Log-periodic oscillations due to discrete effects in complex networks

Julian Sienkiewicz, Piotr Fronczak, and 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

(Received 11 August 2006; revised manuscript received 6 February 2007; published 13 June 2007)

We show how discretization affects two major characteristics in complex networks: internode distances (measured as the shortest number of edges between network sites) and average path length, and as a result there are log-periodic oscillations of the above quantities. The effect occurs both in numerical network models as well as in such real systems as coauthorship, language, food, and public transport networks. Analytical description of these oscillations fits well numerical simulations. We consider a simple case of the network optimization problem, arguing that discrete effects can lead to a nontrivial solution.

DOI: 10.1103/PhysRevE.75.066102 PACS number(s): 89.75.Hc, 02.50.−r, 05.50.+q

I. INTRODUCTION

Complex networks that represent such sophisticated systems as acquaintances between people [1], collaboration of scientists [2], protein interactions [3], Internet [4] or city public transport [5] have drawn much attention during the last few years. Among other network characteristics, the most frequently examined features are internode distance defined as a number of edges along the shortest path connecting two vertices and an average path length—the mean value of above quantity. Several authors [6–10] have dealt with this problem using different approaches to obtain analytical expressions for average path lengths. Such a high concern regarding distance features can be justified by the following reasons: (i) the fundamental purpose of many networks is to link distant nodes in a way that enables efficient flow of information, traffic, etc., (ii) average path length is considered as the most natural observable in complex networks, (iii) most of the examined systems exhibit the property of small average path length $\langle l\rangle$ in comparison to the network size N (i.e., $\langle l\rangle \sim \ln N$ or even $\langle l\rangle \sim \ln\ln N$), (iv) identifying mathematical description of the distance between two nodes (i.e., virus infected individuals [11], crucial chemical compounds [3], stock assets [12]) allows us to judge on the stability of the system and its tolerance to failures [13].

In this paper we study a simple effect of log-periodic oscillations in average path lengths which we observe in several real-world examples. Using a formalism developed in [10] we give a theoretical explanation of this feature supported by numerical simulations of scale-free networks with different scaling exponents. We show that such oscillations are due to discrete effects of path length distributions for networks with large average degree values. We also study a fundamental and well-known problem of optimal network density taking into account the shortest average path length and the smallest number of links in a network [15]. We find that the oscillations substantially influence the solution of this problem.

II. HIDDEN VARIABLES

Recently it has been shown [16,17] that the average distance $\langle l_{ij}\rangle$ between nodes i and j characterized by degrees k_i and k_j can be expressed as

$$\langle l_{ij}\rangle = a - b\ln(k_i k_j). \tag{1}$$

This relation is fulfilled in a wide spectrum of real-world networks and their models such as random graphs or Barabási-Albert evolving networks [17], however our recent research shows deviations from this scaling law which take a form of regular oscillations. This can be clearly seen in Fig. 1 where four real-world networks and two common known models have been gathered.

To explain differences between Eq. (1) and plots in Fig. 1 we will use and modify results obtained by Fronczak *et al.* in [10]. In the cited paper exact expressions for average path length using hidden variables formalism have been received. Assuming that each node i is characterized by its hidden variable h_i randomly drawn from a given distribution $\rho(h)$, and a connection probability between any pair of nodes is proportional to $h_i h_j$ one can show [19] that a degree distribution $P(k)$ is

$$P(k) = \sum_h \frac{e^{-h}h^k}{k!}\rho(h). \tag{2}$$

The probability $p^*_{ij}(x)$ that vertices i and j are xth neighbors can be expressed [10] as $p^*_{ij}(x)=F(x-1)-F(x)$, where

$$F(x) = \exp(-AB^x) \tag{3}$$

and $A=\frac{h_i h_j}{\langle h^2\rangle N}$, $B=\frac{\langle h^2\rangle}{\langle h\rangle}$. One should keep in mind that the parameter B is a "global" one (i.e., its value is determined only by the first and second moment of a hidden variable distribution), while A can be called "local," it depends on a specific product $h_i h_j$. As the expectation value of average distance between i and j can expressed as $\langle l_{ij}\rangle=\sum_{x=1}^{x=\infty} x p^*_{ij}(x)$ $=\sum_{x=0}^{x=\infty}F(x)$, one can write the following equation using Poisson summation formula (see Appendix B in [10]):

