Available online at www.sciencedirect.com
SCIENCE DIRECT
Physica A 351 (2005) 167–174
PHYSICA A
www.elsevier.com/locate/physa

Scaling of distances in correlated complex networks

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

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

Received 7 November 2004
Available online 1 January 2005

Abstract

The influence of node–node degree correlations on distances in complex networks has been studied. We have found that even the presence of strong correlations in complex networks does not break a universal scaling of distances between vertices of such networks as science collaboration networks, biological networks, Internet Autonomous Systems and public transport systems. A mean distance between two nodes of degrees k_i and k_j in such networks equals to $\langle l_{ij}\rangle = A - B\log(k_i k_j)$ for a fixed value of the product $k_i k_j$. The scaling is valid over several decades. Parameters A and B depend on the mean value of a node degree $\langle k\rangle_{nn}$ calculated for the nearest neighbors. We have found that extending our simple theory basing on a random branching tree by the first-order node degree correlations improves theoretical predictions for parameters A and B in assortative networks, while it fails in disassortative ones.

PACS: 89.75.−k; 02.50.−r; 05.50. + q

Keywords: Complex networks; Distances; Scaling; Correlations

*Corresponding author. Tel.: + 48 22 6607133; fax: + 48 22 6282171.
E-mail address: jholyst@if.pw.edu.pl (J.A. Hołyst).

0378-4371/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2004.12.018

During the last few years random, evolving networks have become a very popular research domain among physicists [1–3]. A lot of efforts were put into the investigation of such systems in order to recognize their structure and to analyze emerging complex properties. It was observed that despite network diversity, most of real web-like systems share three prominent structural features: small average path length, high clustering and scale-free degree distribution [1–3]. The empirical analysis of many real complex networks has revealed the presence of several universal scaling laws. Scale-free behavior of degree distributions $P(k) \sim k^{-\gamma}$ observed in a number of social, biological and technological systems [4] is probably the most amazing. Another interesting scaling law is related to clustering effects [5]. It was observed that, depending on the behavior of the local clustering coefficient $c(k)$, real-world networks can be grouped into two universality classes of hierarchical $c(k) \sim k^{-\alpha}$ and non-hierarchical $c(k) \simeq const$ systems [6]. Recently, there has also been found that the distribution of connection strengths $s(k)$ in many real weighted networks follows power-law $s(k) \sim k^{\beta}$ [7,8]. It is not surprising, if we consider the fact, that connection loads (also called betweenness) scale similarly [9] $g(k) \sim k^{\eta}$ while load distributions is $P_L(\ell) \sim \ell^{-2}$ [10]. The cumulative community size distribution received for different social networks is also given by a universal scaling law $P(S>s) \sim s^{-\alpha}$ where the exponent $\alpha \approx 0.5$ or 1 in different scaling regions [11]. The first value is very near to a characteristic exponent of drainage area distribution in rivers [12] that was studied as a problem related to allometric relations [13]. At the macro-scale one can describe a whole network by a dependence of mean distance l between any pair of nodes in the system on the total number N of nodes. It was found that such distances behave in different ways for scale-free networks with different exponents of node degree distributions [14–17].

Recently, we have observed [18] universal scaling for distances l_{ij} between nodes possessing degrees k_i and k_j. The distances behave as

$$\langle l_{ij} \rangle = A - B \log(k_i k_j) , \tag{1}$$

where the mean value is taken over all pairs of nodes having a *fixed* product $k_i k_j$. We have found a simple model that explains such a scaling [18] that is in agreement with the results presented in paper [19]. The model bases on the concept of a random tree exploring the network. In the present paper we consider influence of node degree correlations on the scaling (1).

At Figs. 1–3 we show distances $\langle l_{ij} \rangle$ for several networks [18]. One can see that the scaling is very well fulfilled although the networks belong to very different classes (Erdős–Rényi random graphs, Barabási–Albert scale-free evolving networks, co-authorship networks, biological networks, Internet Autonomous Systems and public transport networks). This amazing scaling (1) can be connected with the small-world effect that exists in all these systems.

