Vol. **114** (2008) *ACTA PHYSICA POLONICA A* No. 3

Proceedings of the 3rd Polish Symposium on Econo- and Sociophysics, Wrocław 2007

How Nodes and Groups Properties Influence Assortativity in Social Networks?

P. Hetman, P. Magnuszewski, J. Stefańska, L. Bujkiewicz and K. Ostasiewicz

Institute of Physics, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50–370 Wrocław, Poland

A model of social network construction taking into account both social and individual influences on the distribution of links is proposed. The balance between social and individual factors is regulated through a "flexibility" parameter, reflecting how strong the initial individual sociability is altered by groups structure. The main interest is focused on the effect of groups on degree–degree correlation. Both numerical and analytical results on the relationship between assortativity and flexibility are presented.

PACS numbers: 89.75.Hc, 87.23.Ge, 02.50.Cw

1. Introduction

Networks are the backbone of social structure. The long tradition of social network research in sociology has been recently supported by statistical physics community [1–4]. New models and measures of complex networks have been applied successfully to social networks [5–9]. The question has been raised about differences between social networks and other types of networks — are they different only because of the properties of their nodes, or maybe there is a difference that can be detected in a network structure? Newman and Park [10], basing on empirical evidence, posited that social networks are characterized by positive assortativity [11, 12], while for other network types assortativity is zero or negative. Whitney and Alderson [13, 14] showed with expanded empirical evidence that data do not fully support this hypothesis. Further progress in this debate was reached through specific models (mechanisms) showing how positive assortativity is generated in social networks [10, 11, 15–19].

In this article we focus on the effect of groups on social network structure which as has been shown by Newman and Park [10] can lead to positive assortativity. However, the model proposed in [10] reflects only one psychological mechanism of relationships' formation. In their model, the group structure fully controls individuals' properties in such a way that every person is connected to a given fraction p of other common group members, and is not connected to anyone outside. Such framing has two major consequences if the situation is interpreted in social terms. Firstly, every person only maintains in-group relations and does not connect to non-members — therefore getting new social contacts can occur only through joining groups. The second consequence of Newman and Park model is related to the infinite flexibility of their nodes — if the group is big enough and p value is high, each node may end up with thousands and thousands of connections. It is very unlikely to maintain so many social ties because maintaining them does consume cognitive resources that are limited — people cannot infinitely increase the number of relationships*.

The goal of this paper is to introduce more psychologically realistic rules for how groups affect network structure. What we understand as "groups affecting network structure" in this case is that belonging to a group offers opportunities to form social ties with other group members with fairly high probability.

We propose a new model where individuals are only to a certain degree affected by the groups they belong to, but also their individual characteristics are taken into account. The individual characteristics are conceptualized as a pre-set node degree. In other words — pre-set node degree reflects that every person has an inborn tendency to socialize — ranging from recluses to "party animals". The flexibility parameter in our model allows people to change their individual inborn tendency and use social opportunities (here manifested as group membership) to change the number of relationships. To simplify the algorithm we set the level of flexibility as uniform for the population.

The paper is organized as follows. In the next section the concept of assortativity is briefly reviewed. In Sect. 3 we present our model of network construction. In Sect. 4 we present the numerical results of how network assortativity depends on individuals' flexibility. We explain the obtained effects using analytical calculations for a simplified model.

2. Assortativity

The network is called assortative if it exhibits positive degree–degree correlation. There are two commonly used measures of assortativity. One possibility is to measure the mean degree of nearest neighbors of a node as a function of node degree, referred to as nearest neighbor average connectivity (neighbor connectivity) [21, 22]:

*For example, data from the International Social Survey 2001 [20] indicate that the average number of friends remains fairly small (number of friends at work place $\mu = 2.16$, $\sigma = 5.00$; number of friends living near you $\mu = 3.30$, $\sigma = 7.26$; number of other close friends $\mu = 6.10$, $\sigma = 11.55$).

$$\overline{k}_{nn}(k) = \sum_{k'} k' \Pr(k'|k), \tag{1}$$

where $\Pr(k'|k)$ is the conditional probability that a node of degree k (k-node) is connected to k'-node. If the curve (1) falls, the network tends to be disassortative whereas ascending curve indicates assortative behavior. Another possible measure is assortativity coefficient, r, introduced as a normalized (ranging in $[-1, 1]$) connected degree–degree correlation function. For a given network of M edges, r is calculated as follows [11]:

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i (j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \tag{2}$$

where j_i and k_i are the degrees of nodes at the end of i-th edge, $i = 1, 2, \ldots, M$.

