Path Length Scaling and Discrete Effects in Complex Networks

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

Faculty of Physics and Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland
julas|agatka|jholyst|suchecki@if.pw.edu.pl

1 Introduction

Complexity has thousands of faces. This mere fact is commonly known to wide group of scientists working in such different areas as physics, chemistry, meteorology or logistics. It is rather impossible to see at first glance if a problem we are tackling with can be counted as a complex one. In the 18th century, when the problem of seven bridges in Königsberg was solved by Leonhard Euler [1], people gained an extremely useful and simple tool for difficult problems - the *graph theory* - which has now been dynamically developing since 1950's works of Erdös [2]. Late 1990's brought us another "revolution" in understanding of sophisticated problems - *complex networks*. The discoveries of Watts and Strogatz [3] as well as those of Barabási and Albert [4] set up a new direction in complexity. The networks have shown us directly how a structure which initially seems to be complicated, might be modelled by simple rules if we map the ingredients of this system onto nodes and links in a complex network. The power of the complex networks lays in the fact that such distant relations as acquaintances between people [5], collaboration of scientists [3], protein interactions [6], Internet [7], stock assets [8] or city public transport [9] may all be treated as one topological object.

In this work we would like to show two different studies: the first devoted to explanation of certain scaling laws observed in real-world systems and the second one that was set up to give exact analytical expression for average path length in some network models. Those two studies combined together enabled us to discover and to explain effects of oscillations on average path length - a phenomenon emerging from the discrete structure of complex networks . At the end we present a direct application of the oscillation effect to a simple optimization problem [10] that may be used in real-life situations.

2 Basic Network Characteristics

An object called *network* consists of *nodes* (vertices) and *edges* (links) that connect nodes. The most natural definition is that of *degree* - the number of edges connected to a vertex. To describes tatistical properties of network one uses *degree distribution* $p(k)$ - the probability that a randomly chosen vertex is characterized by degree k. As any other science, also network science has its own fundamental observables which can be used to distinguish between different types of networks [11].

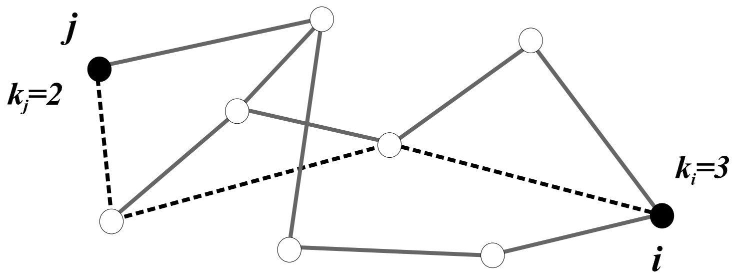

Fig. 1. Internode distance $l_{ij} = 3$ between nodes i and j - marked as a dashed line.

The *internode distance* l_{ij} between nodes i and j is depicted at Fig. 1 and can be defined as the shortest number of edges one has to pass getting from one node to another. The sum of all internode distances in the network divided by the total number of pairs is called *average path length*.

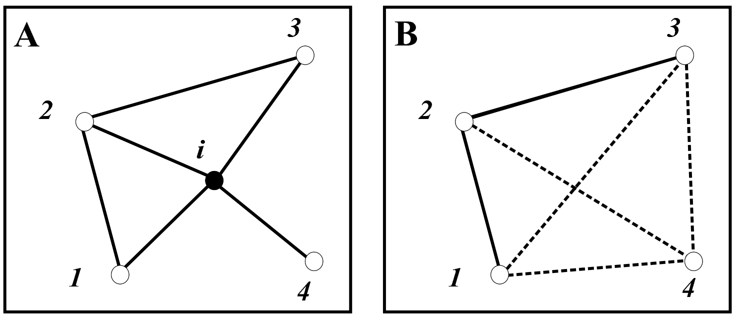

Fig. 2. Construction used to calculate clustering coefficient for node i: (A) node i with its nearest neighbors (B) connections among nearest neighbors of node i - existing are solid lines, possible ones are dashed lines. One can calculate that in this example $c_i = 2/6 = 1/3$.

The *Clustering coefficient* is a parameter that describes the local transitivity of the network. For a single node such coefficient is defined as the ratio of the number of connections e_i among the first neighbors of node i to the maximal possible number of such links. It can be written as

$$c_i = \frac{2e_i}{k_i\,(k_i - 1)}. \tag{1}$$

An average value taken over all nodes in the network gives the global clustering coefficient .

To investigate correlations between nodes' degrees one uses *assortativity coefficient* [12] that takes a following form:

$$r = \frac{\sum_i j_i k_i - \frac{1}{M}\sum_i j_i \sum_i k_i}{\sqrt{\sum_i j_i^2 - \frac{1}{M}(\sum_i j_i)^2}\sqrt{\sum_i k_i^2 - \frac{1}{M}(\sum_i k_i)^2}}, \tag{2}$$

where M - number of pairs of nodes (twice the number of edges), j_i, k_i - degrees of vertices at both ends of i-th pair and index i goes over all pairs of nodes in the network. A positive value of r means that nodes with high degree tend to link among each other (see Fig. 3A) and similarly low degree nodes are connected to other low degree nodes, while negative values - that high degree nodes are mostly connected to low degree ones (see Fig. 3B).

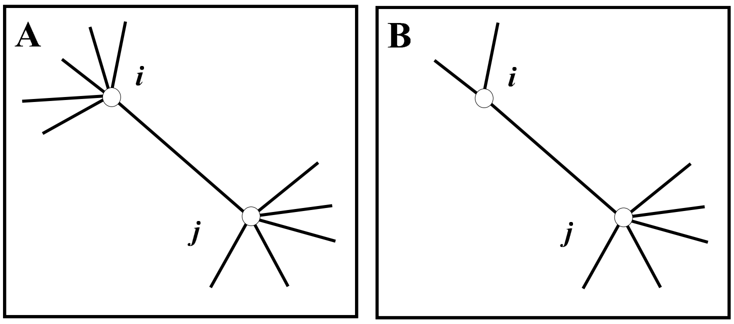

Fig. 3. Visualization of typical connections in (A) assortative (B) disassortative network.