$$\langle l_{ij}\rangle = \frac{-\ln A - \gamma}{\ln B} + \frac{1}{2} + R,$$

$$R = \sum_{n=1}^{\infty} R_n \equiv 2\sum_{n=1}^{\infty}\left(\int_0^{\infty} F(x)\cos(2n\pi x)dx\right), \tag{4}$$

where γ=0.5772 is Euler's constant. If the average number

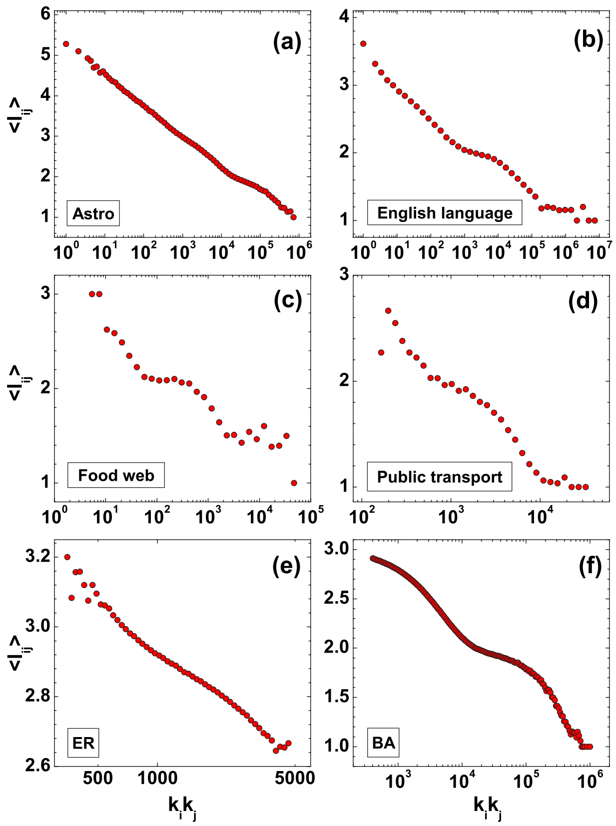

FIG. 1. (Color online) Mean distance $\langle l_{ij} \rangle$ between pairs of nodes i and j as a function of a product of their degrees $k_i k_j$ for four real-world networks and two models. (a) Astro coauthorship network, $N=13\,986$, $\langle k \rangle=25.56$; (b) English language word co-occurrence network, $N=7381$, $\langle k \rangle=11.98$; (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, $N=10\,000$, $\langle k \rangle=40$. (f) Barabási-Albert network, $N=10\,000$, $m=20$. All data are logarithmically binned. In the case of real-world networks an edge between two nodes exists if (a) one scientists cites another, (b) two words are consecutive, (c) one specie devours another, and (d) there is a direct route connecting two bus or tram stops. For data sources see [18,14].

of links is relatively small then the function $F(x)$ decreases *slowly*, terms in the cosine transform of such a smooth function are negligible and the term R can be neglected. Otherwise one must take into account, at least for the first term from the infinite series in Eq. (4), what leads to log-periodic oscillation $\langle l_{ij} \rangle$ with the period $\Delta \ln(h_i h_j)=\ln B$ (see the dicussion below).

III. OSCILLATIONS

Figure 2 shows a comparison of such oscillations for networks with different link density [given by the parameter m in the formula $\rho(h)$ below]: low ($m=2$, upper row) and high ($m=40$, lower row). The examined systems are 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 [20] and represent the class of random networks with asymptotic scale-free connectivity distributions characterized by an arbitrary scaling exponent $\alpha>2$. In Fig. 2(b) $F(x)$ (dotted line) and p_{ij}^* (solid line) are presented together with points corresponding to discrete values of those functions. It is clearly seen that for $m=40$ probability p_{ij}^* is much more narrow than for $m=2$, thus the slope of $F(x)$ decays more rapidly. Figure 2(c) shows the cosine transform of $F(x)$ given by the integral in Eq. (4). Depending on the shape of $F(x)$, the amplitude of this transform can take small (large) values resulting in small (large) values of R. One should keep in mind that because R is in fact 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 a well-approximated value of R [cf. points corresponding to discrete values of R_n in Fig. 2(c)]. Figure 2(d) shows resulting average distance $\langle l_{ij} \rangle$ between nodes i and j as a function of hidden variables $h_i h_j$ without (dotted lines) and with (solid lines) included term R. In the case of sparse network the R term can be omitted (curves overlap), while for a dense one its value modifies the shape of $\langle l_{ij} \rangle$ a lot.

