Networks with given two-point correlations: Hidden correlations from degree correlations

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

(Received 17 June 2006; published 31 August 2006)

This paper orders certain important issues related to both uncorrelated and correlated networks with hidden variables, in which hidden variables correspond to desired node degrees. In particular, we show that networks being uncorrelated at the hidden level are also lacking in correlations between node degrees. The observation supported by the depoissonization idea allows us to extract a distribution of hidden variables from a given node degree distribution. It completes the algorithm for generating uncorrelated networks that was suggested by other authors. In this paper we also carefully analyze the interplay between hidden attributes and node degrees. We show how to extract hidden correlations from degree correlations. Our derivations provide a mathematical background for the algorithm for generating correlated networks that was proposed by Boguñá and Pastor-Satorras.

DOI: 10.1103/PhysRevE.74.026121 PACS number(s): 89.75.Hc, 64.60.Ak

I. INTRODUCTION

Recently, the techniques of equilibrium and nonequilibrium statistical physics were developed to study complex networks [1–3]. This paper is devoted to equilibrium correlated networks, with a special attention given to networks with two-point correlations [4–8].

What does it mean that a network is correlated In simple words one can say that there exist certain relationships between network nodes. For example, when one considers a social network, i.e., a group of people with links given by acquaintance ties one may expect that young people are mostly surrounded by other young people. One may also expect that wealthy individuals are more often associated with other wealthy individuals than with poor ones. In some sense, the above examples let one suppose that social networks are positively correlated, at least when one considers an individual's age or income. The situation is more contentious if one asks for the relationship between genders of the nearest neighbors. Now, it is difficult to assess if social networks are positively or negatively correlated. The above examples show that even taking into account a single network one can observe correlations at different levels of the system complexity. Each node i in such a system has assigned a set of different attributes such as, gender g_i, age a_i, education e_i, *attractiveness* k_i, etc. The last property may be quantified as a number of nearest neighbors of the considered individual. In the graph theory [9] the quantity is known as node's degree. The above outlined network correlations and multilevel structure constitute two main issues undertaken in this paper.

Multilevel topology has been recently considered by several groups of researchers [10–14] and, at present, the proposed models are known as random networks with hidden variables. In general, random networks with hidden variables have a fixed number of vertices N. Each node in a network belonging to this class of models has assigned a hidden variable h_i (fitness, tag) randomly drawn from a fixed probability distribution $R(h_i)$ (throughout the paper we use the symbol R with reference to distributions at a hidden level and P with reference to node degrees). Edges are assigned to pairs of vertices $\{i,j\}$ with a given connection probability r_{ij}. In the simplest case r_{ij} depends only on values of the hidden variables h_i and h_j, but in a more general situation it may, for example, involve hidden variables characterizing the nearest or the next-nearest neighbors of both considered nodes i and j. In fact, the first case represents networks with at most two-point correlations at the level of hidden variables, while the latter one allows for higher-order correlations.

Below we outline the concept of network correlations in a more rigorous way. The introduced ideas will be completed and widely discussed in the next section.

From a mathematical point of view, the lack of two-point correlations means that the probability $R(h_j|h_i)$ that an edge departing from a vertex i of property h_i arrives at a vertex j of property h_j and is independent of the initial vertex i [15,16]. The above translates into the fact that the nearest neighborhood of each node is the same (in statistical terms). On the other hand, when $R(h_j|h_i)$ depends on both h_i and h_j, one says that the studied network has two-point correlations [5,7]. To characterize this type of correlation one usually takes advantage of the joint, two-dimensional distribution $R(h_i,h_j)$, which describes the probability of a randomly chosen edge to connect vertices labeled as h_i and h_j.

In order to characterize a network in a more detailed way the concept of higher-order correlations, given by multidimensional probability distributions, should be used. In this paper we limit ourselves to two-point correlations. The lack of higher-order correlations is ensured by the factorization of the conditional distribution $R(h_{j1},h_{j2},\ldots,h_{jn}|h_i) = R(h_{j1}|h_i)R(h_{j2}|h_i)\ldots R(h_{jn}|h_i)$, which describes the probability of a node of property h_i to have n neighbors labeled as $h_{j1},h_{j2},\ldots,h_{jn}$. Such networks, with the only two-point correlations are called *Markovian*, due to the reason that they are completely defined by the joint distribution $R(h_i,h_j)$, which on its own turn completely determines the conditional distribution $R(h_j|h_i)$ and the distribution of hidden variables $R(h_i)$. Relationships between $R(h_i,h_j)$, $R(h_j|h_i)$, r_{ij}, and $R(h_i)$ will be analyzed later.

After this short introduction to the concept of hidden variables and to the problem of network correlations we may go back to the main topic of the paper, i.e., the interplay between different levels of the network complexity. The approach has been initiated by Söderberg [12] and developed by Boguñá and Pastor-Satorras [7]. Boguñá and Pastor-Satorras have concentrated on the question: how correlations at the level of hidden variables affect pattern of connections at the level of node degrees. Given $R(h_i,h_j)$ the authors have derived an analytical expression for the joint distribution of degrees of the nearest neighbors $P(k_i,k_j)$. In this paper, we concentrate on the simplest case of the considered networks in which hidden attributes mimic nodes degrees (see Sec. V in Ref. [7]), and we ask the reverse question: what kind of hidden correlations $R(h_i,h_j)$ produce the given pattern of node degree correlations $P(k_i,k_j)$.

As a matter of fact, since most of us are better acquainted with node degree notation than with abstract hidden variables, the *reverse* approach seems to be very interesting, at least from the methodological point of view. It is already well known that there exist degree correlations in real networks [4]. On the other hand, due to the lack of data, nothing is known about correlation at the hidden level, from which the observed network structure arises. The paper represents a small step towards understanding the phenomenon of self-organization in complex networks beyond the predominant approach of the so-called evolving networks [17,18].

The paper is organized as follows. In Sec. II we review the general results on correlated random networks with hidden variables. The salient issues concerning two-point correlations, which have been originally presented in Refs. [7,8], are discussed in this section. Section III is devoted to theoretical aspects of our inverse problem. Some remarks on uncorrelated networks and a practical algorithm for generating random networks with two-point degree correlations are also given in this section. Finally, in Sec. IV we draw our conclusions.

II. CORRELATED NETWORKS WITH HIDDEN VARIABLES

A. Construction procedure

Probably the most attractive feature of random networks with hidden variables is the construction procedure that consists of only two steps.