3 Internode Distances

In 2005 our group observed [13] a universal scaling for distances $\langle l_{ij} \rangle$ between nodes possessing degrees k_i and k_j. The distances behave as

$$\langle l_{ij} \rangle = a - b \log(k_i k_j), \tag{3}$$

where the mean value is taken over all pairs of nodes having a *fixed* product $k_i k_j$. Figure 4 presents mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j versus the product of their degrees $k_i k_j$ in various complex networks belonging to very different types: Erdős-Rényi random graphs, Barabási-Albert evolving networks, biological networks [14, 15, 16] (*Silwood*, *Ythan*, *Yeast*), social networks [17, 18] (co-authorship groups *Astro* and *Cond-mat*), Internet Autonomous Systems [19] as well as selected networks for public transport in Polish cities [20, 9](Gorzów Wlkp., Łódź, Zielona Góra). The relation (3) is very well fulfilled over several decades for all our data. We would like to underline that the networks mentioned above display a wide variety of basic characteristics such as degree probability, clustering coefficient or degree-degree correlations and their origins are quite different in each single case. The only apparent common feature of all above systems is the small-world effect.

3.1 Model

To justify the relation (3) we use a simple approach basing on the concept of branching trees exploring the space of a random network. The goal is to find the mean shortest path between a node i of degree k_i and a node j of degree k_j. One can immediately notice that following a random direction of a randomly chosen edge one arrives at node j with a probability $p_j = k_j/(2E)$, where $2E = N\langle k \rangle$ is a double number of links. In consequence one needs (in average) $M_j = 1/p_j = 2E/k_j$ of random trials to arrive at the node j.

Now let us consider a branching process represented by the tree T_i (Fig. 5) 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

$$k_i k_j \kappa^{x-1} = N\langle k \rangle. \tag{4}$$

Since the mean distance from the node i to the node j is $\langle l_{ij} \rangle = x$ thus we get the scaling relation (3) with

$$a = 1 + \frac{\log(N\langle k \rangle)}{\log \kappa} \quad \text{and} \quad b = \frac{1}{\log \kappa}. \tag{5}$$

The result (5) is in agreement with the paper [21] where the concept of generating functions for random graphs has been used. One should be aware of the fact that in the above presented estimations we have omitted degree-degree correlations or loops, and we have treated the branching level x as a continuum variable to fulfill the relation (4).

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

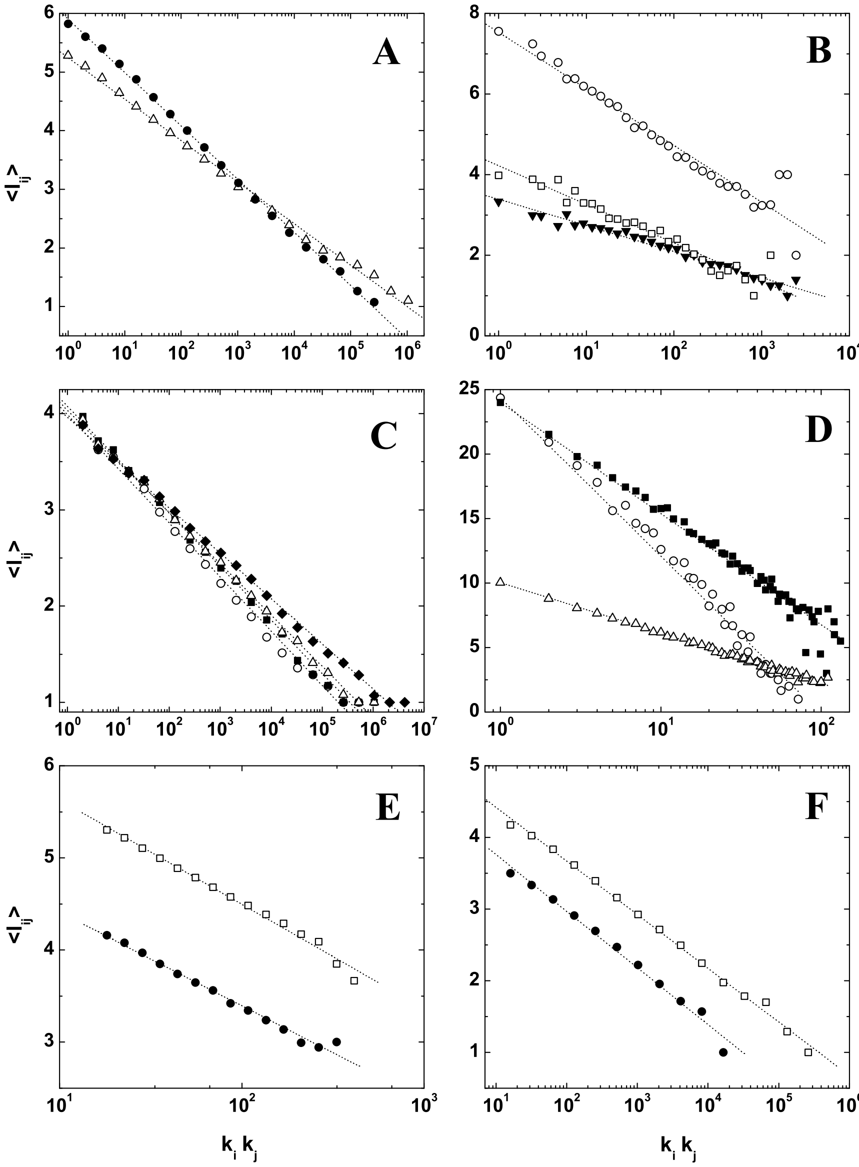

Fig. 4. Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$. (A) Co-authorship networks: *Astro* (triangles), *Cond-mat* (circles). (B) Biological networks: *Silwood* (squares), *Yeast* (circles), *Ythan* (triangles). (C) Internet Autonomous Systems, years: *1997* (circles), *1998* (squares), *1999* (triangles), *2001* (diamonds). (D) Public transport networks in Polish cities: *Gorzów Wlkp.* (circles), *Łódź* (squares), *Zielona Góra* (triangles). (E) *Erdős-Rényi* random graphs: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares), (F) *Barabási-Albert* networks: $\langle k \rangle = 8$ and $N = 1000$ (circles) $N = 10000$ (squares). In (A), (C), (E) and (F) data are logarithmically binned with the power of 2, in case of (B) with the power of 1.25 and in case of (D) data are not binned.