To obtain more quantitative results one should perform the integral in Eq. (4), however it is not analytical, so in order to calculate the term R one can approximate $F(x)$ with the following piecewise linear function $\tilde{F}(x)$:

$$\tilde{F}(x) = \begin{cases} 1, & x < x_0, \\ \frac{1}{e}(1 - \ln A - x \ln B), & x \in \langle x_0, x_1 \rangle, \\ 0, & x > x_1, \end{cases} \tag{5}$$

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 order to simplify the calculation we have chosen the inflexion point x_i of $F(x)$. Functions $\tilde{F}(x)$ and $F(x)$ are presented in Fig. 3. Using Eq. (5) one can approximate terms R_n with

$$\tilde{R}_n = -\frac{\ln B \sin\left(\frac{\pi n e}{\ln B}\right)}{\pi^2 n^2 e} \sin\left(\frac{\pi n}{\ln B}(2\ln A - 2 + e)\right). \tag{6}$$

As one can see taking only the first term in the above series is justified because the next terms decay as $1/n^2$. Equation (6) allows us to make an immediate observation that deviations from Eq. (1) take the form of regular oscillations along the $h_i h_j$ axis with period equal to $\ln B$ which increases with the heterogeneity of the networks (see the inset in Fig. 3). This very value is connected with the *discrete nature of distance* in network—the period along $\langle l_{ij} \rangle$ is equal to 1 (see explanation in Appendix A) and the tangent of the function $\langle l_{ij} \rangle(h_i h_j)$ is $(\ln B)^{-1}$ [see Eq. (4)]. One can also easily calculate that the deviation vanishes as long as $\langle l_{ij} \rangle \approx k/2$ where k is an integer (see Appendix B). 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. Figure 4 presents a comparison of average

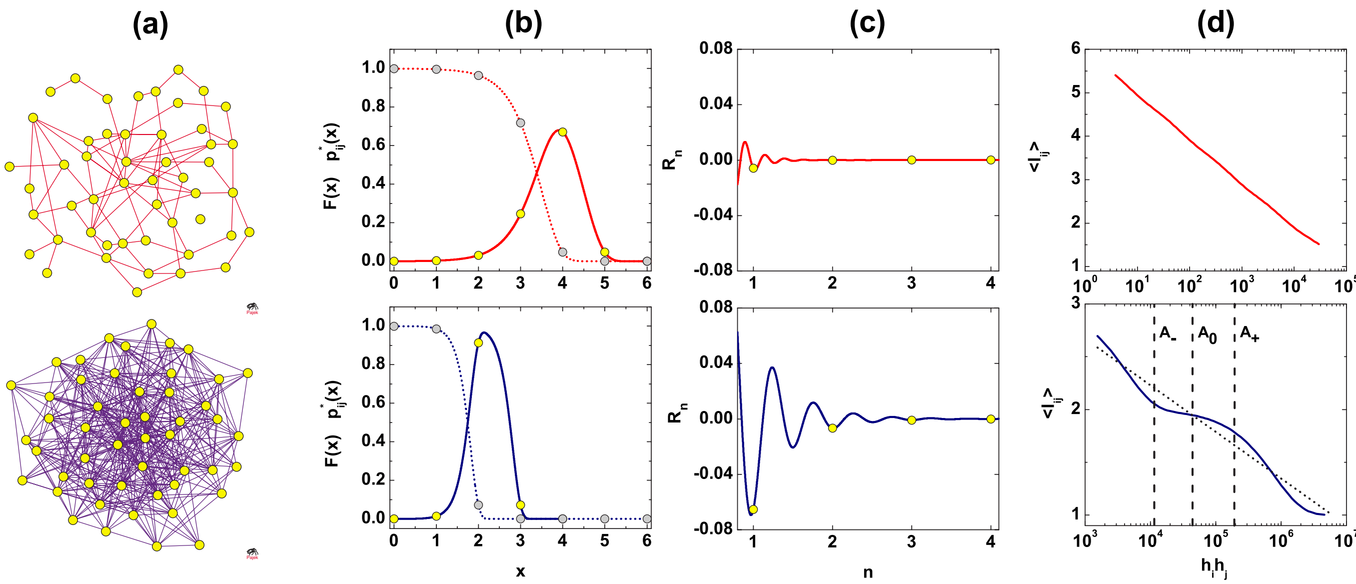

FIG. 2. (Color online) Comparison of two networks characterized by hidden variable distribution $\rho(h)=(\alpha-1)m^{\alpha-1}h^{-\alpha}$ for $\alpha=3.0$ and $N=10\,000$—upper row $m=2$, lower row $m=40$. (a) Samples of sparse (upper) and dense (lower) networks; (b), (c), (d) detailed description in text and in caption of Fig. 3. In case of plots (b) and (c) values of A have been chosen in such a way that the deviation is maximal.

