

Universal scaling of distances in complex networks

Janusz A. Hołyst, Julian Sienkiewicz, Agata Fronczak, Piotr Fronczak and Krzysztof Suchecki
Faculty of Physics and Center of Excellence for Complex Systems Research,
Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland
(Dated: November 5, 2004)

Universal scaling of distances between vertices of Erdős-Rényi random graphs, scale-free Barabási-Albert models, science collaboration networks, biological networks, Internet Autonomous Systems and public transport networks are observed. A 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 scaling is valid over several decades. A simple theory for the appearance of this scaling is presented. Parameters A and B depend on the mean value of a node degree $\langle k \rangle_{nn}$ calculated for the nearest neighbors and on network clustering coefficients.

PACS numbers: 89.75.-k, 02.50.-r, 05.50.+q

Recently, a lot of efforts have been put into investigation of network systems, in order to recognize their structures and emerging complex properties (for a review see [1, 2, 3, 4]). The empirical analysis of many real complex networks has revealed the presence of several universal scaling laws. The scale-free behavior of degree distributions $P(k) \sim k^{-\gamma}$ [5] observed in a number of social, biological and technological systems is probably the most amazing. Aside from that, many further scaling laws have been found, such as a dependence of clustering coefficient on node degree in hierarchical networks $c(k) \sim k^{-\alpha}$ [6], scaling of connection weight distribution [7, 8], connection load distribution [9], load dependence on degree [10] and others [11, 12, 13].

At the macro-scale one can describe a whole network by a dependence of a mean distance between any pair of nodes on the network size and in many real networks the small-world effect is observed [14], i.e. the mean distance l between nodes of such networks increases not faster than logarithmically with their size N . In scale-free networks the small-world effect changes to ultra small-world effect ($l \sim \log \log N$) when $\lambda < 3$ [15, 16, 17, 18]. It was also observed that if a network disorder is present, optimal paths become much longer and the small-world effect disappears [19]. The recent research on complex networks is slowly shifting from problems of network topology to directed and weighted networks [20, 21, 22], network dynamics [23], as well as to the issue of network efficiency [24].

In the present paper we come back to networks geometry and analyze surprising empirical scaling that has not been considered before. We think that our observations can be important for understanding of network structures and for processes driving their evolution as well as for constructing search algorithms in real web-like systems. We show that the mean distance between nodes with degrees k_i and k_j is given by the following relation

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

The above scaling law is shown to be correct not only for network models but also for many real networks regard-

less of their degree distribution and correlation profiles.

Fig. 1 presents mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$ in selected complex networks. Analyzed systems belong to very different classes ranging from generic models of random graphs and scale-free networks, through natural systems such as food webs and metabolic networks to man-made like the Internet and public transport networks. We include data for Erdős-Rényi random graphs, Barabási-Albert evolving networks, biological networks [25, 26, 27] (*Silwood*, *Ythan*, *Yeast*), social networks [28, 29] (co-authorship groups *Astro* and *Cond-mat*), Internet Autonomous Systems [30] and selected networks for public transport in Polish cities [31, 32] (Gorzów Wlkp., Łódź, Zielona Góra). One can see, that the relation (1) 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 [34], 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 [14], the scaling (1) is nearly absent and it is visible only for large rewiring probability, and only for larger degrees, where nodes have many shortcuts.

Although the scaling (1) 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. 2 presents the dependence of average path length l_{ij} on k_i , for a fixed product $k_i k_j$ in the case of *Astro* network and for the Internet Autonomous Systems in 1999. One can see that although the *Astro* network is assortative [34] (short-range attraction), pairs of nodes with similar degrees are in average further away than different degree pairs (long-range repulsion). For the disassortative network AS [34] 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 [33].