(i) First, prepare N nodes and assign them hidden variables independently drawn from the probability distribution $R(h_i)$,

(ii) Next, each pair of nodes $\{i,j\}$ link with probability r_{ij}. Note that the above procedure does not include any ambiguous instructions like *in the case of…do…*. The lack of such ambiguities enables the analytical treatment of the discussed models.

The particularly simple class of sparse random networks with hidden variables has been recently considered by Boguñá and Pastor-Satorras [7]. Provided that the connection probability scales according to

$$r_{ij} = \frac{f(h_i,h_j)}{N}, \tag{1}$$

where $f(h_i,h_j)=f(h_j,h_i)$, the authors have shown that the degree distribution of nodes characterized by a fixed value of hidden variable h_i is given by the Poisson distribution

$$g(k_i|h_i) = \frac{e^{-\langle k(h_i)\rangle}\langle k(h_i)\rangle^{k_i}}{k_i!}, \tag{2}$$

where $\langle k(h_i)\rangle$ stands for the average degree of nodes with hidden attribute h_i, and in the considered case of attributes, which correspond to hidden (desired) node degrees

$$\langle k(h_i)\rangle = N\int r_{ij}R(h_j)dh_j = h_i. \tag{3}$$

The result expressed by Eq. (2) is very important because it joins the two levels of network complexity. Moreover, as we will show later, the above-mentioned result together with the formula (3) allows us to freely move between the notation of hidden variables and the notation of nodes degrees. In particular, the degree distribution $P(k_i)$ gains a simple form of Poissonian convolution

$$P(k_i) = \int \frac{e^{-h_i}h_i^{k_i}}{k_i!}R(h_i)dh_i, \tag{4}$$

that can be inverted providing us with the formula for the distribution of hidden variables $R(h_i)$ resulting in the desired $P(k_i)$. (Note that the last relation between both distributions $P(k_i)$ and $R(h_i)$ implies the following relation between their moments: $\langle h^n\rangle=\langle k(k-1)\cdots(k-n+1)\rangle$, and in particular, $\langle h\rangle=\langle k\rangle$, $\langle h^2\rangle=\langle k(k-1)\rangle$.) As a matter of fact, in the case of other pair-related distributions, like $R(h_i,h_j)$ and $P(k_i,k_j)$, it is also possible to obtain relations similar to Eq. (4), and to invert them. However, before we proceed with the inverse problems, we have to recall the most important issues related to two-point correlations.

B. Two-point correlations

One usually thinks about a network as about a collection of nodes and links. The distribution of hidden variables $R(h_i)$ is the basic characteristic of the set of nodes, whereas the joint distribution $R(h_i,h_j)$, describing two-point correlations, applies to the set of edges (see Fig. 1). Both distributions express the probability that a random representative of its own ensemble has assigned a given attribute, i.e., property h_i in the case of node, and a pair of hidden variables $\{h_i,h_j\}$ in the case of edge.

At the moment, let us briefly comment on Fig. 1. The second stage of the figure shows the set of nodes and the set of links, both corresponding to the simple network presented at the first stage of the same figure. Obviously, based only on the two sets it is almost impossible to recreate the original network. Such a representation related to a single network neglects much information. What is lacking are higher-order correlations. On the other hand, the joint distribution $R(h_i,h_j)$ characterizing the set of edges conveys all the information that is required to construct an *ensemble* of Markovian random networks (see the third stage of Fig. 1). In fact, all the calculations performed in this paper are related to such ensembles.

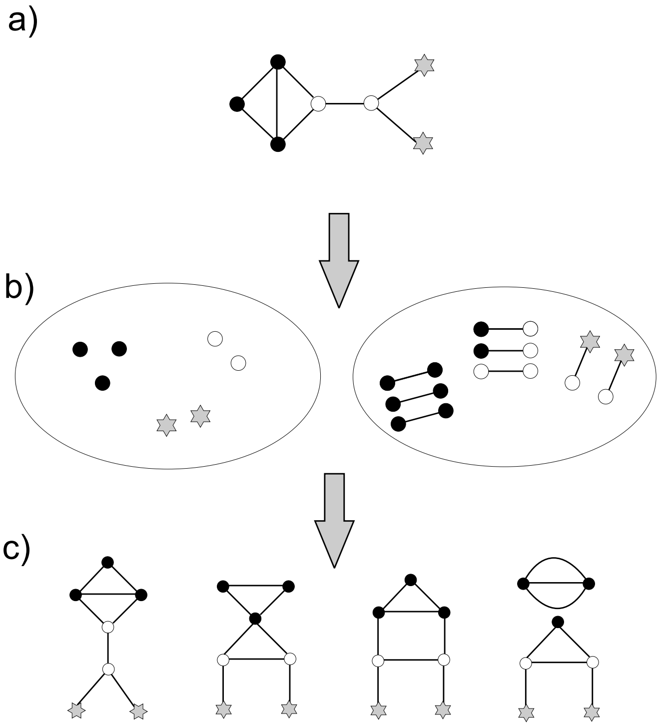

FIG. 1. Schematic representation of an ensemble of random networks with a given two-point correlation at the level of hidden variables. A detailed description of the figure is given in the text.

For a given network in the thermodynamic limit, both distributions $R(h_i)$ and $R(h_i,h_j)$ may be defined in the following way:

$$R(h_i) = \frac{N(h_i)}{N} \tag{5}$$

and

$$R(h_i,h_j) = \frac{E(h_i,h_j)}{2E} = \frac{E(h_i,h_j)}{\langle k\rangle N}, \tag{6}$$

where $N(h_i)$ gives the number of nodes assigned by h_i and $N=\int N(h_i)dh_i$, whereas $E(h_i,h_j)$ stands for the number of links labeled as $\{h_i,h_j\}$, and due to the normalization condition $E=\int\int E(h_i,h_j)dh_idh_j/2=\langle k\rangle N$.