3. Description of the model

The central idea of Newman and Park model [10] is that node degree is equal to a fixed fraction p of nodes in all groups to which the given node belongs. As it has been motivated in Sect. 1 our model combines inborn characteristics of an individual (pre-set degree) with tunable influence of its social environment. The network construction proceeds in the following steps:

1. *Assign node degrees according to pre-set node degree distribution.* There are N nodes in the system. Each node is assigned its pre-set node degree $D_i^{\mathrm{pre-set}}$ from degree distribution d, common for all nodes, $d_k^{\mathrm{pre-set}} = \Pr(D_i^{\mathrm{pre-set}} = k)$, $k = 0, 1, 2, \ldots$ These degrees can be modified later depending on node flexibility.

2. *Assign node group degrees according to affiliation distribution.* The model allows for multiple group affiliations, that is, one node can belong to none, one or many groups. Each node is assigned its node group degree A_i from the affiliation distribution a, common for all nodes, i.e. $a_j = \Pr(A_i = j)$, $j = 1, 2, \ldots$ In this article the analysis is restricted to the case where each node belongs to at least one group.

3. *Create groups and assign nodes to groups.* N_{G} groups are created and nodes are assigned to groups randomly (each group is equally probable) to match each node's group degree distribution.

4. *Adjust node degree based on its flexibility.* Flexibility makes it possible to manipulate to what extent nodes change their degrees in the process of link formation, as this ability may vary from no change at all (nodes maintain their initial degree regardless the number of potential "friends" in groups they belong to) to extreme flexibility (nodes change their initial degree to match the number of potential "friends" in groups they belong to). We introduce flexibility as a parameter f ranging from 0 to 1. Let node i be assigned to A_i groups: $G_{i1}, G_{i2}, \ldots, G_{iA_i}$, then its resulting degree equals

$$D_i = \left[(1-f) \times D_i^{\text{pre-set}} + f \times \left(\sum_{j=1}^{N} \mathbf{1}_{G_{i1} \cup G_{i2} \cup \ldots \cup G_{iA_i}}(j) - 1 \right) \right], \quad (3)$$

where $\mathbf{1}$ is an indicator function and $[x]$ denotes the nearest integer to x. We shall denote $d_k = \Pr(D_i = k)$, $k = 0, 1, 2, \ldots$

5. *Create links.* Links are created using a modification of the algorithm for the construction of configuration network [23, 24]. The difference is that nodes that belong to the same group have higher probability (so-called in-group probability, p_{in}) of being connected to each other. Nodes that do not share a group affiliation also may be linked, but the probability (out-group probability, p_{out}) is significantly smaller than p_{in}.

4. Numerical results

Within this paper we examine the mechanism of flexibility. We establish the total number of nodes N (the simulations are performed for $N = 1000$). The pre-set degree is drawn from generalized zero-truncated Burr distribution, $\text{Burr}(x_0, B, c, k)$ characterized by four parameters: location x_0, scale B, and shape $c, k^\dagger$ (Fig. 1). Figures 1 and 2 present assortativity coefficient (2) as a function of flexibility for three different values of N_{G}. Our intuition prompts that if all nodes adjust their degrees to the sizes of their groups, the degree correlation increases. Indeed, if we do not allow multiple affiliation, degree-correlation versus flexibility is an ascending curve, see Fig. 1, top. However, as soon as nodes can be assigned to more than a single group, it is clearly seen that assortativity is not trivially related with flexibility, see Fig. 1, bottom. For multiple affiliations all curves exhibit a peak for some intermediate value of f. It suggests that there are two competing mechanisms related to flexibility that affect degree-correlation. On the one hand, the mechanism of adjustment to the group size is necessary so that the network exhibits assortativity. The second mechanism is related to an effect of multimodal ("torn") distribution of resulting degree distribution. For multimodal we understand that some intermediate values between modes have very low probability. This effect is especially strong for small number of groups N_{G}. For high values of f the pre-set degree becomes irrelevant, the degree distribution depends on the size(s) of group(s) that given node is assigned to. The probability to be assigned to a given group is given as $p = (a_1 + 2a_2 N_{\text{G}} + 3a_3 + \ldots)/N_{\text{G}}$ and therefore the distribution of the group size is Poissonian with intensity (mean value) $\lambda = Np = N\langle A\rangle/N_{\text{G}}$, where $\langle A\rangle$ denotes the average number of affiliations. Clearly, λ increases with decreasing N_{G} but the ratio of the standard deviation to the mean value decreases as $\lambda^{-1/2}$. The degree distribution is approximately concentrated around λ, $2\lambda, \ldots$ The distance between λ and 2λ, 2λ and 3λ, ...