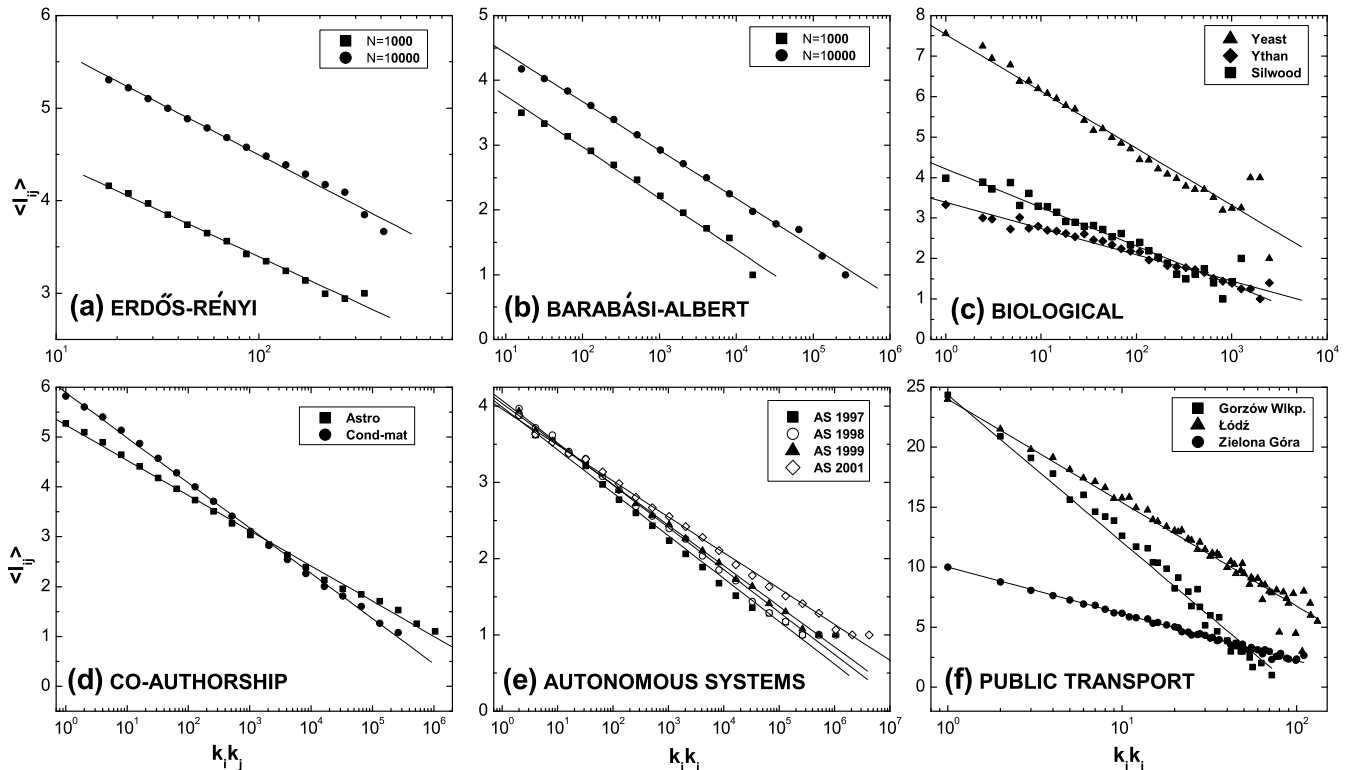


FIG. 1: Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. (a) *Erdős-Rényi* random graphs: $\langle k \rangle = 8$ and $N = 1000$ (squares) $N = 10000$ (circles), (b) *Barabási-Albert* networks: $\langle k \rangle = 8$ and $N = 1000$ (squares) $N = 10000$ (circles), (c) Biological networks: *Silwood* (squares), *Yeast* (triangles), *Ythan* (diamonds), (d) Co-authorship networks: *Astro* (squares), *Cond-mat* (circles), (e) Internet Autonomous Systems: *Year 1997* (squares), *Year 1998* (circles) *Year 1999* (triangles), *Year 2001* (diamonds), (f) Public transport networks in Polish cities: *Gorzów Wlkp.* (squares), *Łódź* (triangles), *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.

To justify the relation (1) let us consider a simple approach that bases on a concept of branching trees exploring the space of a random network. We need to estimate the mean shortest path between a node i of degree k_i and a node j of degree k_j . Let us notice that following a random direction of a randomly chosen edge one approaches node j with a probability $p_j = k_j / (2E)$, where $2E = N\langle k \rangle$ is a double number of links. It follows 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 (Fig. 3) 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

$$k_i k_j \kappa^{x-1} = N\langle k \rangle. \quad (2)$$

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

$$A = 1 + \frac{\log(N\langle k \rangle)}{\log \kappa} \quad \text{and} \quad B = \frac{1}{\log \kappa}. \quad (3)$$

The result (3) is in agreement with the paper [35] 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 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 (2). The last approximation can be improved if one finds a probability distribution for $P(l_{ij})$ and a corresponding average value $\langle l_{ij} \rangle$. Such an approach has been performed in our papers [17, 18] where we have applied the concept of hidden variables and have received the same value of the parameter B and small corrections to A .

