Eur. Phys. J. B **38**, 305–309 (2004)
DOI: 10.1140/epjb/e2004-00018-0

THE EUROPEAN PHYSICAL JOURNAL B

Multi-component static model for social networks

D.-H. Kim, B. Kahng [a], and D. Kim

School of Physics and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea

Received 27 October 2003
Published online 17 February 2004 –

Abstract. The static model was introduced to generate a scale-free network. In the model, N number of vertices are present from the beginning. Each vertex has its own weight, representing how much the vertex is influential in a system. The static model, however, is not relevant, when a complex network is composed of many modules such as communities in social networks. An individual may belong to more than one community and has distinct weights for each community. Thus, we generalize the static model by assigning a q-component weight on each vertex. We first choose a component (μ) among the q components at random and a pair of vertices is linked with a color μ according to their weights of the component (μ) as in the static model. A $(1-f)$ fraction of the entire edges is connected following this way. The remaining fraction f is added with $(q+1)$-th color as in the static model but using the maximum weights among the q components each individual has. The social activity with such maximum weights is an essential ingredient to enhance the assortativity coefficient as large as the ones of real social networks.

PACS. 89.65.-s Social and economic systems – 89.75.Hc Networks and genealogical trees – 89.75.Da Systems obeying scaling laws

1 Introduction

Recently there have been considerable efforts to understand complex systems in terms of random graph, consisting of vertices and edges, where vertices (edges) represent individuals (acquaintances or their interactions) [1–4]. In such complex networks, the emergence of a power-law degree distribution, $P(k) \sim k^{-\gamma}$, is an interesting feature. Such networks are called scale-free (SF) networks. To illustrate such SF behavior, many *in silico* models have been introduced, whose examples include the Barabási and Albert model [5], the Huberman and Adamic model [6], etc. In those models, the number of vertices grows with time.

The static model [7] is another type of *in silico* model designed to generate SF networks, where the number of vertices is fixed. Each vertex is indexed by an integer i ($i = 1, \cdots, N$) and assigned its own weight $w_i = i^{-\alpha}$, where α is a tunable parameter. Next, two different vertices (i, j) are selected with probabilities equal to normalized weights, $w_i / \sum_k w_k$ and $w_j / \sum_k w_k$, respectively, and are connected via an edge unless one exists already. This process is repeated until mN edges are present in the system, so that the mean degree is $2m$. Then it follows that the degree distribution is SF with the exponent $\gamma = 1 + 1/\alpha$. Thus, tuning the parameter α in $[0.5, 1)$, we can obtain a continuous spectrum of the exponent γ in the range $2 < \gamma \leq 3$, for which the degree distribution has finite mean and diverging variance. Since the number of vertices does not grow, one may wonder if this model can be applied to evolving real world network. However, since the model network can be easily generated and exhibits little hump in the degree distribution, it is useful to study many aspects of SF networks.

In this paper, we generalize the static model by allowing a q-component weight $(w_i^{(1)}, w_i^{(2)}, \ldots, w_i^{(q)})$ to each vertex i. We suppose that the μ-th component $w_i^{(\mu)}$ of a vertex i represents its own weight or fitness to a subgroup (μ) ($\mu = 1, \ldots, q$) in a society. For example, we suppose that two persons i and j are alumni of a high school, a subgroup (μ). They would have different weights $w_i^{(\mu)}$ and $w_j^{(\mu)}$ in the subgroup (μ), determined by their school activities. The person i and another person k are colleagues in a company, another subgroup (ν). They have also different weights, $w_i^{(\nu)}$ and $w_k^{(\nu)}$, by their positions in the company, the subgroup (ν). Then the person i has weights $w_i^{(\mu)}$ and $w_i^{(\nu)}$ in different subgroups, which are not the same in general. We make an edge between the pair (i,j) in one color representing the subgroup (μ) and the pair (i,k) in another color for the subgroup (ν). Vertices in the system are connected with edges in q different colors representing different subgroups. Subgroups are then connected each other by weak ties as explained later. Since our society comprises many different subgroups and a person can be acquainted with other people belonging to diverse

[a] e-mail: kahng@phya.snu.ac.kr

subgroups, this generalized static model is useful for modeling social networks.