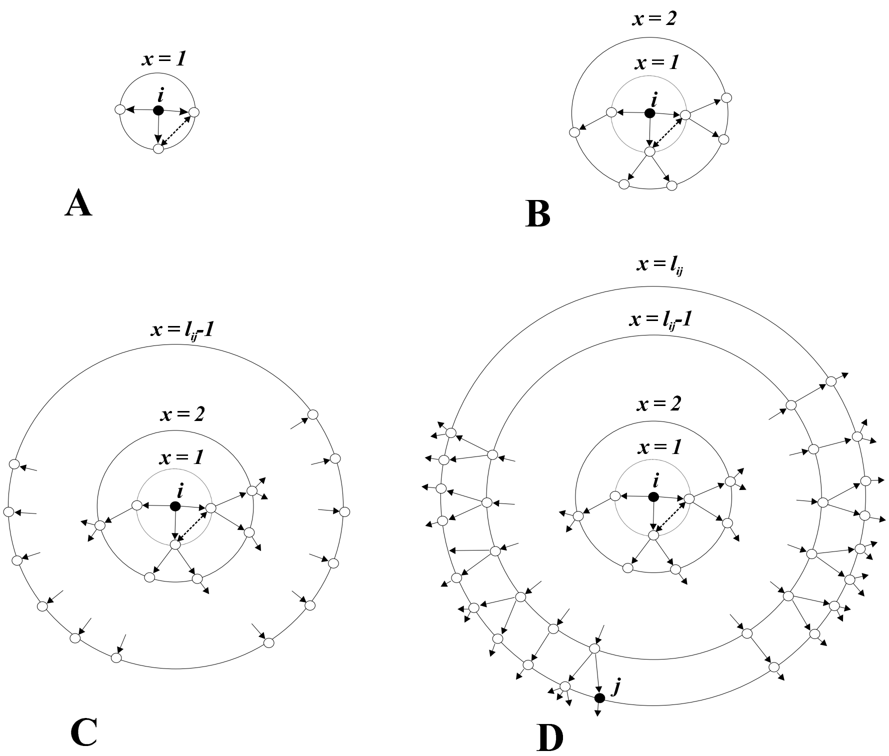

Fig. 5. Tree formed by a random process, starting from the node i and approaching the node j. Parts A, B, C and D depict consecutive steps of the random tree formation process.

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 (4). 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 1) and a, b calculated from our model.

The good agreement between theory based on random networks and empirical data suggests that the considered real networks exhibit a large level of randomness.

3.2 Clustering Coefficient

The influence of loops of the length three on the relation (3) 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. 5). 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

$$k_i k_j [\kappa(1-c')]^2 [\kappa(1-c)]^{x-3} = N\langle k \rangle, \tag{6}$$

where $c' = c(\langle k \rangle - 1)/\kappa$. It follows that instead of (5) we have

$$a' = 3 + \frac{\log(N\langle k \rangle) - 2\log[\kappa(1-c')]}{\log[\kappa(1-c)]} \text{ and } b' = \frac{1}{\log[\kappa(1-c)]}. \tag{7}$$

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' (see Table 1).

3.3 Degree-degree Correlations

One can see from Table 1 some of the examined systems share the property of high degree-degree correlations, either negative or positive. We would like to take that fact into account in our considerations [22]. Degree-degree 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

$$\kappa_i \equiv k_i^{(nn)} - 1 = D k_i^{\phi - 1}. \tag{8}$$

The value of ϕ being over 1 means the network is assortative, while $\phi < 1$ is a sign of disassortative nature of the system. If we neglect higher order correlations then Eq. (4) should be replaced by

$$k_i k_j \kappa_i \kappa_j \kappa^{x-3} = N\langle k \rangle. \tag{9}$$

Taking into account Eq. (8) we can replace parameters a and b given by the Eq. (5) with

$$a_\phi = a + 2 - 2b \log D \quad \text{and} \quad b_\phi = \phi b. \tag{10}$$

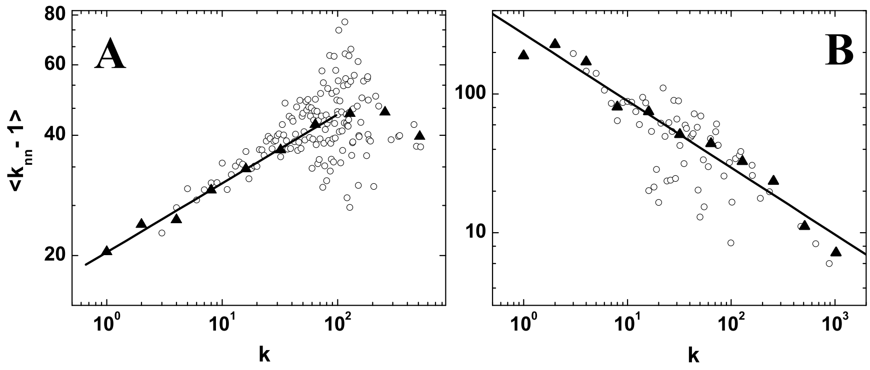

Fig. 6. Estimation of ϕ coefficient for (A) *Cond-mat* co-authorship network, (B) *Internet Autonomous System 1998*.

Figure 6 illustrates the estimation of ϕ coefficient for two different networks: assortative *Cond-mat* collaboration network (Fig. 6A) and disassortative *Internet Autonomous System* network taken in year 1998 (Fig. 6B). 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 the scaling (8) is not so obvious as for the relation (3).

Table 1 shows the comparison between experimental data collected from the examined networks and the results obtained from Eqs (5) and (10). 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).

3.4 Rescaling

We are also able to present all the data points at one plot only - this can be done using simple re-scaling methods which as a reference points take *effective degree value* k_{eff} and *maximal intervertex distance* l_{max}. Defining rescaled values $\widetilde{\langle l_{ij} \rangle}$ and $\widetilde{k_i k_j}$ as:

$$\widetilde{\langle l_{ij} \rangle} = 2 \frac{\langle l_{ij} \rangle}{l_{max}} \log k_{eff} \qquad (11)$$

$$\widetilde{k_i k_j} = \frac{k_i k_j}{k_{eff}^2} \qquad (12)$$

we get (see Appendix A) a simplified formula for the scaling relation:

$$\widetilde{\langle l_{ij} \rangle} = - \log \widetilde{k_i k_j}, \qquad (13)$$

Table 1. 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.*, *Łódź* and *Zielona Góra* are public transport networks in corresponding Polish cities. c is the clustering coefficient, r - assortativity value, and ϕ coefficient is obtained using procedure described in the text. a_e and b_e mean experimental values (Fig. 4), a and b are given by (5), a' and b' by (7) while a_ϕ and b_ϕ by (10). In case of models due to the lack of correlations ϕ parameters should be equal to 1, so we omitted values of a_ϕ and b_ϕ.