†The cumulative distribution function is of the form $F(x) = 1 - \left[1 + \left(\frac{x-x_0}{B}\right)^c\right]^{-(k+1)}$ for $x > \max(0, x_0)$.

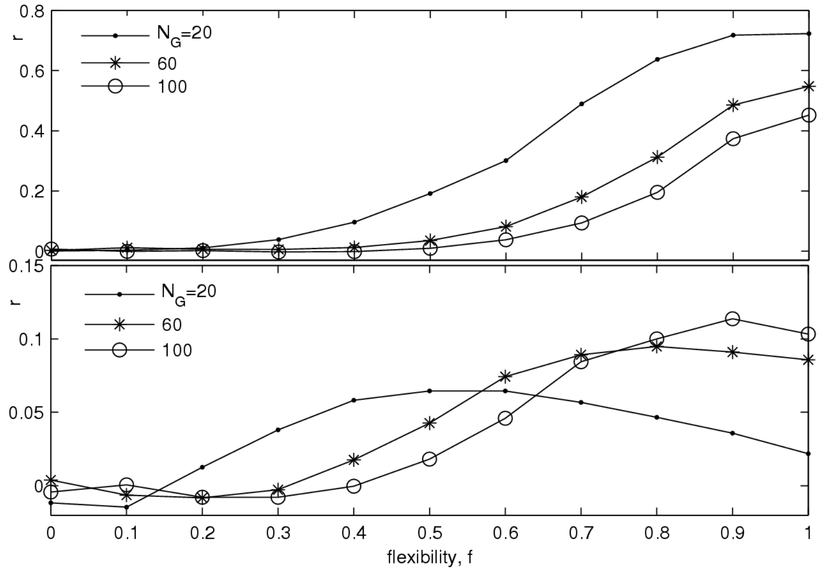

Fig. 1. Assortativity coefficient as a function of flexibility. Top plot: single affiliation, $a_1 = 1$; $p_{\rm out} = 0.01$, $p_{\rm in} = 1$. Bottom plot: multiple affiliation distribution, Burr(0.5,1.5,3,4), $p_{\rm out} = 0.01$, $p_{\rm in} = 1$.

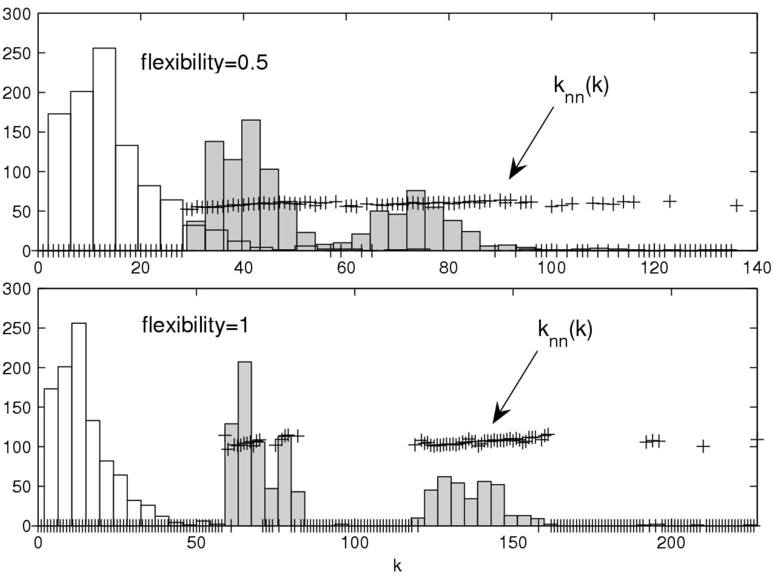

Fig. 2. Histograms of pre-set (white boxes) and resulting (gray) degree distribution for $N_{\rm G} = 20$, $N = 1000$, multiple affiliation distribution Burr(0.5,1.5,3,4), $p_{\rm out} = 0.01$, $p_{\rm in} = 1$. Top plot: $f = 0.5$; bottom plot: $f = 1$.