So far, there have been many attempts to explain the structures and the properties of social networks [8]. Recently, Watts et al. [9] introduced a hierarchical model for social network. In the model, individuals belong to groups that in turn belong to groups of groups and so on, creating a tree-like hierarchical structure of social organization. Here an individual can belong to more than one group, as a result of which the distance between two persons is shorter than the ultrametric distance between them. Such hierarchical model illustrates well the small-world property of social network as implied in the Milgram's "six degrees of separation" [10]. Another simple social network model, introduced by Newman [11], is based on the concept of bipartite graph [12] and community structure [13]. The assortativity of this model was studied in reference [14]. While this model is simple, it reproduces successfully large clustering coefficient and positive value of assortative coefficient. While our q-component model is similar to those models in the spirit of dividing people into subgroups, however, we assign weights to each person for each subgroup, and connections are made following those weights. Also our model is meaningful in the aspect that each vertex is assigned "multi-component" weights. Thus, it is noteworthy that the approach we use is different from those used in references [9,11].

Social network exhibits an interesting feature in the degree-degree correlation function, different from biological or information networks. Newman [15] studied the degree-degree correlation in terms of the correlation function between the remaining degrees of the two vertices on each side of an edge, where the remaining degree means the degree of that vertex minus one. He introduced the assortativity coefficient r, defined as

$$r = \frac{1}{\sigma_q^2} \sum_{j,k} jk(e_{jk} - q_j q_k), \tag{1}$$

where e_{jk} is the joint probability that the two vertices on each side of a randomly chosen link have j and k remaining degrees, respectively. q_k is the normalized distribution of the remaining degree $q_k = (k+1)P(k+1)/\sum_j jP(j)$, and $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$. Interestingly, complex networks can be classified into three types, having $r < 0$, $r \approx 0$ and $r > 0$, called the dissortative, the neutral, and the assortative network, respectively [15]. The assortativity or dissortativity can also be identified by a quantity, denoted by $\langle k_{\rm nn}\rangle(k)$, the average degree of a neighboring vertex of a vertex with degree k [16]. For the dissortative, the neutral, and the assortative mixing, $\langle k_{\rm nn}\rangle(k)$ decreases, remains constant, and increases with respect to k, respectively. Most social networks are assortative as shown in table 1, while the Internet and the protein interaction network are dissortative. While many *in silico* models have been introduced, most of them are neutral. An exception is the growing network model introduced by Callaway et al. [17]. Thus it would be interesting to introduce an *in silico* model having the assortativity coefficient as positive and large as empirical values, which would enable one to understand a basic mechanism of social network formation. We will show that such assortative networks can be generated via the q-component static model.

Table 1. The size N, the mean degree $\langle k\rangle$, the diameter d, and the assortativity coefficient r for a number of social networks.

Name	N	$\langle k\rangle$	d	r	Ref.
cond-mat	16,264	5.85	6.628	0.185	[18]
arXiv.org	52,909	9.27	6.188	0.363	[18]
Mathematics	78,835	4.16	8.455	0.672	[19]
Neuroscience	205,202	11.79	5.532	0.604	[19]
Video movies	29,824	33.69	4.789	0.222	[20]
TV miniseries	33,980	73.04	3.845	0.379	[20]
TV cable movies	117,655	55.48	3.796	0.135	[20]
TV series	79,663	118.44	4.595	0.529	[20]