To justify relation (1) we introduced [18] a very simple approach that bases on a concept of branching trees exploring the space of a random network. The problem is to find the mean shortest path between a node i of degree k_i and a node j of degree k_j. If N is the total number of nodes in the system, $\langle k \rangle$ is a node mean degree and then the total number of all *directed links* (double number of all connections) is $2E =$

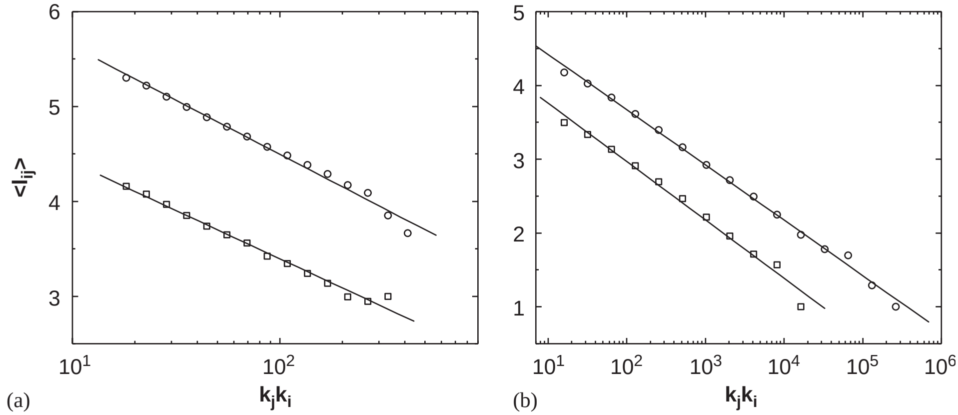

Fig. 1. (a) Erdős–Rényi random graphs: $\langle k \rangle = 8$ and $N = 1000$ (squares) $N = 10\,000$ (circles). Data are logarithmically binned with the power of 2. (b) Barabási–Albert networks: $\langle k \rangle = 8$ and $N = 1000$ (squares) $N = 10\,000$ (circles). Data are logarithmically binned with the power of 2.

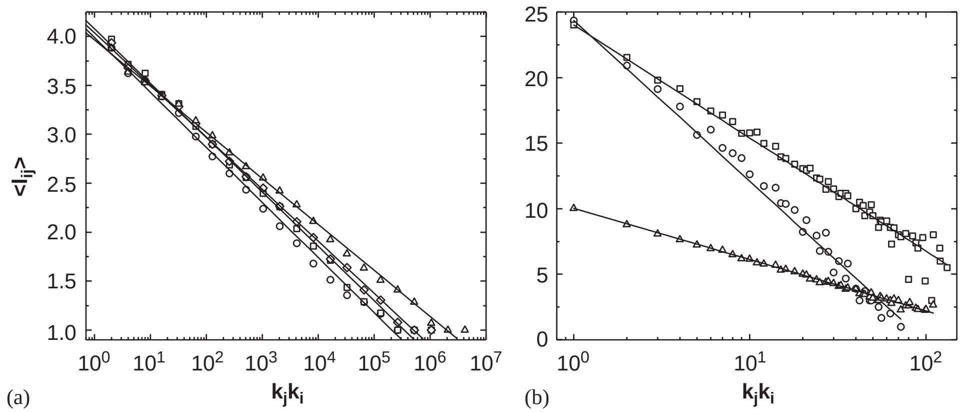

Fig. 2. (a) Autonomic systems: Year 1997—$N = 3015$, $\langle k \rangle = 3.42$ (circles), Year 1998—$N = 4180$, $\langle k \rangle = 3.72$ (squares) Year 1999—$N = 5861$, $\langle k \rangle = 3.86$ (diamonds), Year 2001—$N = 10\,515$, $\langle k \rangle = 4.08$ (triangles). Data are logarithmically binned with the power of 2. (b) Transport networks in Polish cities: Gorzów Wlkp.—$N = 269$, $\langle k \rangle = 2.48$ (circles), Łódź—$N = 1023$ (squares), $\langle k \rangle = 2.83$, Zielona Góra—$N = 312$, $\langle k \rangle = 2.98$ (triangles).

$N\langle k \rangle$. The probability that a randomly chosen directed link ends at the node j equals to

$$p_j = k_j/2E \,. \tag{2}$$

It follows that on average one needs

$$M_j = 1/p_j = 2E/k_j \tag{3}$$

random trials among all directed links to find any link that ends at the node j.