distance $\langle l_{ij}\rangle$ versus h_ih_j for scale-free networks with different scaling exponents α. As expected, the amplitude of oscillations rises with decaying α, which can be easily understood as $\ln B_{\alpha_1} > \ln B_{\alpha_2}$ for $\alpha_1 < \alpha_2$. Similar oscillation effects can also be observed for average path length $\langle l\rangle$, which value is obtained by integration of Eq. (4) over all pairs of products h_ih_j [see inset (b) in Fig. 5].

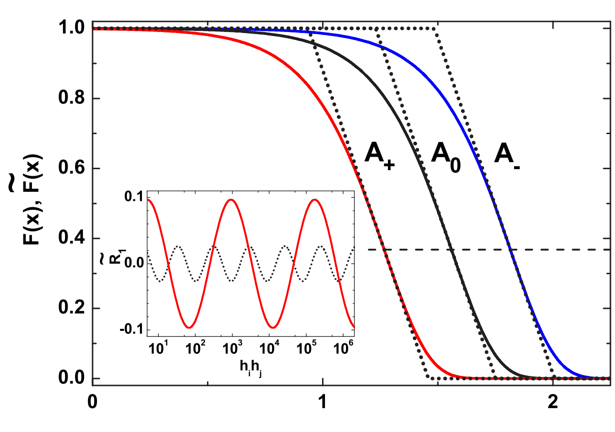

FIG. 3. (Color online) Function $F(x)$ (solid lines) and its linear approximation $\widetilde{F}(x)$ (dotted lines) for scale-free network with $\alpha=3$, $N=10\,000$ and $m=40$ calculated for three different values of the product h_ih_j (see labeled dashed lines in Fig. 2): (A_-) $h_ih_j=11\,389$, maximal negative deviation from the $\langle l_{ij}\rangle$ trend; (A_0) $h_ih_j=43\,249$ minimal (zero) deviation from the $\langle l_{ij}\rangle$ trend; (A_+) $h_ih_j=198\,730$, maximal positive deviation from the $\langle l_{ij}\rangle$ trend. Dashed line represents the point of inflexion x_i of $F(x)$ $[F(x_i)=1/e]$ used to calculate tangent of $F(x)$. Inset shows $\widetilde{R}_1$ versus product h_ih_j in the case of $m=2$ (dotted line) and $m=40$ (solid line).

IV. COST FUNCTION

Let us now focus on possible applications of the presented phenomenon. One of them can be a network optimization process which has been widely studied in recent years [15,21,22]. Such an optimization is of common interest in many different areas, among them electrical engineering, telecommunication, road construction, and trade logistics. The simplest model is based on the assumption of minimal transport costs. These costs include two main aspects of network performance: a price of constructing and maintaining links between nodes and a price caused by communication delays of information transfer. The former one is proportional to the total number of links (we assume the same price for every link), while the latter one should be proportional to the sum of the shortest existing connections between each two nodes:

$$C=(1-\lambda)\frac{N}{2}\langle k\rangle+\lambda\binom{N}{2}\langle l\rangle. \qquad (7)$$

Here λ is a parameter controlling a ratio between prices of a single link and costs of communication delays. In fact one must find an optimal link density considering two contradictive demands: a fully connected network with the shortest connections and a tree with the smallest number of links. A typical solution of this problem is a unimodal cost function with minimum at some intermediate value of $\langle k\rangle$.

Discrete effects in networks studied above can lead to reshaping of the total cost function. As an example let us consider the scale-free network generated by the method described in [20] with parameters $N=10^6$ and $\alpha=3$. The cost function for this network is presented in Fig. 5 (we also show how this function could look if we neglected discrete effects). One can see that neglecting the correction term can