2 Model

The q-component static model network is constructed as follows. Initially, N vertices are present in the system, representing N people in a society. Each vertex is assigned a q-component weight $(w_i^{(1)}, w_i^{(2)}, \ldots, w_i^{(q)})$, where i is the vertex index. $w_i^{(\mu)}$, the μ-th weight of a vertex i, is given as $\ell_{i,\mu}^{-\alpha_\mu}$, where $\ell_{i,\mu}$ is the rank of the vertex i in the μ-th subgroup. We take $\{\ell_{1,\mu}, \ldots, \ell_{N,\mu}\}$ be a random permutation of the integers $\{1, \ldots, N\}$. α_μ is also taken to be a real random number distributed uniformly in the range $[0.5, 1)$. In general, ranks of a person for different subgroups should be correlated in real society; however, we take them as independent in this work for simplicity. As the number of people N becomes large, the number of distinct subgroups q can increase in real world. Thus, q is presumed to depend on N linearly.

Edges are connected as follows: First, among the q components, we choose a component μ at random. Second, we choose two vertices (i, j) with probabilities equal to normalized weights, $p_i^{(\mu)} \equiv w_i^{(\mu)}/\sum_k w_k^{(\mu)}$ and $p_j^{(\mu)} \equiv w_j^{(\mu)}/\sum_k w_k^{(\mu)}$, and attach an edge between them with the μth color unless an edge in that color exists already. Edge color is distinct for each component. Note that the pair (i, j) can be connected via more than one edge in different colors. Edges in different component are distinguished by their own colors. The process of attaching such edges is repeated until $(1-f)mN$ edges are added to the system. f is a parameter between 0 and 1. We will see that m is related to the average degree. Since the component was chosen randomly, the number of edges in one color is $(1-f)mN/q$ on average.

To construct a minimal model mimicking social relations, we need elements playing the role of "weak ties" [21]. So, to take into account of social relationships among people having different backgrounds, we suppose that additional social relationships are formed following the maximum weights among the q components each individual has. The normalized maximum weight of vertex i is defined

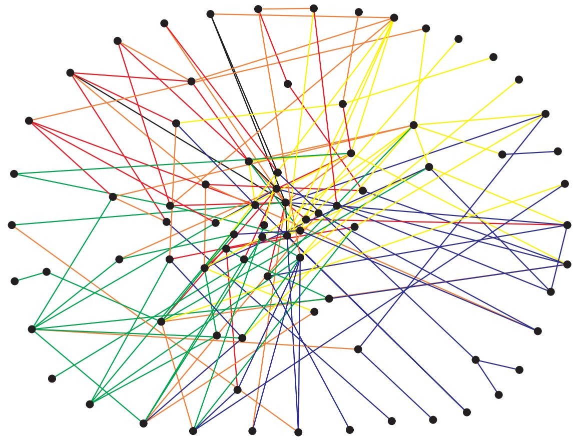

Fig. 1. A network of the q-component static model with parameters $N = 80$, $m = 2$, $q = 4$ and $f = 0.2$. Edges in four colors (red, yellow, green and blue) are the connections within each group. Edges in orange are those formed by maximum weights. Edges in more than two colors are colored in black.

as $w_{i,m} = \max(p_i^{(1)}, \ldots, p_i^{(q)})$. Then two distinct vertices i and j are chosen with probabilities, $w_{i,m}/\sum_k w_{k,m}$ and $w_{j,m}/\sum_k w_{k,m}$, respectively, and are linked by a new color different from the previous q colors unless such an edge exists already. This process is repeated until fmN edges are formed. Edges formed by such maximum weights are the similar to the activity that people make acquaintances with strangers belonging to different groups by exchanging their own name cards where social status representing their own maximum weights is printed. Such edges can thus be regarded as weak ties, introduced by Granovetter [21] which play an important role in social networks, connecting different subgroups. We find that the assortativity coefficient is enhanced by the presence of such weak ties.