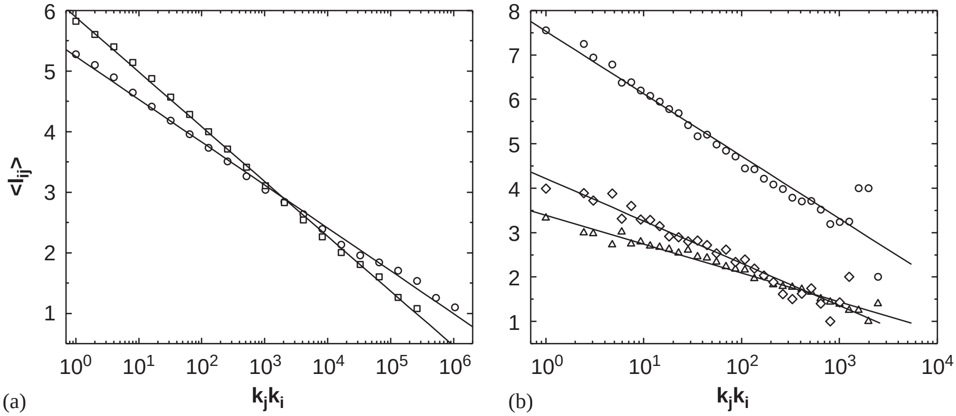

Fig. 3. (a) Co-authorship networks: Astro—$N = 13\,986$, $\langle k \rangle = 25.56$ (circles), Cond-mat $N = 17\,013$, $\langle k \rangle = 9.46$ (squares). Data are logarithmically binned with the power of 2. (b) Biological networks: Silwood—$N = 153$, $\langle k \rangle = 4.77$ (diamonds), Yeast—$N = 1846$, $\langle k \rangle = 2.39$ (circles), Ythan—$N = 135$, $\langle k \rangle = 8.83$ (triangles). Data are logarithmically binned with the power of 1.25.

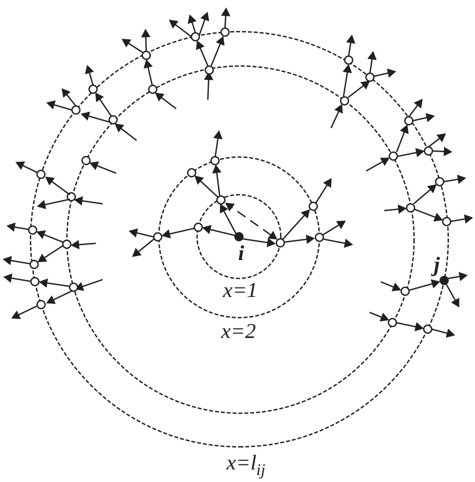

Fig. 4. Tree formed by a random process, starting from the node i and approaching the node j.

Now let us consider a branching process represented by a tree T_i (Fig. 4) that starts at the node i where an average branching factor is κ (all loops are neglected). The factor κ can be calculated as $\kappa = \langle k_{nn} - 1 \rangle$, i.e., as a mean value of nearest-neighbors degrees minus one, averaged geometrically over the whole network to take into account the product κ^{x-1} in (4).

If a distance between the node i and the surface of the tree equals to x, then on average there are

$$N_i = k_i \kappa^{x-1} \tag{4}$$

nodes at the surface of the tree and there is the same number of directed links ending at these nodes (Fig. 4). It follows that on average the tree T_i touches the node j when $N_i = M_j$, i.e., when

$$k_i k_j \kappa^{x-1} = N\langle k\rangle \; . \tag{5}$$

Since the mean distance from the node i to the node j is $\langle l_{ij}\rangle = x$ thus we get the scaling relation (1) with

$$A = 1 + \frac{\log(N\langle k\rangle)}{\log \kappa} \quad \text{and} \quad B = \frac{1}{\log \kappa} \; . \tag{6}$$

Eqs. (6) are in agreement with the paper [19] and slightly differ from our previous results [16,17].

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. 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 [20]. If ϕ is smaller than one, then the network is disassortative and high-degree nodes are mostly connected with low-degree nodes, which is typical for the Internet Autonomous Systems [20]. If we neglect higher-order correlations, then Eq. (5) 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 Eq. (6) with

