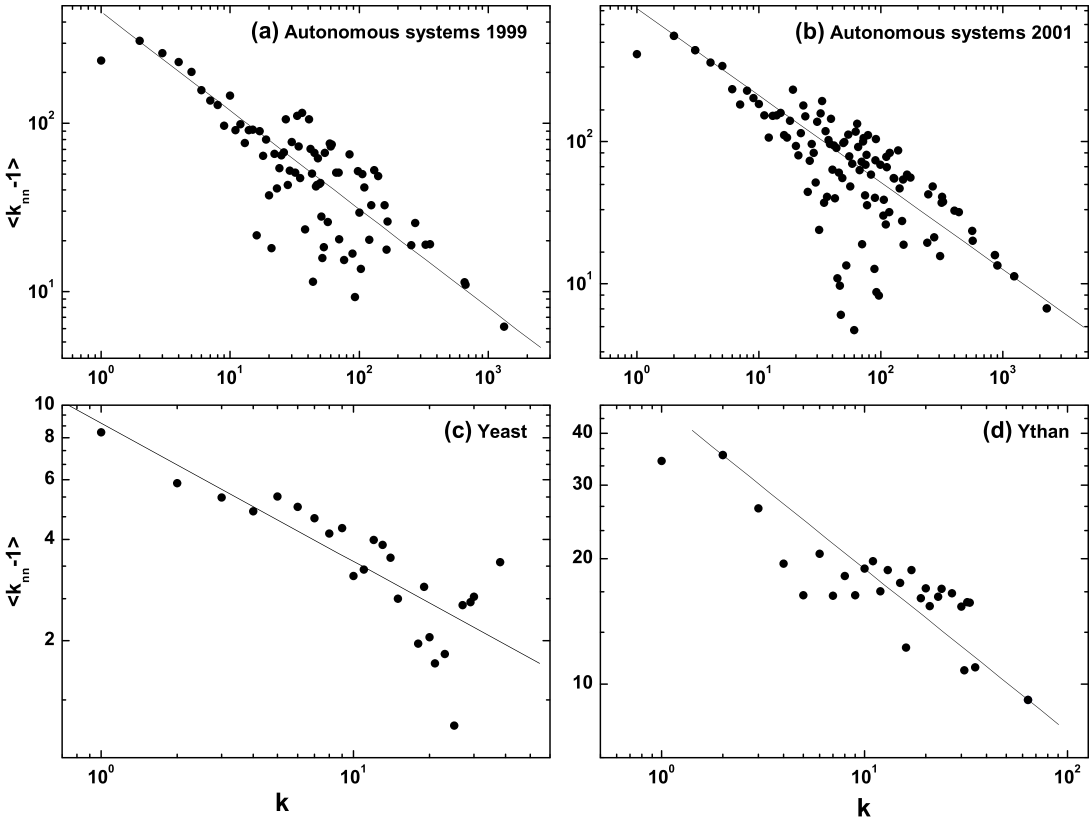

FIGURE 7. Estimation ϕ value for disassortative networks. (a) *Autonomous systems 1999*: the slope corresponds to exponent $\phi - 1 = -0.51$, (b) *Autonomous systems 2001*: the slope corresponds to exponent $\phi - 1 = -0.55$, (c) *Yeast*: the slope corresponds to exponent $\phi - 1 = -0.41$, (a) *Ythan*: the slope corresponds to exponent $\phi - 1 = -0.39$.

Let us assume that this relation can be written as

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

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 [27]. 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 [27]. 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

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

The plots 6 and 7 show the degree correlations for several different networks and illustrate the estimation of ϕ coefficient. After obtaining the histogram of $\langle k_{nn} - 1 \rangle$ for

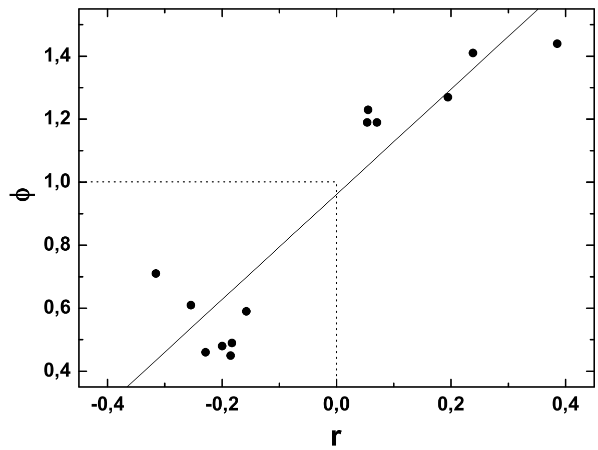

FIGURE 8. Dependency of ϕ on assortativity coefficient r.

each node i and plotting the dependence of those values on the node degree we perform the linear approximation in the log-log scale. A slope calculated in this way corresponds to the exponent $\phi - 1$ (with accordance to the formula $\langle k_{nn} - 1 \rangle \sim k^{\phi - 1}$). One can see, that the scaling (7) is not so obvious as for the relation (2). We have compared the resulting ϕ coefficient with standard assortativity coefficient introduced in [27] (see fig.8). We need to point out, that positive values of r correspond to ϕ larger than one. The linear fit of all points at the diagram (r, ϕ) nearly crosses the point $(r = 0, \phi = 1)$, what means that our definition of ϕ parameter is correct.