The connection probability r_{ij} [Eq. (1)] may be simply expressed by the ratio between the actual number of edges connecting vertices of tags h_i and h_j, and the maximum value of this quantity

$$r_{ij} = \frac{E(h_i,h_j)}{E_{max}(h_i,h_j)}. \tag{7}$$

Since, during the network construction process one analyzes all pairs of nodes, the maximum number of the considered connections is given by $E_{max}=N(h_i)N(h_j)$. Now, taking advantage of Eq. (6), the expression for r_{ij} may be rewritten in the following form:

$$r_{ij} = \frac{\langle k\rangle R(h_i,h_j)}{NR(h_i)R(h_j)}. \tag{8}$$

To fulfill the discussion on two-point correlations it is necessary to define the relation between the two distributions (5) and (6). The relation is encoded in the so-called *hidden (desired) degree* detailed balance condition

$$R(h_j|h_i) = \frac{E(h_i,h_j)}{\langle k(h_i)\rangle N(h_i)} \overset{\text{def}}{=} \frac{\langle h\rangle R(h_i,h_j)}{h_iR(h_i)}, \tag{9}$$

which is, in fact, the key feature that justifies Eqs. (3) and (4). At the moment, note the analogy between the last identity and the *real degree* detailed balance condition [7] that holds in any Markovian network

$$P(k_j|k_i) = \frac{\langle k\rangle P(k_i,k_j)}{k_iP(k_i)}. \tag{10}$$

C. Interplay between hidden variables and node degrees

Each node of the considered networks is characterized by two parameters: hidden variable h_i and degree k_i. The probability $p(k_i\cap h_i)$ that two given parameters h_i and k_i meet together in a certain node is described by the identity

$$p(k_i \cap h_i) = R(h_i)g(k_i|h_i) = P(k_i)g^*(h_i|k_i). \tag{11}$$

The meaning of both conditional distributions $g(k_i|h_i)$ and $g^*(h_i|k_i)$ is simple. The first distribution $g(k_i|h_i)$ [Eq. (2)] has just been introduced. It describes the probability that a node labeled by h_i has k_i nearest neighbors. The second distribution $g^*(h_i|k_i)$ is complementary to the first one, and gives the probability that a node with k_i nearest neighbors is labeled by h_i.

The knowledge of $g(k_i|h_i)$ and $g^*(h_i|k_i)$ allows one to find the relation between the joint distributions $R(h_i,h_j)$ and $P(k_i,k_j)$ characterizing pair correlations, respectively, at the level of hidden variables and at the level of node degrees. Simple arguments let one describe the interplay by the following relation [19]:

$$P(k_i,k_j) = \int\int g(k_i-1|h_i)R(h_i,h_j)g(k_j-1|h_j)dh_idh_j. \tag{12}$$

The last expression states that if one knows hidden correlations then it is possible to calculate degree correlations. Relating the problem to the present *state of the art* in the field of complex networks it is a very artificial situation. The concept of networks with hidden variables is still little known because most of the researchers are strongly attached to the node degree notation. From this point of view, the reverse problem that would answer the question: What kind of hidden correlations produce given degree correlations, seems to be very attractive. In addition, since one knows how to construct Markovian networks with hidden variables, the solution of the above problem would also allow us, in a simple way, to generate networks with given two-point degree correlations. We work out the problem in Sec. III.

In fact, the idea of generating networks with given degree correlations by means of networks with hidden variables originates from Boguñá and Pastor-Satorras (see Sec. V in

Ref. [7]). The authors have argued that since the conditional probability $g(h_i|k_i)$ is Poissonian [Eq. (2)], the joint distribution for node degrees [Eq. (12)] approaches the joint distribution for hidden variables for $k_i,k_j \gg 1$,

$$P(k_i,k_j) \simeq R(k_i,k_j), \tag{13}$$

and, respectively, asymptotic behavior of the degree distribution is given by

$$P(k_i) \simeq R(k_i). \tag{14}$$

The range of convergence of the two distributions has been estimated by Dorogovtsev [8], who has shown that both approximations (13) and (14) are only acceptable when $R(h_i,h_j)$ and $R(h_i)$ are sufficiently slowly decreasing.

III. HIDDEN CORRELATIONS FROM DEGREE CORRELATIONS

A. Depoissonization

At the moment, let us examine Eq. (4) more carefully. Depending on whether h_i is a discrete or continuous variable, the expression for $P(k_i)$ is, respectively, a discrete or integral transform with a Poissonian kernel [20]. The issue of determining $R(h_i)$ from $P(k_i)$ is simply the problem of finding the inverse transform [21], and one can show (see Appendix A) that for some $P(k_i)$ there indeed exists a unique $R(h_i)$, which satisfies Eq. (4). When h_i is continuous one gets

$$R(h_i) = e^{h_i}\mathcal{F}^{-1}[G(ix)], \tag{15}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform and G stands for the generating function for the degree distribution $P(k_i)$,

$$G(ix) = \sum_{k_i} (ix)^{k_i} P(k_i). \tag{16}$$

In the case of discrete h_i, the inverse Poisson transform is given by the formula

$$R(h_i) = e^{h_i}\mathcal{Z}^{-1}[G(-\ln x)], \tag{17}$$

where $\mathcal{Z}^{-1}$ describes the inverse Z transform.

The same applies for the joint degree distributions characterizing two-point correlations. Since

$$g(k_i - 1|h_i) = \frac{k_i}{h_i} g(k_i|h_i), \tag{18}$$

the formula (12) may be rewritten as

$$\frac{P(k_i,k_j)}{k_ik_j} = \int\int g(k_j|h_j)\frac{R(h_i,h_j)}{h_ih_j} g(k_i|h_i)dh_idh_j. \tag{19}$$

Now, it is easy to see that $P(k_i,k_j)$ is connected with $R(h_i,h_j)$ by the two-dimensional Poisson transform and

$$\frac{R(h_i,h_j)}{h_ih_j} = e^{h_i+h_j}\mathcal{F}^{-1}\{\mathcal{F}^{-1}[G^*(ix,iy)]\}, \tag{20}$$

where

$$G^*(ix,iy) = \sum_{k_i}\sum_{k_j} (ix)^{k_i}(iy)^{k_j}\frac{P(k_i,k_j)}{k_ik_j}. \tag{21}$$

The last two expressions can be simply translated for the case of discrete hidden variables. Since, however, Fourier transforms are more convenient to work with than Z transforms, we have decided to pass over such a reformulation.