is covered by $\lambda^{1/2}$ standard deviations. Thus, increasing λ results in multimodal character of degree distribution, see Fig. 2. We attempt to clarify how multimodality affects assortativity in the section to follow.

5. The effect of bimodality

In this section we simplify the assumptions and study the effect of multimodal distribution on assortativity, in particular neighbor connectivity (1), using analytical approach. First let the affiliation distribution be supported in two points only, i.e. $a_1 + a_2 = 1$. Moreover, let $f = 1$, $p_{\text{out}} = 0$, $p_{\text{in}} = 1$. We shall derive the analytical form of (1) and examine its behavior in relation to bimodality of node degree distribution. Let us denote by $j_1(k)$ and $j_2(k)$ the mean degree of neighbors of a k-node that belongs to one group and two groups, respectively. Then $\overline{k}_{nn}(k)$ may be expressed as $j_1(k)w_1(k)+j_2(k)w_2(k)$, where $w_1(k)$ and $w_2(k)$ are appropriate weights representing probabilities that k-node belongs to one (w_1) or two (w_2) groups. In order to assign weights w_1 and w_2 we need the formula for node degree distribution. The group size is Poissonian with intensity $\lambda = N(a_1 + 2a_2)/N_{\text{G}}$. Immediately, the sum of the sizes of two groups is also Poissonian with doubled intensity. However, the groups may overlap and the distribution of the overlap size is also Poissonian with intensity $\lambda' = 2a_2N/[N_{\text{G}}(N_{\text{G}}-1)]$. If a k-node belongs to two groups and the overlap of these groups contains i nodes (which is at least one node and at most $k+1$), then the sum of the sizes of these groups n_1 and n_2 equals $n_1 + n_2 = k + 1 + i$ yielding,

$$d_k = a_1\tilde{p}_1(k+1;\lambda) + a_2\sum_{i=1}^{k+1}\tilde{p}_1^{k+1}(i;\lambda')\tilde{p}_{2i}(k+i+1;2\lambda). \tag{4}$$

We use the notation $\tilde{p}(i,\lambda)$ to denote Poisson probability function with intensity λ, i.e. $\tilde{p}(i,\lambda) = \exp(-\lambda)\lambda^i/i!$ and $\tilde{p}_K^L(i,\lambda) = \tilde{p}(i,\lambda)/\sum_{j=K}^{L}\tilde{p}(j,\lambda)$ to express truncated Poisson distribution, supported in $\{K, K+1, \ldots, L\}$ (if L is infinite we simply write $\tilde{p}_K(i,\lambda)$). Consequently,

$$\overline{k}_{nn}(k) = j_1(k)\frac{a_1\tilde{p}_1(k+1;\lambda)}{d_k} + j_2(k)\frac{a_2\sum_{i=1}^{k+1}\tilde{p}_1^{k+1}(i;\lambda')\tilde{p}_{2i}(k+i+1;2\lambda)}{d_k}. \tag{5}$$

Now, we focus on j_1 and j_2. First, note that the proportion of nodes within a group that belong to one and to two groups, say α_1 and α_2, does not coincide with the affiliation distribution a_1 and a_2. Indeed, there is on average $N(a_1+2a_2)/N_{\text{G}}$ nodes in a group and there is on average Na_1/N_{G} nodes with a single affiliation per group, thus $\alpha_1 = a_1/(a_1+2a_2)$. Let us note that all neighbors of the node belonging to single group are in neighborhood relation with one another. Moreover, α_2 of them have on average $\lambda - i$ additional neighbors in some other group, namely

$$j_1(k) = \frac{1}{k}\sum_{i=1}^{k+1}\tilde{p}_1^{k+1}(i;\lambda')\left[k^2 + \alpha_2 k(\lambda - i)\right] \approx k + \alpha_2\left(\lambda - \frac{\lambda'}{\sum_{j=1}^{k+1}\tilde{p}(j;\lambda')}\right). \tag{6}$$