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 it could be estimated as the

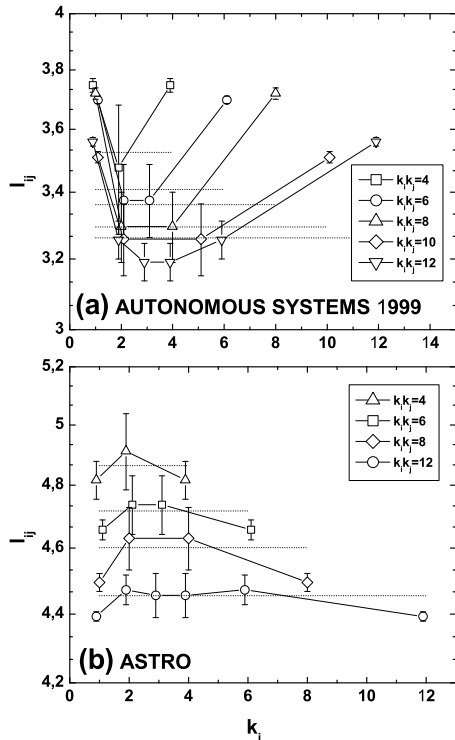


FIG. 2: Dependence of average path length on k_i , for fixed $k_i k_j$ product. The lines connecting the symbols are there for clarity. The bars show point weight, meaning relative numbers of pairs ij . The horizontal lines are weighted averages over k_i , and are just average path lengths for given $k_i k_j$. The top shows data for the Internet Autonomous Systems, while bottom for *Astro* co-authorship network. 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.

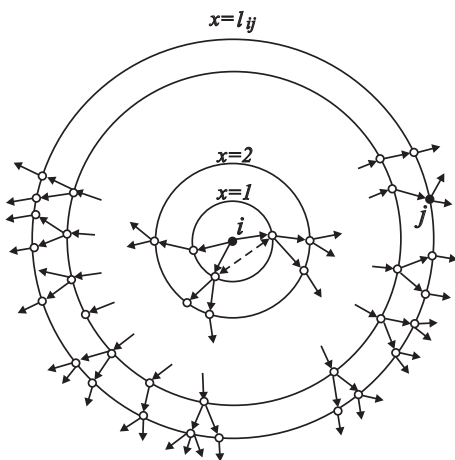


FIG. 3: Tree formed by a random process, starting from the node i and approaching the node j .

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 (2). 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. We calculate arithmetic mean branching factor over nearest neighborhood of every node m , i.e. $\kappa^{(m)} = \langle k \rangle_{nn}^{(m)} - 1$, and then average it geometrically over all nodes m , i.e. $\kappa = \langle \kappa^{(m)} \rangle_m$. Although our approach is not exact, it leads to a good agreement between coefficients A_e , B_e taken from real networks (see Table I) and A , B calculated from our model.

The influence of loops of the length three on the relation (1) 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.3). In fact such a situation 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 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

$$k_i k_j [\kappa(1 - c')]^2 [\kappa(1 - c)]^{x-3} = N \langle k \rangle, \quad (4)$$

where $c' = c(\langle k \rangle - 1)/\kappa$. It follows that instead of (3) we have

$$A' = 3 + \frac{\log(N \langle k \rangle) - 2 \log[\kappa(1 - c')]}{\log[\kappa(1 - c)]}, \quad B' = \frac{1}{\log[\kappa(1 - c)]}. \quad (5)$$

The Table I contains a comparison between the experimental data (Fig. 1) and theoretical predictions given by Eq. (3) and (5). One can observe that the range of parameters A and B for different networks is very broad. Our approximate approach fits very well to random Erdős-Rényi graphs and BA models and is fairly good 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' . The good agreement between theory based on random networks and empirical data suggests that the considered real networks exhibit a large level of randomness.

In conclusions we have observed universal path length scaling for different classes of real networks and models.