B. Some remarks on uncorrelated networks

As we said at the beginning of this paper, the lack of pair correlations at the level of hidden variables means that the conditional probability $R_0(h_j|h_i)$ does not depend on h_i (to differentiate between uncorrelated and correlated networks the characteristics related to the former case have been denoted by the subscript "0"). In fact, it is simple to show that

$$R_0(h_j|h_i) = \frac{h_j}{\langle h\rangle}R_0(h_j), \tag{22}$$

and, respectively, the joint distribution [Eq. (9)] gains a factorized form

$$R_0(h_i,h_j) = \frac{h_iR_0(h_i)h_jR_0(h_j)}{\langle h\rangle^2}. \tag{23}$$

Inserting the above expression into Eq. (7) one gets the formula for the connection probability in uncorrelated networks

$$r_{ij}^0(h_i,h_j) = \frac{h_ih_j}{\langle h\rangle N}. \tag{24}$$

Now, the question is, does the lack of pair correlations at the hidden level translate to the lack of pair correlations between degrees of the nearest neighbors? Inserting Eq. (23) into Eq. (12) and then taking advantage of the degree detailed balance condition (10), one gets the answer

$$P_0(k_j|k_i) = \frac{k_j}{\langle k\rangle}P_0(k_j), \tag{25}$$

i.e., the lack of hidden correlations results in the lack of node degree correlations. Thus, in order to generate an uncorrelated network with a given degree distribution $P_0(k_i)$ one has to (i) prepare the desired number of nodes N, (ii) label each node by a hidden variable randomly taken from the distribution $R_0(h_i)$ given by Eq. (15) or Eq. (17). (iii) Each pair of nodes links with the probability r_{ij}^0 [Eq. (24)].

(See also other methods for generating random uncorrelated networks with a given degree distribution [15,22]).

Note also that due to Eq. (3), the connection probability $r_{ij}^0(h_i,h_j)$ [Eq. (24)] may be calculated as the density of connections $p_{ij}^0(k_i,k_j)$ among the degree classes k_i and k_j averaged over all pairs of nodes possessing hidden variables, respectively, equal to h_i and h_j,

$$r_{ij}^0(h_i,h_j) = \sum_{k_i}\sum_{k_j} g(k_i|h_i)p_{ij}^0(k_i,k_j)g(k_j|h_j), \tag{26}$$

where

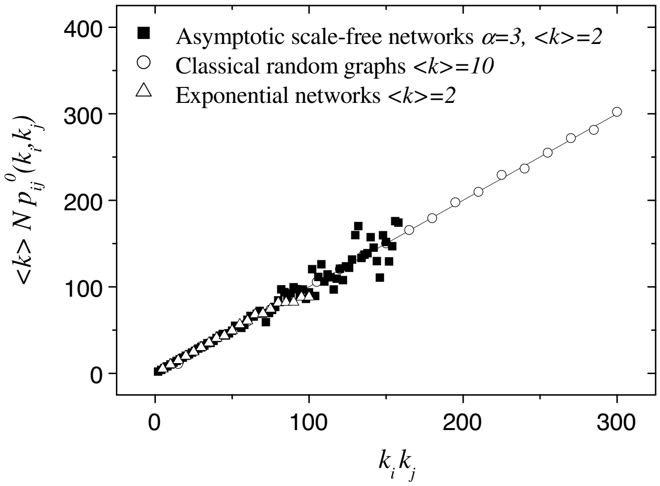

FIG. 2. Density of connections $p_{ij}^0(k_i,k_j)$ among the degree classes k_i and k_j in different uncorrelated networks with hidden variables. The points stand for the results of numerical simulations averaged over 500 networks of size $N=5000$, whereas the solid line corresponds to Eq. (27).

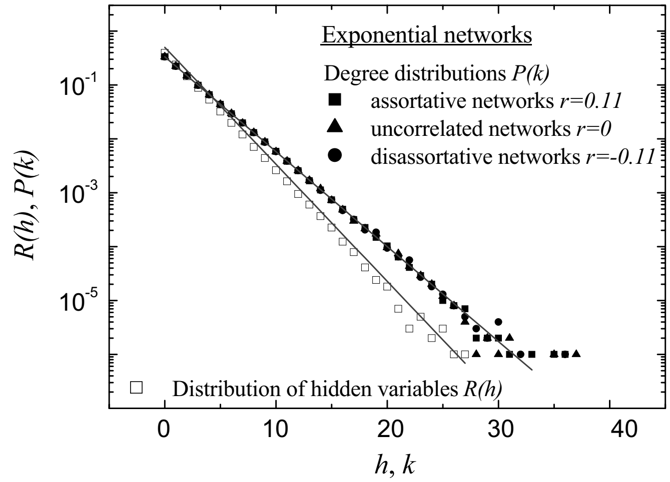

FIG. 3. Distribution of hidden variables, and degree distributions obtained from numerical simulations of uncorrelated (Sec. III C) and correlated (Sec. III E) exponential networks.

$$p_{ij}^0(k_i,k_j) = \frac{k_i k_j}{\langle k\rangle N}. \tag{27}$$

The last expression describing the density of connections between nodes of degrees k_i and k_j in uncorrelated sparse networks has already been used by several authors (in particular, see Refs. [23,25]).

C. Examples of uncorrelated networks

1. Classical random graphs of Erdös and Rènyi

The degree distribution in the Erdös-Rènyi (ER) model is Poissonian,

$$P_0(k_i) = \frac{e^{-\langle k\rangle}\langle k\rangle^{k_i}}{k_i!}, \quad k_i \geqslant 0. \tag{28}$$

The first step towards the calculation of the required distribution $R_0(h_i)$ is finding a characteristic function for Poisson distribution. Taking advantage of Eq. (16) one gets

$$G(ix) = e^{\langle k\rangle(ix-1)}. \tag{29}$$

Now, inserting the exponential function into Eq. (15) or Eq. (17), one can see that the distribution of hidden variables in classical random graphs is, respectively, given by the Dirac's delta function (in the case of continuous h_i) or the Kronecker delta (in the case of discrete h_i)

$$R_0(h_i) = \delta(h_i - \langle k\rangle). \tag{30}$$

The above result means that all vertices are equivalent, and the connection probability at the level of hidden variables [Eq. (24)] is given by

$$r_{ij}^0 = \frac{\langle k\rangle}{N}. \tag{31}$$

Figure 2 presents the density of connections $p_{ij}^0(k_i,k_j)$ between nodes characterized by degrees k_i and k_j in the ER model, and other uncorrelated networks. As one can see, there exists a very good agreement between the formula (27) and the results of numerical simulations.