The formula that approximates $j_2(k)$ is more complicated

$$j_2(k) = \frac{1}{k} \sum_{i=1}^{k+1} \frac{\tilde{p}(i;\lambda')}{\sum_{j=1}^{k+1} \tilde{p}(j;\lambda')} \sum_{n_1=i}^{k+1} P_k(n_1,i) A_k(n_1,i), \tag{7}$$

where

$$P_k(n_1,i) = \frac{\tilde{p}_i^{k+1}(n_1;\lambda)\tilde{p}_i^{k+1}(k+1+i-n_1;\lambda)}{\sum_{n=i}^{k+1} \tilde{p}_i^{k+1}(n;\lambda)\tilde{p}_i^{k+1}(k+1+i-n;\lambda)} \tag{8}$$

reflects the probability that for given overlap i a k-node is affiliated to two groups of sizes n_1 and $n_2 = k+i+1-n_1$. In $A_k(n_1,i)$ we sum up the degrees of k-node's neighbors, i.e.

$$A_k(n_1,i) = (n_1-i)(n_1-1) + (n_2-i)(n_2-1) + (i-1)k$$
$$+[(k+1+i)\alpha_2 - 2i](\lambda-1).$$

In the two first summands we exclude the overlap and sum up all the neighbors that nodes have within their own group. Additionally, α_2 of all nodes belong to two groups: all nodes in the overlap have k neighbors (third summand) and all nodes outside the overlap have on average $\lambda - 1$ additional neighbors.

We obtain the analytic form of neighbors connectivity if we combine (6) and (7) in (5). The saw-like shapes of neighbors connectivity, and simultaneously small values of assortativity coefficient, are particularly distinct for small values of N_{G}. The values of w_1 as a function of k are presented in Fig. 3 (dotted lines). As the number of groups decreases $w_1(k)$ exhibits a sharp fall from 1 to 0 which is related to strongly bimodal character of degree distribution. Moreover, the difference between the values of $j_1(k)$ and $j_2(k)$ for k near the fall of w_1 exhibits the tendency to decrease as N_{G} increases, see Fig. 3.

6. Conclusions

We have shown that group structure can lead to positive assortativity with psychologically sound network construction rules. The crucial parameter in our model is individuals' flexibility controlling the balance between social and individual influences on the number of links. Degree correlation is initially growing with flexibility as flexibility allows individuals to adjust their degrees, which leads to the situation where group members have similar degrees. Together with higher probability of in-group link formation it increases degree correlation. On the other hand, high flexibility splits degree distribution because, on average, individuals belonging to one group have lower degree than individuals belonging to two groups, and they have lower degree than individuals belonging to three groups etc. This creates modes in degree distribution consisting with individuals belonging to one, two or more groups. Once such almost-separate modes are created degree correlation is declining because of the saw-like shape of the average neighbors' degrees.

The debate whether positive assortativity is specific only for social networks may not be fully resolved soon. We hope that this article can contribute to this