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

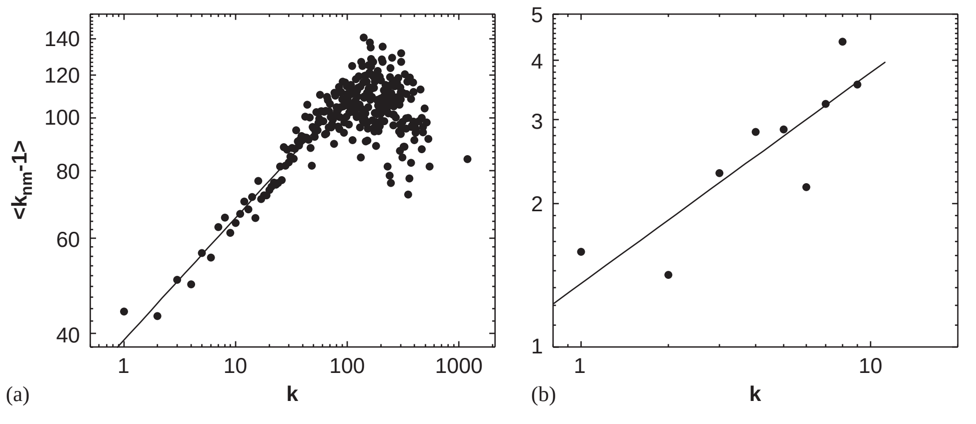

Fig. 5. (a) Estimation ϕ value for Astro network: the slope corresponds to exponent $\phi - 1 = 0.23$. (b) Estimation ϕ value for Gorzów Wlkp. network: the slope corresponds to exponent $\phi - 1 = 0.44$.

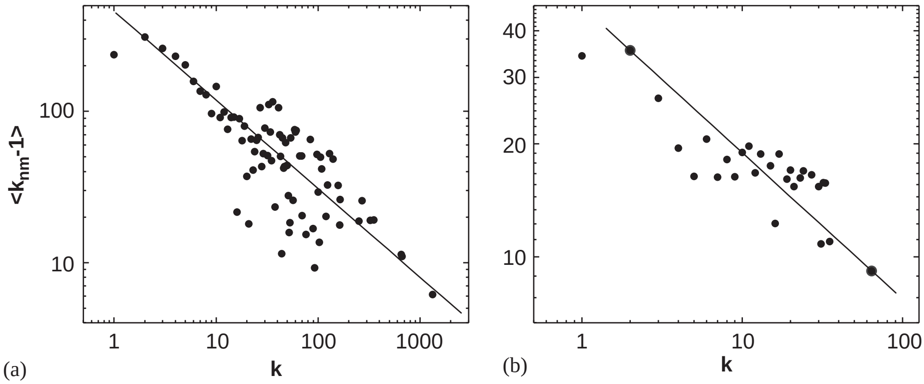

Fig. 6. (a) Estimation ϕ value for AS 1999 network: the slope corresponds to exponent $\phi - 1 = -0.49$. (b) Estimation ϕ value for Ythan network: the slope corresponds to exponent $\phi - 1 = -0.39$.

Figs. 5 and 6 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 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 scaling (7) is not so obvious as for relation (1). We have compared the resulting ϕ coefficient with standard assortativity coefficient introduced in Ref. [20] (see Fig. 7). One has 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 $(r = 0, \phi = 1)$, which means that our definition of ϕ parameter is correct.

Table 1 shows the comparison between experimental data collected from the examined networks and the results obtained from Eqs. (6) and (9). One can note that the values of A_ϕ and B_ϕ are more accurate for the networks characterized by a ϕ coefficient above unity (assortative).

We suppose that to get a better agreement between experimental data and theoretical results, higher-order correlations should be taken into account.

In conclusion, we have considered the influence of the first-order degree–degree correlations on the scaling law observed for distances between nodes with a fixed degree product $k_i k_j$. We have found that inclusion of such correlations improves theoretical predictions for assortative networks, while it fails for disassortative ones. We would like to stress that the observed scaling holds over many decades even for strongly correlated networks with the correlation coefficients $|r| > 0.3$. It seems that the observed dependence is a universal property of many complex networks.

We are thankful to S.N. Dorogovtsev for useful comments and to P. Wójcicki for help with data collection. The work has been supported by the special grant Dynamics of Complex Systems at Warsaw University of Technology and by the ESF programme STOCHDYN. JAH is thankful to the Complex Systems Network of Excellence *EXYSTENCE* for the financial support during the Thematic Institute at MPI-PKS Dresden.

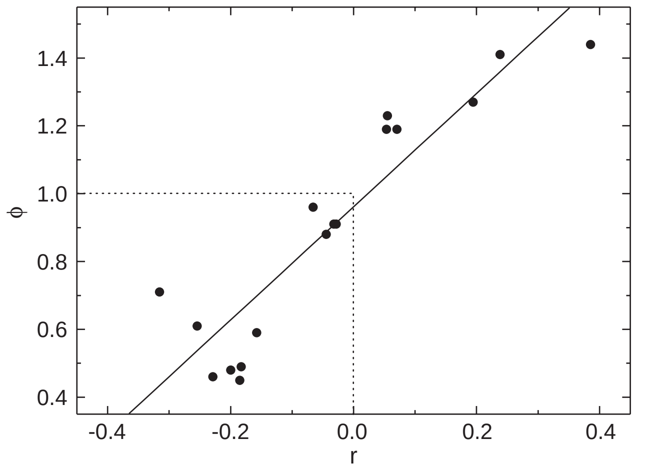

Fig. 7. Dependency of ϕ on assortativity coefficient r.