2. Networks with exponential degree distribution

Now, let us suppose that

$$P_0(k) = \frac{\langle k\rangle^{k_i}}{(1+\langle k\rangle)^{k_i+1}}, \quad k_i \geqslant 0. \tag{32}$$

Generating a function for the above degree distribution is given by an infinite geometric series, which converges to

$$G(ix) = \frac{1}{1+\langle k\rangle - i\langle k\rangle x} \tag{33}$$

for $x < x_c = (\langle k\rangle+1)/\langle k\rangle$, and diverges for $x \geqslant x_c$. Now, taking into account the discussion given in Appendix A, one can show that Eq. (15) provides a reasonable distribution of hidden variables

$$R_0(h_i) = \frac{e^{-h_i/\langle k\rangle}}{\langle k\rangle}, \quad h_i \geqslant 0, \tag{34}$$

which produces such a degree distribution that is very similar to the desired one [Eq. (32)] (see Fig. 3).

3. Scale-free networks

In mathematical terms, the scale-free property translates into a power-law degree distribution

$$P_0(k_i) = \frac{A}{k_i^{\alpha}}, \quad k_i \geqslant 1, \tag{35}$$

where α is a characteristic exponent and A represents a normalization constant. Generating a function for this distribution is given by the polylogarithm

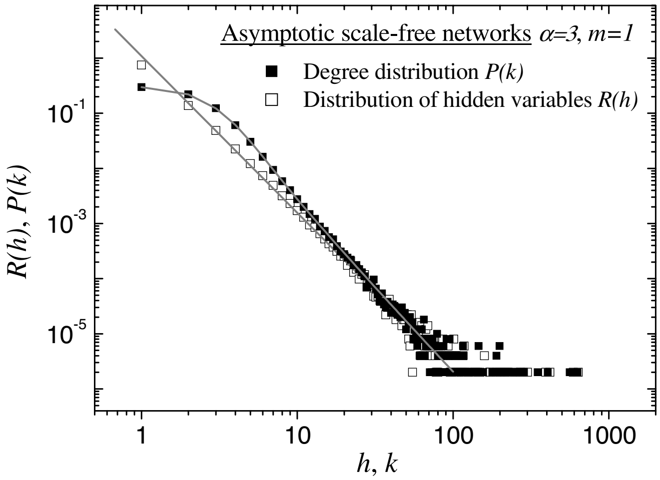

FIG. 4. Degree distribution in uncorrelated networks with scale-free distributions of hidden variables [Eq. (38)]. The points represent the results of numerical simulations whereas the solid lines correspond to Eq. (39).

$$G(ix) = A\sum_{k_i=1}^{\infty}\frac{(ix)^{k_i}}{k_i^{\alpha}} = A\,\mathrm{Li}_{\alpha}(ix). \tag{36}$$

To derive the distribution of hidden variables that lead to an uncorrelated scale-free network, one has to find the inverse Fourier transform of the polylogarithm with the imaginary argument

$$R_0(h_i) = Ae^{h_i}\mathcal{F}^{-1}[\mathrm{Li}_{\alpha}(ix)], \tag{37}$$

or the adequate inverse Z transform.

Unfortunately, it does not appear that closed-form solutions for both inverse transforms can be simply obtained. Nevertheless, some asymptotic results for scale-free networks can be derived. In particular, one can show that the power-law distribution of hidden variables

$$R_0(h_i) = \frac{(\alpha-1)m^{(\alpha-1)}}{h_i^{\alpha}}, \quad h_i \geqslant m, \tag{38}$$

leads to asymptotic scale-free networks with degree distribution given by

$$P(k_i) = (\alpha-1)m^{(\alpha-1)}\frac{\Gamma(k_i-\alpha+1,m)}{k_i!}, \tag{39}$$

where $\Gamma(x,y)$ stands for incomplete gamma function. In the limit of large degrees $k_i \gg 1$ the above degree distribution decays as $P(k_i) \sim k_i^{-\alpha}$ (see Fig. 4).

The effect of structural cutoffs in power-law distributions of hidden variables with $\alpha < 3$ [Eq. (38)], imposing the largest hidden variable to scale as $h_{max} \sim \sqrt{N}$ (the relation follows from $r_{ij}^0 \leqslant 1$ [Eq. (24)]), does not represent any problem in the studied formalism. The effect of h_{max} in the scale-free $R(h_i)$ may be considered as an *exponential* cutoff

$$R(h_i) = \frac{(\alpha-1)m^{(\alpha-1)}}{h_i^{\alpha}}\exp\left(-\frac{h_i}{h_{max}}\right). \tag{40}$$

Due to the properties of the Poisson transform [20], the above $R(h_i)$ results in a truncated power-law degree distribution [Eq. (39)]

$$P(k_i) = \frac{\Gamma(k_i-\alpha+1,m)}{k_i!}\frac{(\alpha-1)m^{(\alpha-1)}}{(1/h_{max}+1)^{k_i-\alpha+1}}. \tag{41}$$

As expected [see the discussion of Eq. (14)], in the limit of large degrees the last formula approaches Eq. (40).

Before we finish with scale-free networks let us, once again, concentrate on Fig. 4. The figure presents degree distribution $P(k_i)$ in sparse networks with scale-free distributions of hidden attributes $R(h_i)$ [Eq. (38)]. One can observe that although both distributions converge in the limit of large degrees, there exist serious deviations between Eqs. (38) and (39) in the limit of small degrees. The relative behavior of the two distributions lets one expect that the correct $R(h_i)$ reproducing the *pure* scale-free degree distribution $P(k_i)$ should describe a kind of *condensate* with a huge number of nodes characterized by very low values of hidden attributes. On the other hand, despite the *ambiguous* behavior of $P(k_i)$ for small degrees, the power-law tail is interesting in its own right. A number of real networks have fat-tailed degree distributions. The above allows us to deduce on fat-tailed distributions of underlying hidden attributes assigned to individuals cocreating real networks.

D. How to generate correlated networks with a given degree correlation

Here, we again make use of the dePoissonization idea proposed in Sec. III A. The procedure of generating random networks with a given two-point, degree correlation $P(k_i,k_j)$ is as follows:

(i) First, prepare N nodes;

(ii) next, label each node by a hidden variable randomly taken from the distribution $R(h_i)$ [Eq. (15)] and;

(iii) finally, link each pair of nodes with the probability r_{ij} [Eq. (7)], where $R(h_i,h_j)$ is calculated according to the formula (20).

Although very clear, the above procedure suffers a certain inconvenience: given a joint degree distribution $P(k_i,k_j)$, the closed-form solution for $R(h_i,h_j)$ [Eq. (20)] is often hard to get. Since, however, there exists a number of algorithms for the numerical inversion of Fourier transform, the above does not represent a real problem.