network	c	r	ϕ	a_e	a	a'	a_ϕ	b_e	b	b'	b_ϕ
ER $N = 10^3$	0.007	0	-	5.43	5.46	5.48	-	1.017	1.143	1.147	-
ER $N = 10^4$	0.001	0	-	6.77	6.60	6.61	-	1.136	1.143	1.143	-
BA $N = 10^3$	0.038	0	-	4.54	4.24	4.27	-	0.813	0.830	0.842	-
BA $N = 10^4$	0.007	0	-	5.17	4.81	4.81	-	0.778	0.777	0.779	-
Astro	0.609	0.055	1.23	5.24	4.30	4.98	4.41	0.707	0.595	0.786	0.732
Cond-mat	0.604	0.053	1.19	5.90	5.09	6.38	5.05	0.908	0.786	1.150	0.935
Silwood	0.142	-0.316	0.71	4.22	3.69	3.78	3.19	0.955	0.941	1.004	0.668
Yeast	0.068	-0.158	0.59	7.53	6.66	6.87	5.71	1.406	1.552	1.629	0.916
Ythan	0.216	-0.254	0.61	3.39	3.35	3.45	2.81	0.649	0.765	0.832	0.466
AS 1997	0.182	-0.229	0.46	3.99	3.39	3.42	2.58	0.562	0.596	0.629	0.274
AS 1998	0.250	-0.200	0.48	4.08	3.41	3.45	2.65	0.555	0.575	0.620	0.276
AS 1999	0.250	-0.183	0.49	4.03	3.35	3.38	2.55	0.532	0.540	0.579	0.265
AS 2001	0.289	-0.185	0.45	3.96	3.23	3.25	2.50	0.471	0.481	0.518	0.217
Gorzów Wlkp.	0.082	0.385	1.44	24.36	16.06	19.76	16.67	12.270	5.333	6.651	7.679
Łódź	0.065	0.070	1.19	24.01	11.67	12.70	11.89	8.621	3.084	3.389	3.670
Zielona Góra	0.067	0.238	1.41	10.03	8.96	9.63	9.62	3.908	2.682	2.917	3.781

which enables us to present several different datasets at one plot (see Fig. 7) grouping along one common line.

4 Log-periodic Oscillations on Path Lengths

In the previous sections we have shown that our results nicely cover the relation (3), both in case of network models as well as for real-world examples. However, our recent results (see [23]) leave no doubt that for some specific situations the relation between $\langle l_{ij} \rangle$ and $k_i k_j$ is not so straightforward. As it is presented at Fig. 8, in several examples we have encountered signs of a log-periodic behavior of internode distances, that is:

$$\langle l_{ij} \rangle = a - b \log k_i k_j + c \sin(\log k_i k_j). \tag{14}$$

We will try to explain that phenomenon using hidden variable formalism and its applications we have used in our previous work [24].

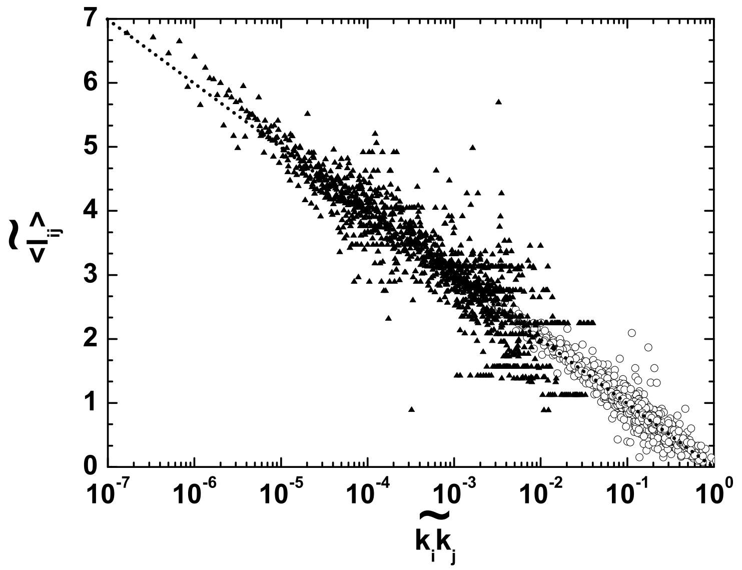

Fig. 7. Rescaled data from biological networks (filled triangles) and Public Transport Networks (open circles).

4.1 Hidden Variables

In [24] Fronczak et *al.* have derived exact expressions for average path lengths in case of uncorrelated random networks using a hidden variable approach. In this method every node i is assigned with a *hidden variable* h_i randomly drawn from a distribution $\rho(h)$ and the connection probability between any pair of nodes is proportional to $h_i h_j$. The resulting node degree distribution $P(k)$ is given by [25]:

$$P(k) = \sum_h \frac{\mathrm{e}^{-h} h^k}{k!} \rho(h). \tag{15}$$

It has been proved [24] that probability $p_{ij}^*(x)$ of nodes i and j being exactly x-th neighbors can be written as $p_{ij}^*(x) = F(x-1) - F(x)$, where:

$$F(x) = \exp\left[-\frac{h_i h_j}{\langle h^2 \rangle N} \left(\frac{\langle h^2 \rangle}{\langle h \rangle}\right)^x\right] \tag{16}$$

In the above equation N is network size and:

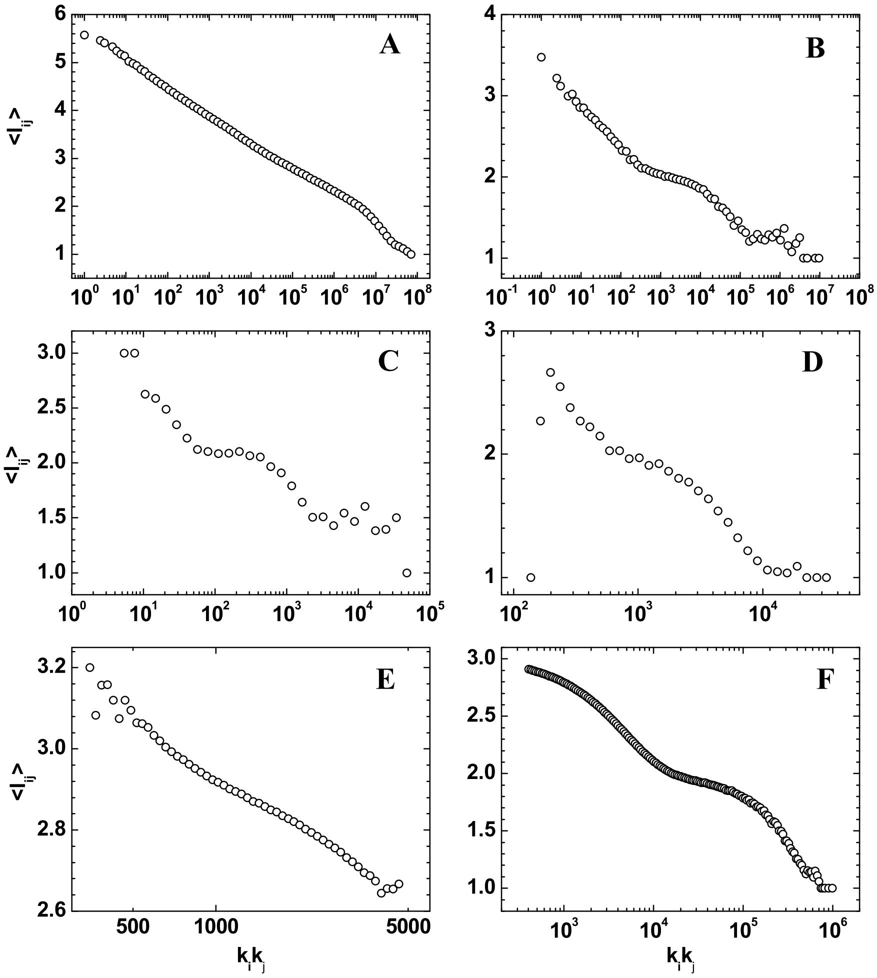

Fig. 8. Average distance $\langle l_{ij} \rangle$ between nodes i and j for real-world networks (A - D) and network models (E - F). (A) Movie actor network $N = 382219$, $\langle k \rangle = 86.64$ (B) Spanish language word cooccurrence network $N = 11857$, $\langle k \rangle = 7.43$ (C) Caribbean food web network $N = 249$, $\langle k \rangle = 25.73$ (D) Opole public transport network $N = 205$, $\langle k \rangle = 50.19$ (E) Erdös-Rényi random graph $N = 10000$, $\langle k \rangle = 40$. (F) Barabási-Albert evolving network $N = 10000$, $m = 20$. All data are logarithmically binned. Sources for data are following: data for movie actor have been taken from A.-L. Barabási web page http://www.nd.edu/ alb, Spanish cooccurrence dataset has been downloaded from V. Batagelj WWW page http://vlado.fmf.uni-lj.si/pub/networks/pajek/ and Caribbean food web data have been taken from The Integrative Ecology Group http://ieg.ebd.csic.es. Opole data have been collected for so called "space P" [9] which is defined as follows: nodes are bus and tram stops and an edge means that there is a direct route linking them.

$$\langle h\rangle = \int_{h_{min}}^{h_{max}} h\rho(h)\mathrm{d}h\,, \tag{17}$$

$$\langle h^2\rangle = \int_{h_{min}}^{h_{max}} h^2\rho(h)\mathrm{d}h\,. \tag{18}$$

As consequence of these assumptions, Fronczak et *al.* have been able to express the average distance $\langle l_{ij}\rangle$ between nodes i and j as:

$$\langle l_{ij}\rangle = \frac{-\ln h_i h_j + \ln N\langle h\rangle - \gamma}{\ln\frac{\langle h^2\rangle}{\langle h\rangle}} + \frac{3}{2} + R, \tag{19}$$

where $\gamma \simeq 0.5772$ is the Euler's constant and R is a sum of cosine Fourier transforms of $F(x)$ function:

$$R = \sum_{n=1}^{\infty} R_n \equiv 2\sum_{n=1}^{\infty}\left(\int_0^{\infty} F(x)\cos(2n\pi x)\mathrm{d}x\right), \tag{20}$$

After integrating Eq. (19) over all pairs $h_i h_j$ one can obtain an equation for average path length $\langle l\rangle$ which is in fact dependent only on N and on specific distribution $\rho(h)$:

$$\langle l\rangle = \frac{-2\langle\ln h\rangle + \ln N\langle h\rangle - \gamma}{\ln\frac{\langle h^2\rangle}{\langle h\rangle}} + \frac{3}{2} + S, \tag{21}$$

here S is the term R integrated over all hidden variables $h_i h_j$.

4.2 Average Internode Distance

Using appropriate hidden variable distributions $\rho(h)$ for different types of networks, one can obtain exact expressions for average internode distance from Eq. (3):

- *Scale-free (SF) networks*. A hidden variable distribution $\rho_{sf}(h)$ for asymptotically SF networks ($k \gg 1$) is given by the following expression [26]:

$$\rho_{SF}(h) = \frac{(\alpha-1)m^{\alpha-1}}{h^{\alpha}} \tag{22}$$

As in this work we pay attention only to the case, where $\alpha = 3$, thus after using Eqs (17) and (18) and substituting $h_{min} = m$ and $h_{max} = m\sqrt{N}$ one can easily show that:

$$\langle l_{ij}^{SF}\rangle = \frac{-\ln h_i h_j + \ln 2mN - \gamma}{\ln(m\ln\sqrt{N})} + \frac{3}{2} + R, \tag{23}$$

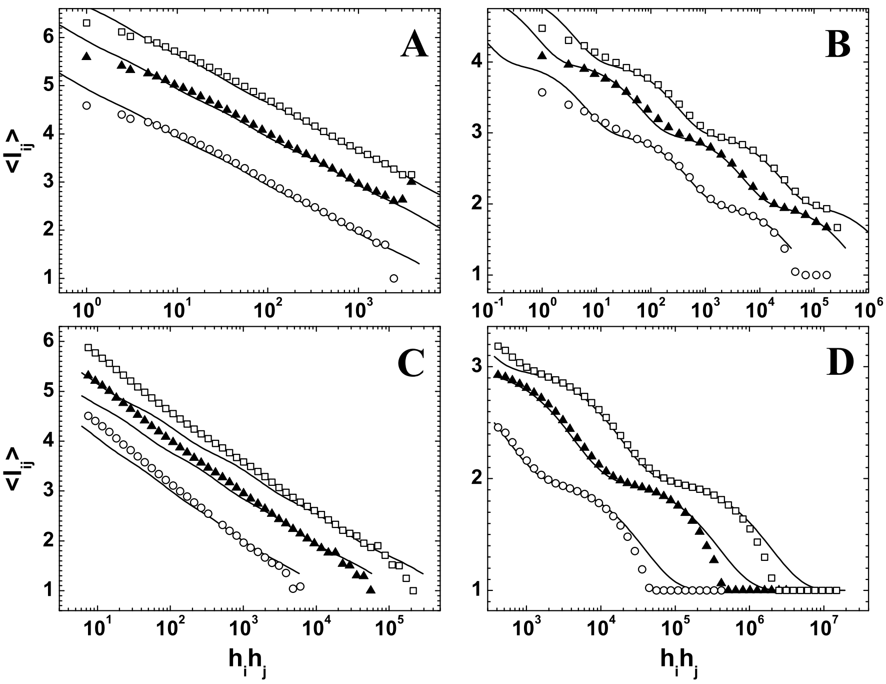

Fig. 9. Average internode distance $\langle l_{ij} \rangle$ between nodes i and j as a function of nodes' hidden variables product $h_i h_j$ for exponential networks (A - B) and scale-free networks with $\alpha = 3$. Solid lines are calculated using Eq. (3) (with numerical integration in order to obtain term R) while scatter data (circles - $N = 1000$, triangles - $N = 10000$ and squares - $N = 50000$) are numerical simulations using algorithm presented in [26]. Average degree $\langle k \rangle = 5$ for (A) and (C) and $\langle k \rangle = 40$ for (B) and (D).