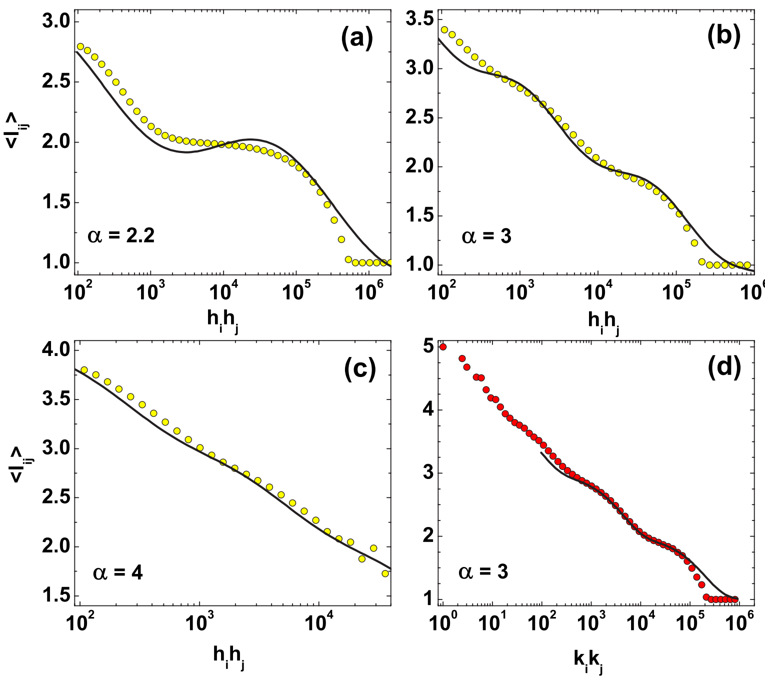

FIG. 4. (Color online) Average distance $\langle l_{ij} \rangle$ between nodes i and j versus their hidden variable product $h_i h_j$ [plots (a), (b), and (c)] or $k_i k_j$ (d) for scale-free networks (m=10) of N =10 000 nodes and α=2.2 (a), α=3 (b) and (d), and α=4 (c). Scatter data are obtained using the algorithm presented in [20] while solid lines have been calculated from Eq. (4) where R is taken directly from Eq. (6).

lead to about 30% underestimation of optimal network density. Inset (a) in Fig. 5, obtained for another value of the parameter λ, shows a different situation—instead of one global minimum we have now two well separated minima. The network administrator who tries to operate in accordance with the economic rule (7) must to remember that the improvement of network efficiency can lead to a temporal increase of costs and can be discouraging since one must pass over the cost barrier. Much simpler application of the observed phenomenon is presented in Fig. 5(b), where one can see that during the network growth there are regions where average path length increases slower (faster) which can encourage (discourage) the network administrator for further network expansion.

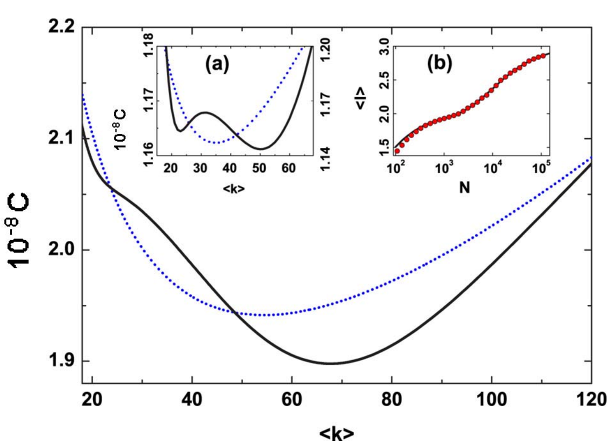

FIG. 5. (Color online) Cost function C versus average degree $\langle k \rangle$ for the scale-free network characterized by N=10^6 nodes, α=3.0 and λ=10^{-4}. Solid line is obtained assuming oscillations' correction while the dotted line neglects it. Inset (a) shows cost function C for identical network parameters N and α but with λ=5.4×10^{-4}. Left-hand Y axis corresponds to cost function with oscillations' correction (solid line) while right-hand Y axis corresponds to the function that neglects the correction (dotted line). Inset (b) presents average path length $\langle l \rangle$ versus system size N for scale-free network with α=3 and m=40: solid line is theory, while scatter data have been obtained using the hidden variable algorithm [20].

V. CONCLUSIONS

To summarize, we have observed log-periodic oscillations in a large class of complex networks models as well as in social, biological, and transport networks. The phenomenon emerges due to the discretiztion effects and is significant for dense networks. We have shown that the oscillations can substantially influence a solution of the network optimization problem when the utility function includes both link prices as well as costs of delays of internode communication.