E. Examples of correlated networks

To make our previous derivations more concrete, we should immediately introduce some examples of correlated networks. In order to simplify the task we will take advantage of general patterns for joint degree distributions with two-point assortative (a) and disassortative (d) correlations that were proposed by Newman [5],

$$P^d(k_i,k_j) = \tilde{P}(k_i)\tilde{Q}(k_j) + \tilde{Q}(k_i)\tilde{P}(k_j) - \tilde{Q}(k_i)\tilde{Q}(k_j) \tag{42}$$

and

$$P^a(k_i,k_j) = 2\tilde{P}(k_i)\tilde{P}(k_j) - P^d(k_i,k_j), \tag{43}$$

where $\tilde{P}(k_i)$ and $\tilde{Q}(k_i)$ are arbitrary distributions such that $\Sigma_{k_i}\tilde{P}(k_i)=\Sigma_{k_i}\tilde{Q}(k_i)=1$. (Please note that the above distributions (42) and (43) may become negative in some situations, and consequently, some restrictions on the parameter space are in order.) Assortativity or correlation coefficients [4] for the above distributions are, respectively, equal to

$$r^d = -\frac{(\mu_p - \mu_q)^2}{\sigma_p^2} \quad \text{and} \quad r^a = -r^d, \tag{44}$$

where μ_p represents the expectation value for $\tilde{P}(k_i)$ and μ_q has the same meaning for $\tilde{Q}(k_i)$, whereas σ_p^2 corresponds to the variance of $\tilde{P}(k_i)$. Now, in order to facilitate further calculations let us assume that

$$\tilde{P}(k_i) = \frac{k_i}{\langle k\rangle}P(k_i) \quad \text{and} \quad \tilde{Q}(k_i) = \frac{k_i}{\langle q\rangle}Q(k_i), \tag{45}$$

where

$$\langle k\rangle = \sum_{k_i} k_i P(k_i) \quad \text{and} \quad \langle q\rangle = \sum_{k_i} k_i Q(k_i). \tag{46}$$

Note that there is no conflict of notation in the last assignment. Putting the two expressions [Eq. (45)] into P^a or P^d, and then taking advantage of the degree detailed balance condition (10), one can easily check that $P(k_i)$ corresponds to degree distribution. Now, given $P(k_i)$ one can execute the first two steps [(i). and (ii)] of the construction procedure described in the preceding subsection.

Inserting the relations (45) into Eqs. (42) and (43) one obtains [cf. Eq. (19)]

$$\frac{P^d(k_i,k_j)}{k_ik_j} = \frac{P(k_i)Q(k_j)}{\langle k\rangle\langle q\rangle} + \frac{Q(k_i)P(k_j)}{\langle k\rangle\langle q\rangle} - \frac{Q(k_i)Q(k_j)}{\langle q\rangle^2} \tag{47}$$

and

$$\frac{P^a(k_i,k_j)}{k_ik_j} = 2\frac{P(k_i)P(k_j)}{\langle k\rangle^2} - \frac{P^d(k_i,k_j)}{k_ik_j}. \tag{48}$$

Due to the linearity of the Poisson transform the joint distributions [Eqs. (47) and (48)] turn out to be particularly useful for our purposes. The mentioned usefulness means that whenever closed-form solutions for the inverse Poisson transforms of $P(k_i)$ and $Q(k_i)$ exist, one can also obtain closed-form solutions for the joint hidden distributions $R^d(h_i,h_j)$ and $R^a(h_i,h_j)$ [Eq. (20)].

Generating functions $G^*(ix,iy)$ [Eq. (21)] for Eqs. (47) and (48) are, respectively given by

$$G_d^*(ix,iy) = \frac{G_p(ix)G_q(iy)}{\langle k\rangle\langle q\rangle} + \frac{G_q(ix)G_p(iy)}{\langle k\rangle\langle q\rangle} - \frac{G_q(ix)G_q(iy)}{\langle q\rangle^2} \tag{49}$$

and

$$G_a^*(ix,iy) = 2\frac{G_p(ix)G_p(iy)}{\langle k\rangle^2} - G_d^*(ix,iy), \tag{50}$$

where

$$G_p(ix) = \sum_{k_i}(ix)^{k_i}P(k_i) \quad \text{and} \quad G_q(ix) = \sum_{k_i}(ix)^{k_i}Q(k_i). \tag{51}$$

Although visually quite complicated, all the above formulas are in fact very simple. Now, in order to perform the last step [(iii)] of our procedure, aiming at constructing correlated networks, one has to calculate the joint distribution $R(h_i,h_j)$. Taking advantage of Eq. (20) one gets

$$R^d(h_i,h_j) = \tilde{R}(h_i)\tilde{S}(h_j) + \tilde{S}(h_i)\tilde{R}(h_j) - \tilde{S}(h_i)\tilde{S}(h_j), \tag{52}$$

and

$$R^a(h_i,h_j) = 2\tilde{R}(h_i)\tilde{R}(h_j) - R^d(h_i,h_j), \tag{53}$$

where, similarly to Eq. (45), one has

$$\tilde{R}(h_i) = \frac{h_i}{\langle k\rangle}R(h_i) \quad \text{and} \quad R(h_i) = e^{h_i}\mathcal{F}^{-1}[G_p(ix)], \tag{54}$$

and also

$$\tilde{S}(h_i) = \frac{h_i}{\langle q\rangle}S(h_i) \quad \text{and} \quad S(h_i) = e^{h_i}\mathcal{F}^{-1}[G_q(ix)]. \tag{55}$$

Note that $R(h_i)$, given in Eq. (54), expresses the distribution of hidden variables in the considered correlated networks [i.e., the inverse Poisson transform of the degree distribution $P(k_i)$].