- *Exponential (EXP) networks.* In the case of the exponential networks the distribution of hidden variables takes a form of

$$\rho_{EXP}(h) = \frac{1}{\langle k \rangle} e^{-\frac{h}{\langle k \rangle}}. \tag{24}$$

  Using Eq. (3) one gets:

$$\langle l_{ij}^{EXP} \rangle = \frac{-\ln h_i h_j + \ln N \langle k \rangle - \gamma}{\ln(2\langle k \rangle)} + \frac{3}{2} + R, \tag{25}$$

- *Erdös-Rényi (ER) random graphs.* As the only way to obtain Poisson distribution from Eq. (15) is to define $\rho_{ER}(h) = \delta_{\langle k \rangle, h}$ (i.e. every node i is characterized by a hidden variable $h_i = \langle k \rangle$), thus examining Eq. (3) is useless.

A comparison between theoretical predictions of Eqs (23) and (25) and numerical simulations performed using algorithm [26] for scale free networks with the hidden variable distributions (22) and (24) is presented at Fig. 9. It is to notice that for small values of $\langle k \rangle$ (Fig. 9A and 9B) the scaling relation is well described with Eq. (3) while for larger average degree values oscillations appear.

4.3 Average Path Length

An average path length is a variable of much interest, because it is a main observable for many complex networks.

Once again, applying hidden variable distributions defined in the previous Section to Eq. (21) we can obtain exact expressions for average path length:

- *Scale-free networks.*

$$\langle l_{SF} \rangle = \frac{\ln \frac{2N}{m} - 1 - \gamma}{\ln(m \ln \sqrt{N})} + \frac{3}{2} + S, \tag{26}$$

- *Exponential networks.*

$$\langle l_{EXP} \rangle = \frac{\ln \frac{N}{\langle k \rangle} + \gamma}{\ln 2\langle k \rangle} + \frac{3}{2} + S, \tag{27}$$

- *Erdös-Rényi.*

$$\langle l_{ER} \rangle = \frac{\ln \frac{N}{\langle k \rangle} - \gamma}{\ln\langle k \rangle} + \frac{3}{2} + S. \tag{28}$$

Figure 10 gathers numerical simulations and theoretical predictions for all three types of networks, showing that for small values of average degree the dependence of $\langle l \rangle$ can be well described using Eqs (26-28) with neglected term S (Fig. 10A). However, as $\langle k \rangle$ increases (Fig. 10B), this relation is no longer valid and it necessary to include correction caused by term S. One should notice, that for exponential networks even $\langle k \rangle = 40$ does not imply the appearance of oscillations.

4.4 Oscillations

From the previous sections one can spot that as long as the average number of links remains relatively small then, due to the generalized mean value theorem, the term R can be neglected. Otherwise one must take into account at least the first term from the infinite series in Eq. (20) what leads to log-periodic oscillation$\langle l_{ij} \rangle$ with the period $\Delta \ln(h_i h_j) = \ln B$ (see discussion below). In the next consideration we use symbols $A = (h_i h_j)/(\langle h^2 \rangle N)$ and $B = \langle h^2 \rangle / \langle h \rangle$ to avoid too complicated formulas.

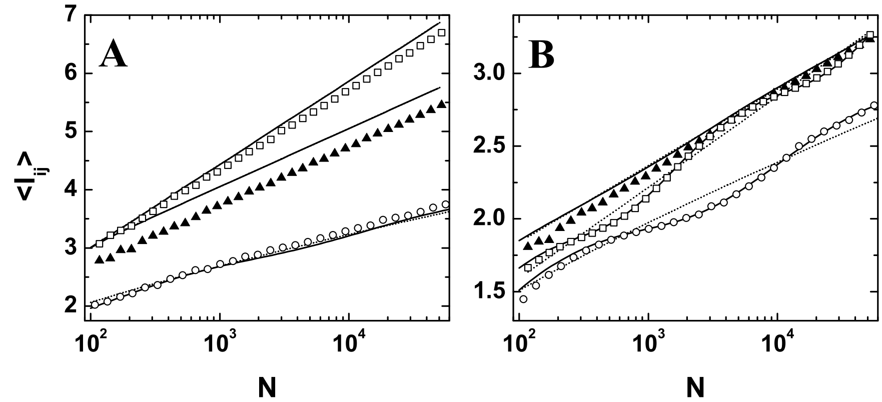

Fig. 10. Average path length $\langle l \rangle$ versus network size N. Scatter data are numerical simulations for scale-free networks with $\alpha = 3$ (circles), exponential networks (triangles) and Erdös-Rényi random graphs (squares), solid lines are calculated using Eq. (21) (with numerical integration in order to obtain term R), dotted lines are also obtained from Eq. (21), but term R is neglected. (A) $\langle k \rangle = 5$ for ER graphs and exponential networks and $\langle k \rangle = 8$ for SF networks. (B) $\langle k \rangle = 40$ for ER graphs and exponential networks and $\langle k \rangle = 60$ for SF networks. Solid and dotted lines corresponding to exponential and ER networks at (A) and to exponential networks at (B) have the same values, thus only the solid ones are visible.

Figure 11 shows a comparison of such oscillations in sparse ($m = 2$, upper row) and dense ($m = 40$, lower row) scale-free networks characterized by a hidden variable distribution $\rho(h) = (\alpha - 1)m^{\alpha-1}h^{-\alpha}$ with $\alpha = 3$. The networks have been generated following the procedure C in [26] and represent the class of random networks with asymptotic scale-free connectivity distributions characterized by an arbitrary scaling exponent $\alpha > 2$. Figure Fig. 11A presents $F(x)$ (dotted line) and p^*_{ij} (solid line) are along with points corresponding to discrete values of those functions. One can notice that for $m = 40$ probability p^*_{ij} is much more narrow than for $m = 2$, so the slope of $F(x)$ decays more rapidly in the first case. Figure 11B shows the cosine transform of $F(x)$ given by Eq. (20). Depending on the shape of $F(x)$, the amplitude of this transform can take small/large values resulting in small/large values of R. As R is a sum of discrete values of a given transform taking only the first term in the sum (i.e. $n = 1$) is sufficient to obtain well approximated value of R (cf. points corresponding to discrete values of R_n at Fig. 11B). Figure 11C shows different behavior of resulting average distance $\langle l_{ij} \rangle$ between nodes i and j versus the product $h_i h_j$ for $m = 2$ (upper plot) and $m = 40$ (lower plot).