ACKNOWLEDGMENTS

Two of the authors (J.S. and J.A.H.) acknowledge support from the EU grant Measuring and Modelling Complex Networks Across Domains—MMCOMNET (Grant No. FP6-2003-NEST-Path-012999) and from Polish Ministry of Science and Higher Education (Grant No. 13/6.PR UE/2005/7). One of the authors (P.F.) acknowledges a support from the EU Grant No. 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). The authors are thankful to Agata Fronczak for fruitful discussions.

APPENDIX A

To give the intuition behind the main argument of the paper and to better understand the discrete nature of distances in networks let us concentrate on the lower row of Fig. 2(b) (i.e., on a case of a dense network). As one can expect, the variance of the probability $p_{ij}^*(x)$ is close to zero because the most probable distance x between nodes is equal to 2 and the influence of other distances $(1,3,4,5,\ldots)$ can be neglected.

Let us assume that the function $p_{ij}^*(x)$ has a maximum at $x=2.5$ (i.e., for different value of parameter A). Now, the variance is maximal since distances $x=2$ and $x=3$ give contribution to it to the same extent. Finally, if $p_{ij}^*(x)$ has a maximum at $x=3$ then again a variance will be minimal. It means that a probability variation emerges because the distances are expressed only by natural numbers. The same discrete values generate oscillations of internode distances with the period along $\langle l_{ij}\rangle$ equal to 1.

APPENDIX B

The obvious condition to obtain values of $\langle l_{ij}\rangle$ for which the oscillations disappear is $\widetilde{R}=0$. However, taking into account that the terms in the series of Eq. (6) decay as $1/n^2$ we approximate this condition as $\widetilde{R}_1=0$. Thus, using this Eq. (6) we have

$$\frac{\pi}{\ln B}(2\ln A-2+e)=-k\pi, \tag{B1}$$

where $k=0,1,2,\ldots$. Extracting $\ln A$ from the above equation and substituting it in Eq. (4) we obtain an equation for $\langle l_{ij}\rangle$,

$$\langle l_{ij}\rangle=\frac{k+1}{2}+\frac{\frac{e}{2}-1-\gamma}{\ln B}. \tag{B2}$$

For sufficiently large and well connected networks the second term is relatively small in comparison to $\langle l_{ij}\rangle=1$ (i.e., for scale-free networks with $\alpha=3$, $N=10\,000$, and $m=20$, the second term is equal to 0.05). Thus we can renumerate k as $\widetilde{k}=k+1$ and write the following approximate equation:

$$\langle l_{ij}\rangle\approx\frac{\widetilde{k}}{2} \tag{B3}$$

for $\widetilde{k}=2,3,\ldots$. This allows us to state, that oscillations disappear as long as the value of $\langle l_{ij}\rangle$ is a multiple of 1/2 [cf. crossings of solid and dotted lines at bottom plot of Fig. 2(d)].

[1] S. Milgram, Psychol. Today **2**, 60 (1967).
[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[3] H. Jeong, B. Tombor, R. Albert, Z. H. Oltavi, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
[5] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).
[6] M. E. J. Newman, C. Moore, and D. J. Watts, Phys. Rev. Lett. **84**, 3201 (2000).
[7] G. Szabó, M. Alava, and J. Kertész, Phys. Rev. E **66**, 026101 (2002).
[8] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[9] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Nucl. Phys. B **653**, 307 (2003).
[10] A. Fronczak, P. Fronczak, and J. A. Hołyst, Phys. Rev. E **70**, 056110 (2004).
[11] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004).
[12] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, Phys. Rev. E **68**, 056110 (2003).
[13] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[14] J. Sienkiewicz and J. A. Hołyst, Phys. Rev. E **72**, 046127 (2005).
[15] F. Schweitzer, *Brownian Agents and Active Particles* (Springer, Berlin, 2003).
[16] A. E. Motter, T. Nishikawa, and Y. C. Lai, Phys. Rev. E **66**, 065103(R) (2002).
[17] J. A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki, Phys. Rev. E **72**, 026108 (2005).
[18] Sources for data are the following: data for Astro network have been collected from publicly avaible database at http://arxiv.org; English coocurrance dataset has been downloaded from V. Batagelj, 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*" [14] which is definied as follows: nodes are bus, tram or underground stops and an edge means that there is a direct route linking them.
[19] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[20] A. Fronczak and P. Fronczak, Phys. Rev. E **74**, 026121 (2006).
[21] M. T. Gastner and M. E. J. Newman, Eur. Phys. J. B **49**, 247 (2006).
[22] R. Ferrer i Cancho and R. Solé, *Statistical Physics of Complex Networks*, Lecture Notes in Physics (Springer, Berlin, 2003).