Table 1
Comparison between experimental and theoretical data

Network	r	A_e	A	B_e	B	ϕ	B_ϕ	A_ϕ
ER $N=10^3$	−0.066	5.43	5.46	1.017	1.143	0.96	1.097	5.36
ER $N=10^4$	−0.032	6.77	6.60	1.136	1.143	0.91	1.040	6.41
BA $N=10^3$	−0.044	4.54	4.24	0.813	0.830	0.88	0.731	4.03
BA $N=10^4$	−0.028	5.17	4.81	0.778	0.777	0.91	0.707	4.57
Astro	0.055	5.24	4.30	0.707	0.595	1.23	0.732	4.41
Cond-mat	0.053	5.90	5.09	0.908	0.786	1.19	0.935	5.05
Silwood	−0.316	4.22	3.69	0.955	0.941	0.71	0.668	3.19
Yeast	−0.158	7.53	6.66	1.406	1.552	0.59	0.916	5.71
Ythan	−0.254	3.39	3.35	0.649	0.765	0.61	0.466	2.81
AS 1997	−0.229	3.99	3.39	0.562	0.596	0.46	0.274	2.58
AS 1998	−0.200	4.08	3.41	0.555	0.575	0.48	0.276	2.65
AS 1999	−0.183	4.03	3.35	0.532	0.540	0.49	0.265	2.55
AS 2001	−0.185	3.96	3.23	0.471	0.481	0.45	0.217	2.50
Gorzów Wlkp.	0.385	24.36	16.06	12.27	5.333	1.44	7.679	16.67
Łódź	0.070	24.01	11.67	8.621	3.084	1.19	3.670	11.89
Zielona Góra	0.238	10.03	8.96	3.908	2.682	1.41	3.781	9.62

Astro and Cond-mat are co-authorship networks, Gorzów Wlkp., Łódź and Zielona Góra are public transport networks in Poland; Silwood, Yeast and Ythan are biological networks and AS stands for autonomic systems with number meaning the year data were gathered. ER stands for Erdős–Rényi graphs and BA for Barabási–Albert networks. r is the assortativity coefficient, A_e and B_e mean experimental values of those coefficients whereas $B = 1/\log\kappa$ and $A = 1 + B\log N\langle k\rangle$. B_ϕ follows $B_\phi = B\phi$ while $A_\phi = A + 2 - 2B\log D$ where ϕ is the scaling exponent for degree correlations.

References

[1] R. Albert, A.L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[2] S. Bornholdt, H.G. Schuster, Handbook of Graphs and Networks, Wiley-VCH, 2002.
[3] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of Networks, Oxford University Press, Oxford, 2003.
[4] A.L. Barabási, R. Albert, Science 286 (1999) 47.
[5] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[6] E. Ravasz, A.L. Barabási, Phys. Rev. E 67 (2003) 026112.
[7] Ch. Li, G. Chen, e-print cond-mat/0311333, 2003.
[8] A. Barrat, et al., Proc. Natl. Acad. Sci. USA 101 (2004) 3747.
[9] M. Barthélemy, Eur. Phys. J. B 38 (2004) 163.
[10] C.M. Ghim, et al., Eur. Phys. J. B 38 (2004) 193.
[11] A. Arenas, et al., Eur. Phys. J. B 38 (2004) 373.
[12] A. Rinaldo, et al., Phys. Rev. Lett. 70 (1993) 822.
[13] J. Banavar, et al., Nature 399 (1999) 130.
[14] R. Cohen, S. Havlin, Phys. Rev. Lett. 90 (2003) 058701.
[15] S.N. Dorogovtsev, et al., Nucl. Phys. B 653 (2003) 307.
[16] A. Fronczak, et al., e-print cond-mat/0212230, 2002.
[17] A. Fronczak, et al., Phys. Rev. E 70 (2004) 056110.
[18] J.A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, K. Suchecki, cond-mat/0411160, 2004.
[19] A.E. Motter, et al., Phys. Rev. E 66 (2002) 065103R.
[20] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701.