Results (9) are presented in corresponding columns of the Table 2. One can notice that the values of A_ϕ and B_ϕ are more accurate for the networks characterized by a ϕ coefficient above unity (assortative).

In conclusions we have observed 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 regardless of network degree distributions, correlations and clustering. We would like to stress, that the observed scaling holds over many decades even for strongly correlated networks with the correlation coefficients $|r| > 0.3$. We expect that the observed scaling law is universal for many complex networks, with applicability reaching far beyond the quoted examples. A simple model of random tree exploring the network explains such a scaling behavior and clustering effects have been taken into account to compare numerical data from model and data from real networks to theoretical predictions. We have also considered influence of first order degree-degree correlations on scaling parameters and we have found that inclusion of such correlations improves theoretical predictions for assortative networks, while it fails for disassortative ones. We suppose that to get better agreement between experimental data and theoretical results, higher order correlations should be taken into account.

ACKNOWLEDGMENTS

We are thankful to S.N. Dorogovtsev and Z. Burda for useful comments. Authors acknowledge the State Committee for Scientific Research in Poland for financial support under grant No. 1P03B04727. This work has also been partially supported by European Commission Projects: CREEN FP6-2003-NEST-Path-012864 and MMCOMNET FP6-2003-NEST-Path-012999. JAH is thankful to the Complex Systems Network of Excellence EXYSTENCE for the financial support that made possible his visit at MPI-PKS Dresden. A.F. acknowledges financial support from the Foundation for Polish Science (FNP 2005).

REFERENCES

1. R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74** 47 (2002).
2. S. Bornholdt and H.G. Schuster, *Handbook of graphs and networks*, Wiley-Vch (2002).
3. S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of networks*, Oxford Univ.Press (2003).
4. R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet : A Statistical Physics Approach*, Cambridge Univ. Press (2004).
5. A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
6. E. Ravasz and A.-L. Barabási, Phys. Rev. E **67** 026112 (2003).
7. Ch. Li and G. Chen, e-print cond-mat/0311333 (2003).
8. A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, Proc. Natl. Acad. Sci. (USA) **101**, 3747 (2004).
9. C.M. Ghim, E. Oh, K.-I. Goh, B. Khnag and D. Kim, Eur. Phys. J. B **38**, 193 (2004).
10. M. Barthélemy, Eur. Phys. J. B **38**, 163 (2004).
11. A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser and R. Guimerà, Eur. Phys. J. B **38**, 373 (2004).
12. A. Rinaldo, I. Rodriguez-Itrube, R. Rigon, E. Ijjasz-Vasquez and R. L. Bras, Phys. Rev. Lett. **70**, 822 (1993).
13. J. Banavar, A. Maritan and A. Rinaldo, Nature **399**, 130 (1999).
14. A. Fronczak, P. Fronczak and J.A. Hołyst, e-prints: cond-mat/0502663 (2005) and cond-mat/0212230 (2002).
15. A. Fronczak, P. Fronczak and J.A. Hołyst, Phys. Rev. E **70**, 056110 (2004).
16. J.A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak and K. Suchecki, to be published.
17. J.A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak and K. Suchecki, e-print cond-mat/0411160 (2004).
18. A.E. Motter, T. Nishikawa and Y.C. Lai, Phys. Rev. E **66** 065103(R) (2002).
19. Data for Foodwebs (*Silwood* and *Ythan*) [20] have been taken from the website http://www.cosin.org /extra/data/foodwebs/ whereas data for protein interaction network (*Yeast*) [21] has been taken from http://www.nd.edu/ networks/database/index.html.
20. D. Garlaschelli, G. Caldarelli and L. Pietronero, Nature **423**, 165 (2003).
21. H. Jeong, S. P. Mason, A.-L. Barabási and Z. N. Oltavi, Nature **411**, 41 (2001).
22. Scientific collaboration network [23] data have been collected by P. Wójcicki 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.
23. M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001).
24. Data for the Internet [4] has been taken from www.cosin.org/extra/data/internet/.
25. In those networks vertices are bus- and tramstops while an edge exists if at least one public transport line crosses two stops. [26].
26. J. Sienkiewicz and J.A. Hołyst, unpublished.
27. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
28. D.J.Watts and S.H.Strogatz, Nature **393** 440 (1998).