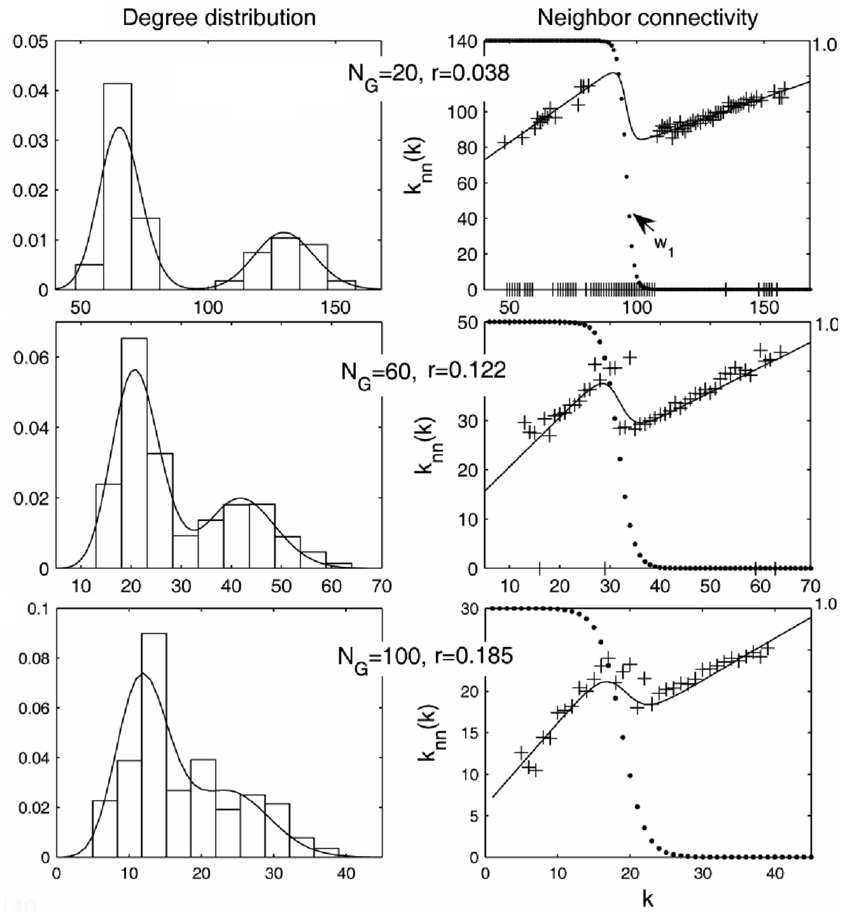

Fig. 3. Left: histograms of degree distribution and corresponding theoretical d_k. Right: analytical (solid line) and numerical (+) results for $k_{nn}(k)$. Analytical values of $w_1(k)$ (dotted line): the w_1 scale is indicated on the right side of the plot; the corresponding values of assortativity coefficient are noted in the top middle part.

debate by showing how positive assortativity can be derived from network creation rules based on group and individual properties.

References

[1] D.J. Watts, S.H. Strogatz, *Nature* **393**, 409 (1998).

[2] R. Albert, A.L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

[3] S.N. Dorogovtsev, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford 2003.

[4] M.E.J. Newman, *SIAM Rev.* **45**, 167 (2003).

[5] M.E.J. Newman, *Phys. Rev. E* **64**, 16131 (2001).

[6] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aaberg, *Nature* **411**, 907 (2001).

[7] M. Girvan, M.E.J. Newman, *Proc. Acad. Natl. Sci.* **99**, 7821 (2002).

[8] A.L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, *Physica A* **311**, 590 (2002).

[9] D.J. Watts, P.S. Dodds, M.E.J. Newman, *Science* **296**, 1302 (2002).

[10] M.E.J. Newman, J. Park, *Phys. Rev. E* **68**, 36122 (2003).

[11] M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).

[12] M.E.J. Newman, *Phys. Rev. E* **67**, 26126 (2003).

[13] D. Whitney, D. Alderson, http://necsi.org/events/iccs6/proceedings.html, *Proc. ICCS*, (2006).

[14] D.L. Alderson, L. Li, *Phys. Rev. E* **75**, 46102 (2007).

[15] M. Catanzaro, G. Caldarelli, L. Pietronero, *Phys. Rev. E* **70**, 037101 (2004).

[16] M. Catanzaro, G. Caldarelli, L. Pietronero, *Physica A* **338**, 119 (2004).

[17] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman, S.H. Strogatz, *Phys. Rev. E* **64**, 041902 (2001).

[18] R. Toivonen, J.P. Onnela, J. Saramäki, J. Hyvönen, K. Kaski, *Physica A* **371**, 851 (2006).

[19] M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera, A. Arenas, *Phys. Rev. E* **70**, 056122 (2004).

[20] http://www.gesis.org/en/data_service/issp/data/2001_Social_Networks_II.htm.

[21] R. Pastor-Satorras, A. Vázquez, A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).

[22] M. Catanzaro, M. Boguna, R. Pastor-Satorras, *Phys. Rev. E* **71**, 027103 (2005).

[23] M. Molloy, B. Reed, *Rand. Struct. Alg.* **6**, 161 (1995).

[24] M. Boguñá, R. Pastor-Satorras, A. Vespignani, *Eur. Phys. J. B* **38**, 205 (2004).