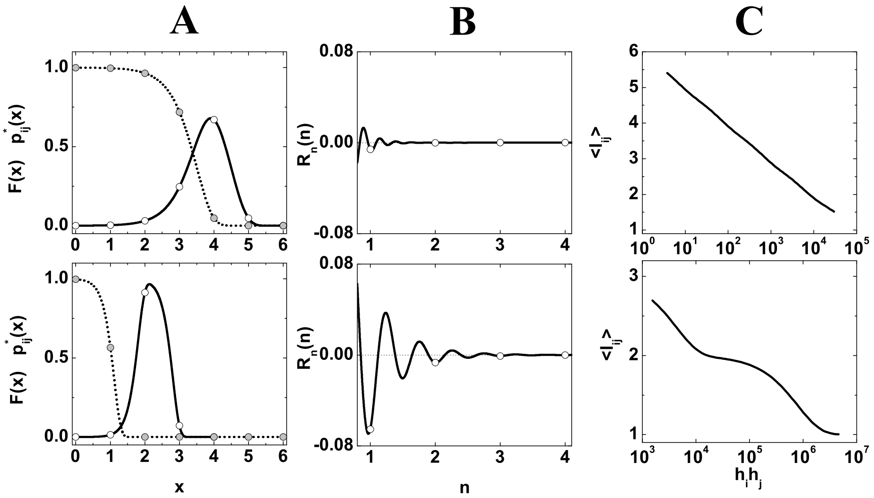

Fig. 11. Comparison of Two Networks Characterized by Hidden Variable Distribution $\rho(h) = (\alpha - 1)m^{\alpha-1}h^{-\alpha}$ for $\alpha = 3.0$ and $N = 10000$ - upper row $m = 2$, lower row $m = 40$. (A, B, C) - detailed description in text and. In case of plots (A) and (B) values of A have been chosen in such a way that the deviation is maximal.

To obtain more quantitative results one should perform the integral in Eq. (20), however it is not analytical, so we made an approximation of $F(x)$ (see Appendix B) receiving a following estimation for R:

$$\widetilde{R}_n = -\frac{\ln B \sin\left(\frac{\pi n \mathrm{e}}{\ln B}\right)}{\pi^2 n^2 \mathrm{e}} \sin\left[\frac{\pi n}{\ln B}\left(2\ln A - 2 + \mathrm{e}\right)\right]. \tag{29}$$

As one can see taking only the first term (i.e. $n = 1$) from Eq. (29) is justified because next terms decay as $1/n^2$. Equation (29) allows us to make an immediate observation that deviations from Eq. (3) take the form of regular oscillations along $h_i h_j$ axis with period equal to $\ln B$ which increases with the heterogeneity of the networks. For dense networks the amplitude of oscillations grows monotonically with B - that is why the effect of oscillations is visible only in sufficiently dense networks. Similar oscillations effects are also observed for average path length $\langle l \rangle$ (see Fig. 10).

4.5 Cost Function

To show a practical way to make use of the phenomenon described in the previous Sections, we will concentrate on network optimization- a problem so far described in many works [10, 27, 28]. Optimization issues are well known and appear in different fields such as telecommunication and road construction. One of the most simplest model that assumes minimal costs in transportation includes two main aspects of network performance: the cost of constructing and maintaining the links between nodes and the cost of communication speed among them. The first part is proportional to the total number of links and the second one is proportional to the sum of shortest connections between each pair of nodes:

$$C = (1-\lambda)\,\frac{N}{2}\langle k\rangle + \lambda \binom{N}{2} \langle l\rangle. \qquad (30)$$

where λ controls the linear combination of two demands in Eq. (30): fully connected network with the shortest connections and a tree with the smallest number of links. A standard solution of this problem is a cost function with minimum at some specific value of $\langle k\rangle$ (see [10]).

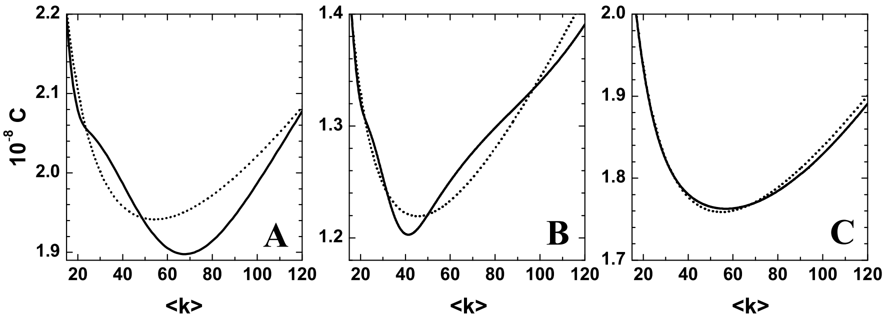

Fig. 12. Cost function C versus average degree $\langle k\rangle$ for three different networks of $N = 10^6$ nodes each. Solid lines represent cost function defined in Eq. (30) while dotted lines neglect the existence of S term. (A) Scale-free network with exponent $\alpha = 3$ $\lambda = 10^{-4}$ (B) Erdös-Rényi random graph $\lambda = 5 \cdot 10^{-5}$ (C) Exponential network $\lambda = 2 \cdot 10^{-5}$

Effects of discretization presented in the previous Sections change the shape of the cost function. Figure 12 depicts cost function for three different types of networks of $N = 10^6$ nodes each: scale-free with exponent $\alpha = 3$, exponential and Erdös-Rényi random graph. One can see that effects of oscillations induce a shift in the position and a decrease in value of the minimum of the cost function in the case of scale-free and Erdös-Rényi networks.

Figure 13 shows that different values of parameter λ can lead to even more interesting situation, where instead of one global minimum we obtain two well separated minima. The network adaptation to lower cost conditions can lead now to a temporal increase of costs since one has to pass over a cost barrier.