Now, let us translate the general considerations into a specific example. Suppose that we are interested in networks with an exponential degree distribution [Eq. (32)]

$$P(k_i) = \frac{\langle k\rangle^{k_i}}{(1+\langle k\rangle)^{k_i+1}}, \quad k_i \geq 0. \tag{56}$$

The distribution of hidden variables in such networks is given by [Eq. (34)]

$$R(h_i) = \frac{e^{-h_i/\langle k\rangle}}{\langle k\rangle}, \quad h_i \geq 0. \tag{57}$$

For mathematical simplicity, let us assume that the distribution $Q(k_i)$ responsible for correlations is also exponential;

$$Q(k_i) = \frac{\langle q\rangle^{k_i}}{(1+\langle q\rangle)^{k_i+1}}, \quad k_i \geq 0. \tag{58}$$

Given $P(k_i)$ and $Q(k_i)$ one has to ensure that the joint hidden distributions (52) and (53), and also the connection probability r_{ij} [Eq. (7)] are positive and smaller than 1. It is easy to check that in our case the condition translates into the relation

$$\frac{1}{\sqrt{2}} \leq \frac{\langle q\rangle}{\langle k\rangle} \leq 1. \tag{59}$$

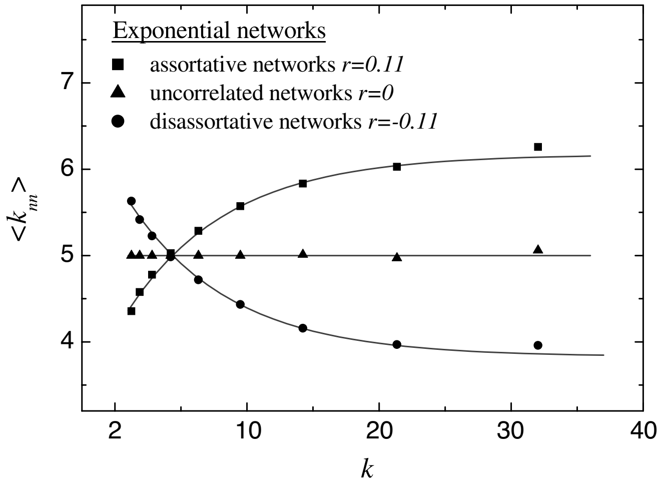

FIG. 5. Average nearest-neighbor degree $\langle k_{nn}\rangle(k)$ for assortative, uncorrelated, and disassortative networks with exponential degree distribution (Sec. III E). Numerical simulations were done for networks of size $N=10^6$, $\langle k\rangle=2$, $\langle q\rangle=\sqrt{2}$, and averaged over 25 network realizations. Note that given $\langle k\rangle=2$, the value of $\langle q\rangle=\sqrt{2}$ corresponds to the maximum value of the correlation coefficient $|r_{max}|=0.11$ [Eq. (61)].

Now, let us briefly examine the role of $Q(k_i)$. Depending on the value of $\langle q\rangle$ one obtains stronger or weaker correlations characterized by [Eq. (44)]

$$r^d = -\frac{2(\langle k\rangle - \langle q\rangle)^2}{\langle k\rangle(\langle k\rangle + 1)} \quad \text{and} \quad r^a = -r^d. \tag{60}$$

In particular, given $\langle k\rangle$, the expression (59) provides possible values for the correlation coefficients that may be reproduced in the considered networks

$$0 \leqslant r \leqslant \frac{\langle k\rangle}{\langle k\rangle + 1}(\sqrt{2}-1)^2, \tag{61}$$

where $r=|r^d|=|r^a|$.

In order to check the validity of our derivations we have performed numerical simulations of correlated networks (42) and (43) with partial distributions given by [Eqs. (56) and (58)]. In Fig. 3 we depict the results corresponding to degree distributions [Eq. (56)] in the considered networks. Figure 5 presents the effects of node degree correlations expressed by the average degree of the nearest neighbor (ANND), which is defined as

$$\langle k_{nn}\rangle(k_i) = \sum_{k_j} k_j P(k_j/k_i). \tag{62}$$

It is already well known that in the case of uncorrelated networks [Eq. (25)] the ANND does not depend on k_i,

$$\langle k_{nn}^0\rangle(k_i) = \frac{\langle k^2\rangle}{\langle k\rangle}, \tag{63}$$

whereas in the case of assortatively (disassortatively) correlated systems it is an increasing (decreasing) function of k_i. One can find that in our example the corresponding functions are given by the below formulas,

$$\langle k_{nn}^0\rangle(k_i) = 1 + 2\langle k\rangle, \tag{64}$$

$$\langle k_{nn}^d\rangle(k_i) = \langle k_{nn}^0\rangle(k_i) - f(k_i), \tag{65}$$

$$\langle k_{nn}^a\rangle(k_i) = \langle k_{nn}^0\rangle(k_i) + f(k_i), \tag{66}$$

where

$$f(k_i) = 2(\langle k\rangle - \langle q\rangle)\left[1 - \left(\frac{\langle q\rangle}{\langle k\rangle}\right)^{k_i-1}\left(\frac{1+\langle k\rangle}{1+\langle q\rangle}\right)^{k_i+1}\right]. \tag{67}$$

As one can see (Fig. 5), the fit between computer simulations and the above analytical expressions is very good, certifying the validity of the proposed algorithm (Sec. III D) for generating random networks with a given degree of correlations.

IV. CONCLUSIONS

In this paper we refer to the set of articles devoted to the so-called random networks with hidden variables. The importance of this paper consists in ordering certain significant issues related to both uncorrelated and correlated networks in which hidden variables mimic desired node degrees. In particular, we show that networks being uncorrelated at the hidden level are also lacking in correlations between node degrees. The observation supported by the depoissonization idea (Sec. III A) allows us to extract the distribution of hidden variables from a given node degree distribution. Until now the distribution of hidden variables required for the generation of a given degree sequence had to be guessed. From this point of view our findings complete the algorithm for generating random uncorrelated networks that was suggested by other authors [12,24]. We also show that the density of connections among degree classes k_i and k_j in sparse uncorrelated networks is a factorized function of node degrees.

In this paper we also carefully analyze the interplay between hidden attributes and node degrees. We show how to extract hidden correlations from degree correlations, and how to freely move between the two levels of the networks complexity. Our derivations provide a mathematical background for the algorithm for generating correlated networks that was proposed by Boguñá and Pastor-Satorras [7].

ACKNOWLEDGMENTS

The authors thank Professor Janusz A. Hołyst and Dr. Mariusz Zając for critical comments and valuable suggestions. A.F. acknowledges financial support from the Foundation for Polish Science (FNP 2005) and the State Committee for Scientific Research in Poland (Grant No. 1P03B04727). The work of P.F. has been supported by European Commission Project No. CREEN FP6-2003-NEST-Path-012864.