network	N	$\langle k \rangle$	c	A_e	A	A'	B_e	B	B'
Erdős-Rényi	1000	8.00	0.007	5.43	5.46	5.48	1.017	1.143	1.147
Erdős-Rényi	10000	8.00	0.001	6.77	6.60	6.61	1.136	1.143	1.143
Barabási-Albert	1000	8.00	0.038	4.54	4.24	4.27	0.813	0.830	0.842
Barabási-Albert	10000	8.00	0.007	5.17	4.81	4.81	0.778	0.777	0.779
Astro	13986	25.56	0.609	5.24	4.30	4.98	0.707	0.595	0.786
Cond-mat	17013	9.46	0.604	5.90	5.09	6.38	0.908	0.786	1.150
Silwood	153	4.77	0.142	4.22	3.69	3.78	0.955	0.941	1.004
Yeast	1846	2.39	0.068	7.53	6.66	6.87	1.406	1.552	1.629
Ythan	135	8.83	0.216	3.39	3.35	3.45	0.649	0.765	0.832
AS 1997	3015	3.42	0.182	3.99	3.39	3.42	0.562	0.596	0.629
AS 1998	4180	3.72	0.250	4.08	3.41	3.45	0.555	0.575	0.620
AS 1999	5861	3.86	0.250	4.03	3.35	3.38	0.532	0.540	0.579
AS 2001	10515	4.08	0.289	3.96	3.23	3.25	0.471	0.481	0.518
Gorzów Wlkp.	269	2.48	0.082	24.36	16.06	19.76	12.270	5.333	6.651
Lódź	1023	2.83	0.065	24.01	11.67	12.70	8.621	3.084	3.389
Zielona Góra	312	2.98	0.067	10.03	8.96	9.63	3.908	2.682	2.917

TABLE I: Comparison between experimental and theoretical data. *Astro* and *Cond-mat* are co-authorship networks, *Silwood*, *Yeast* and *Ythan* are biological networks and *AS* stands for the Internet Autonomous Systems with number meaning the year data were gathered, *Gorzów Wlkp.*, *Lódź* and *Zielona Góra* are public transport networks in corresponding Polish cities. N is the number of nodes, $\langle k \rangle$ - mean degree value, c - clustering coefficient. A_e and B_e mean experimental values (Fig. 1) whereas A and B are given by (3), A' and B' by (5).

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. A simple model of random tree exploring the network explains such a behavior and leads to a good agreement with experimental data. We expect that the observed scaling law is universal for many complex networks, with applicability reaching far beyond the quoted examples.

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 KBN grant No 1P03B04727 and by the COST Action P10 *Physics of Risk*. JAH is thankful to the Complex Systems Network of Excellence *EXYSTENCE* for the financial support during the Thematic Insitute at MPI-PKS Dresden.

[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 et al., *Proc. Natl. Acad. Sci. (USA)* **101**, 3747 (2004).

[9] C.M. Ghim et al., *Eur. Phys. J. B* **38**, 193 (2004).
[10] M. Barthélemy, *Eur. Phys. J. B* **38**, 163 (2004).
[11] A. Arenas et al., *Eur. Phys. J. B* **38**, 373 (2004).
[12] A. Rinaldo et al., *Phys. Rev. Lett.* **70**, 822 (1993).
[13] J. Banavar et al., *Nature* **399**, 130 (1999).
[14] D.J.Watts and S.H.Strogatz, *Nature* **393** 440 (1998).
[15] R. Cohen and S. Havlin, *Phys. Rev. Lett.* **90** 058701 (2003).
[16] S.N.Dorogovtsev et al., *Nucl. Phys. B* **653**, 307 (2003).
[17] A. Fronczak et al., e-print cond-mat/0212230 (2002).
[18] A. Fronczak et al., *Phys. Rev. E* **70**, XXXXXX (2004).
[19] L.A. Braunstein et al., *Phys. Rev. Lett.* **91**, 168701 (2003).
[20] A. Barrat et al., *Phys. Rev. Lett.* **92**, 228701 (2004).
[21] T. Antal and P.L. Krapivsky, e-print cond-mat/0408285 (2004).
[22] S.N. Dorogovtsev and J.F.F. Mendes, e-print cond-mat/0408343 (2004).
[23] M. A. de Menezes and A.-L. Barabási, *Phys. Rev. Lett.* **92**, 028701 (2004).
[24] M.T.Gastner and M.E.J.Newman, cond-mat/0409702 (2004).
[25] Data for Foodwebs (*Silwood* and *Ythan*) [26] have been taken from www.cosin.org/extra/data/foodwebs/ whereas data for protein interaction network (*Yeast*) [27] has been taken from www.nd.edu/networks/database/index.html.
[26] D. Garlaschelli et al., *Nature* **423**, 165 (2003).
[27] H. Jeong et al., *Nature* **411**, 41 (2001).
[28] Scientific collaboration network [29] 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.
[29] M.E.J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
[30] Data for the Internet [4] has been taken from www.cosin.org/extra/data/internet/.

- [31] In those networks vertices are bus- and tramstops while an edge exists if at least one public transport line crosses two stops. [32].
- [32] J. Sienkiewicz and J.A. Hołyst, unpublished.
- [33] J.A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak and K. Suchecki, to be published.
- [34] M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
- [35] A.E. Motter, T. Nishikawa, Y.C. Lai, Phys. Rev. E **66** 065103(R) (2002).