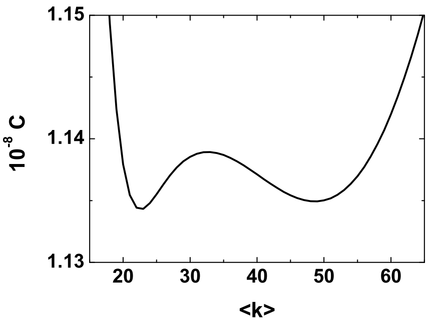

Fig. 13. Cost function C versus average degree $\langle k \rangle$ for scale-free network ($\alpha = 3$) of $N = 10^6$ nodes with $\lambda = 5.4 \cdot 10^{-4}$

5 Summary

In the last Section we presented results that possess applications for commonly known problems - optimization of such structures as telecommunication or public transport networks plays a crucial role in the present life. Without an extended analysis of these systems as complex networks we would not have been able to spot this phenomenon. In other words, the problem itself may not be seen as complex one at first glance, however a closer examination of its ingredients can lead to a reformulation and thus to a significant change of the results.

6 Acknowledgements

J.S., K.S. and J.A.H. acknowledge a support from the EU Grant MMCOMNET No. FP6-2003-NEST-Path-012999 and from Polish Ministry of Science and Higher Education (Grant No. 13/6.PR UE/2005/7). P.F. acknowledges a support from the EU Grant CREEN FP6-2003-NEST-Path-012864 and from Polish Ministry of Science and Higher Education (Grant No. 134/E-365/6.PR UE/DIE 239/2005-2007).

A Rescaling of $\langle l_{ij} \rangle$ and $k_i k_j$.

For the majority of examined networks the average intervertex distance $\langle l_{ij} \rangle$ described with relation (3) takes its maximal value (i.e the maximal intervertex distance) l_{max} for the minimal possible values of k_i and k_j, that is $k_i = k_j = 1$. Then applying the above conditions to (3) we get

$$a = l_{max}. \tag{31}$$

In order to obtain the second parameter of (3), we observe that the straight line described by (3) crosses X-axis for some effective value k_{eff}^2 i.e. $k_i = k_j = k_{eff}$. This allows as to express b as

$$b = \frac{l_{max}}{2 \log k_{eff}}. \tag{32}$$

Applying a and b to Eq. (3) after some algebra we arrive at the following expression:

$$\frac{\langle l_{ij} \rangle}{l_{max}} 2 \log k_{eff} = -\log \left(\frac{k_i k_j}{k_{eff}^2} \right). \tag{33}$$

Introducing characteristic values $\widetilde{\langle l_{ij} \rangle}$ and $\widetilde{k_i k_j}$ (see Eqs (11) and (12)) we get a universal formula for the scaling relation :

$$\widetilde{\langle l_{ij} \rangle} = -\log \widetilde{k_i k_j} \tag{34}$$

B Approximation of function $F(x)$

In order to calculate the term R one can approximate $F(x)$ with the following piecewise linear function $\widetilde{F}(x)$:

$$\widetilde{F}(x) = \begin{cases} 1 & x < x_0, \\ \frac{1}{e} (1 - \ln A - x \ln B) & x \in < x_0, x_1 >, \\ 0 & x > x_1, \end{cases} \tag{35}$$

where $x_0 = (1 - \ln A - e)/\ln B$ and $x_1 = (1 - \ln A)/\ln B$. Since the function $F(x)$ is translationally invariant with respect to the argument x after rescaling the parameter A ($F(x; A) = F(x - x'; A')$) one can freely choose the point in which the slope coefficient is calculated as the tangent of $F(x)$. In our case in order to simplify the calculation we have chosen the inflexion point x_i of $F(x)$. Using Eq. (35) one can approximate terms R_n with

$$\widetilde{R}_n = -\frac{\ln B \sin\left(\frac{\pi n e}{\ln B}\right)}{\pi^2 n^2 e} \sin\left[\frac{\pi n}{\ln B}(2 \ln A - 2 + e)\right]. \tag{36}$$

and thus one obtains an expression responsible for the oscillating term in $\langle l_{ij} \rangle$ scaling.

References

1. Euler, L.: Comment. acad. sc. Petrop. **8** (1741) 128
2. Erdös, P. and Rényi, A.: Publ. Math. Debrecen **6** (1959) 290
3. Watts, D. J. and Strogatz, S.H.: Nature **393** (1998) 440

4. Albert, R., Jeong, H., Barabási, A.–L.: Nature **401** (1999) 130
5. Milgram, S.: Psychology Today **2** (1967). 60
6. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.H. and Barabási, A.–L.: Nature **407** (2000) 651
7. Pastor-Satorras, R. and Vespignani, A.: Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge University Press, Cambridge, 2004
8. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto: Phys. Rev E **68** (2003) 056110
9. Sienkiewicz, J., Hołyst, J.A.: Phys. Rev E **72** (2005) 046127
10. Schweitzer, F.: Brownian Agents and Active Particles. Springer, Berlin (2003)
11. Albert,R. and Barabási, A.–L.: Rev. Mod. Phys. **74** (2002) 47
12. Newman, M.E.J.: Phys. Rev. Lett. **89** (2002) 208701
13. Hołyst, J.A., Sienkiewicz, J., Fronczak, A., Fronczak, P. and Suchecki, K.: Phys. Rev E **72** (2005) 026108
14. Data for Foodwebs (*Silwood* and *Ythan*) [15] have been taken from www.cosin.org/extra/data/foodwebs/ whereas data for protein interaction network (*Yeast*) [16] has been taken from www.nd.edu/ networks/database/index.html.
15. Garlaschelli, D., Caldarelli, G. and Pietronero, L.: Nature **423** (2003) 165
16. Jeong, H., Mason, S.P., Barabási, A.–L. and Oltvai, Z.N.: Nature **411** (2001) 41
17. Scientific collaboration network [18] 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.
18. Newman, M.E.J.: Phys. Rev. E **64** (2001) 016131
19. Data for the Internet [7] has been taken from www.cosin.org/extra/data/internet/.
20. In those networks vertices are bus- and tramstops while an edge exists if at least one public transport line crosses two stops. [9].
21. Motter, A.E., Nishikawa, T. and Lai, Y.C.: Phys. Rev. E **66** (2002) 065103(R)
22. Hołyst, J.A., Sienkiewicz, J, Fronczak, A., Fronczak, P. and Suchecki, K.: Physica A **351** (2005) 167
23. J. Sienkiewicz, P. Fronczak, J. A. Hołyst: e-print cond-mat/0608273 (2006)
24. Fronczak, A., Fronczak, P. and Hołyst, J.A.: Phys. Rev. E **70** (2004) 056110
25. Boguñá, M. and Pastor-Satorras, R.: Phys. Rev. E **68** (2003) 036112
26. Fronczak, A. and Fronczak, P.: Phys. Rev. E **74** (2006) 026121
27. Gastner, M.T. and Newman, M.E.J.: Eur. Phys. J. B **49** (2006) 247
28. Ferrer i Cancho, R. and Solé, R.: Statistical Physics of Complex Networks, Lecture Notes in Physics. Springer, Berlin (2003)