APPENDIX A

The inverse Poisson transform for the case of continuous h [Eq. (15)] has been derived by Wolf and Mehta in 1964 [21]. Below we outline the derivation and we show that it is completely correct in only a very few cases of $P(k)$. It means that although having $R(h)$ one can always find the corresponding $P(k)$ [Eq. (4)], the reverse statement is not true, i.e., there exist such degree distributions, which do not possess a corresponding distribution of hidden variables. We also shortly explain why, even though the formula (15) is not completely correct for a large class of degree distributions $P(k)$, it often provides reasonable distributions of hidden variables $R(h)$. Finally, we adopt the derivation for the case of discrete h [Eq. (17)].

The aim is to inverse the formula

$$P(k) = \int_0^\infty \frac{e^{-h}h^k}{k!}R(h)dh. \tag{A1}$$

In order to do so, one has to consider the below identity,

$$F(x) = \int_0^\infty e^{ixh}R(h)e^{-h}dh \tag{A2}$$

$$= \int_0^\infty \sum_{k=0}^\infty \frac{(ixh)^k}{k!}R(h)e^{-h}dh. \tag{A3}$$

In their derivation Wolf and Mehta [21] assumed that the formula may be rewritten in the following form:

$$F(x) = \sum_{k=0}^\infty (ix)^k \int_0^\infty \frac{h^k}{k!}R(h)e^{-h}dh \tag{A4}$$

$$= \sum_{k=0}^\infty (ix)^k P(k) = G(ix), \tag{A5}$$

where G [Eq. (16)] represents the generating function for the degree distribution $P(k)$. Then, by the Fourier inversion formula they got the expression [Eq. (15)]

$$R(h) = e^h \mathcal{F}^{-1}[G(ix)]. \tag{A6}$$

The above derivation, however, suffers serious analytical inaccuracy. Due to Lebesgue's theorem (see Appendix B) the second formula for $F(x)$ [Eq. (A4)] is only true for $x \leqslant x_c$, where $x_c \geqslant 1$ depends on $R(h)$. On the other hand, since x stands for frequency, and we know that lower frequencies are more important for the reconstruction of a given signal [here $R(h)$], in some cases it is reasonable to forget about mathematical accuracy and use the inverse formula (73) for the whole range of $x \in (0,\infty)$, but still having in mind that the mentioned inaccuracy may result in unacceptable results [e.g., negative values of $R(h)$].

The above reasoning can be simply adopted for the case of discrete transform with a Poissonian kernel, i.e.,

$$P(k) = \sum_{h=0}^\infty \frac{e^{-h}h^k}{k!}R(h). \tag{A7}$$

Let us introduce an auxiliary sequence

$$J(h) = e^{-h}R(h). \tag{A8}$$

It is easy to show that for a certain range of $s \leqslant s_c$ the Z transform of this sequence is equal to a generating function of the degree distribution [Eq. (16)]

$$\mathcal{Z}[J(h)] = \sum_{h=0}^\infty \frac{J(h)}{s^h} \tag{A9}$$

$$= \sum_{h=0}^\infty \left(\sum_{l=0}^\infty \frac{(-h \ln s)^l}{l!} \right) J(h) \tag{A10}$$

$$\overset{s \leqslant s_c}{=} \sum_{l=0}^\infty (\ln s^{-1})^l \sum_{h=0}^\infty \frac{h^l}{l!} J(h) \tag{A11}$$

$$= \sum_{l=0}^\infty (\ln s^{-1})^l P(l) = G(\ln s^{-1}). \tag{A12}$$

Applying the inverse Z transform to the last expression one obtains the formula (17) describing the distribution of hidden variables

$$R(h) = e^h \mathcal{Z}^{-1}[G(\ln s^{-1})]. \tag{A13}$$

APPENDIX B: LEBESGUE'S THEOREM FOR A SERIES STATE [26]

Suppose that $I \subseteq \mathcal{R}$ is an interval. Suppose further that the sequence of functions $g_n \in \mathcal{L}(I)$ (i.e., g_n is a Lebesgue integrable function on I) satisfies the following conditions:

(i) $\Sigma_{n=0}^\infty g_n$ converges almost everywhere on I to a sum function $g: I \to \mathcal{R}$.

(ii) There exists a non-negative function $G \in \mathcal{L}(I)$ such that for every $N \in \mathcal{N}$, $|\Sigma_{n=0}^N g_n(x)| \leqslant G(x)$ for almost all $x \in I$.

Then $g \in \mathcal{L}(I)$, the series

$$\sum_{n=0}^N \int_I g_n(x)dx \tag{B1}$$

converges, and

$$\int_I g(x)dx = \int_I \sum_{n=0}^N g_n(x)dx = \sum_{n=0}^N \int_I g_n(x)dx. \tag{B2}$$

[1] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks* (Wiley, New York, 2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, New York, 2003).
[3] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[5] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[6] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, e-print cond-mat/0206131.
[7] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112

(2003).
[8] S. N. Dorogovtsev, eprint cond-mat/0308336.
[9] R. J. Wilson, *Introduction to Graph Theory* (Longman, Essex, 1985).
[10] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[11] G. Caldarelli, A. Capocci, P. DeLosRios, and M. A. Munoz, Phys. Rev. Lett. **89**, 258702 (2002).
[12] B. Söderberg, Phys. Rev. E **66**, 066121 (2002).
[13] B. Söderberg, Phys. Rev. E **68**, 015102(R) (2003).
[14] B. Söderberg, Phys. Rev. E **68**, 026107 (2003).
[15] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[16] A. Fronczak, P. Fronczak, and J. A. Hołyst, eprint cond-mat/0502663.
[17] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[18] A. L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[19] The formula (12) directly follows from Eq. (25) given in Ref. [7].
[20] H. J. Hindin, in *Theory and Applications of the Poisson Transform*, Conference record of the 2nd Asilomar Conference on Circuits and Systems (Institute of Electrical and Electronics Engineers Inc., New York, 1968), pp. 525–529.
[21] E. Wolf and C. L. Mehta, Phys. Rev. Lett. **13**, 705 (1964).
[22] Z. Burda, J. D. Correia, and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
[23] A. Fronczak, P. Fronczak, and J. A. Hołyst, *How to Calculate to Main Characteristics of Random Uncorrelated Networks*, AIP Conf. Proc. No. 776 (AIP, New York, 2005), p. 52.
[24] F. Chung and L. Lu, Ann. Comb. **6**, 125 (2002).
[25] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[26] W. Kołodziej, *Mathematical Analysis* (PWN, Warsaw, 1978) (in Polish).