Networks constructed in this way have mN edges with $(q+1)$ colors representing internal structure of subgroups. So, some pair of vertices are linked by more than one edges in different colors, albeit such incidences are not so frequent when q is large. However, when we measure the network properties such as the shortest pathways, the degree of a vertex, the assortativity coefficient, and so on, we regard those multiple edges as a single one. Thus the mean degree $\langle k \rangle$ is slightly less than $2m$ by about 5% for typical networks we consider below. A small size network constructed in this way is shown in Figure 1.

3 Simulation results

We perform numerical simulations for various values of q, f, m and N, and examine the degree distribution $P(k)$, the diameter d, the assortativity coefficient r, $\langle k_{\rm nn} \rangle (k)$, and the clustering coefficients $C(N)$ and $C(k)$ as functions of those parameters. Here the diameter is the average distance between a connected pair of vertices along the shortest pathways. All numerical results presented in this paper are averaged over 100 configurations without further specification.

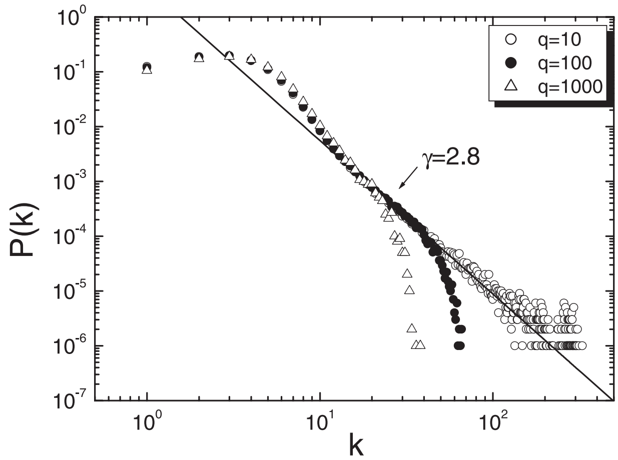

Fig. 2. The degree distribution $P(k)$ vs. the degree k obtained with $N = 10000$, $m = 2$, and $f = 0.2$ for $q = 10$, 100 and 1000.

First, the shape of the degree distribution depends on the number of subgroups. As shown in Figure 2, for small q, the degree distribution follows a power law with $\gamma \approx 2.8$. However, for large q, it has a highly skewed form, approximately obeying a power law for a part of its range, and having an apparently exponential cutoff for larger k. The SF behavior for intermediate k originates from the SF behavior inside each subgroup. The degree k_i of a vertex i is proportional to the sum of $w_i^{(\mu)}$ over $\mu = 1, \ldots, q$ and $w_{i,m}$. Since there is no any correlation in $w_i^{(\mu)}$ for different (μ)s, it is not obvious what the maximum degree the vertex i has. Therefore large fluctuation in degree for large k arises due to the absence of the correlation in $w_i^{(\mu)}$, resulting in the exponential-type decay in $P(k)$. A similar crossover behavior can be found in real social networks, for example, the collaboration networks of physicists, biologists and movie stars (Fig. 1 of Ref. [22]). The degree distribution shows a peak at low degree, which is also shown in several other real social networks such as the coauthorship network in the field of neuro-science (Fig. 2b of Ref. [19]) and the movie star co-playing network (Fig. 1b of Ref. [22]).

Second, we examine the assortative coefficient r as a function of f for a fixed N and several values of m and q. As shown in Figure 3, the assortativity coefficient exhibits a peak around $f \approx 0.2$, meaning that the connections among subgroups are mostly optimized. Thus we limit our further consideration to the case $f = 0.2$. Meanwhile, the diameter gradually decreases with increasing f as shown in the inset of Figure 3.

Third, we study the assortativity coefficient r as a function of N for various m and q. It is likely that r increases with increasing N as $r \sim \ln\ln N$ as shown in Figure 4, but it seems to saturate for larger N. It also increases with increasing m and q, as shown in Figure 4. Thus the q-component static model exhibits r values as large as empirical values listed in Table 1. Some numerical results of r listed in Table 2 show a quantitative agreement

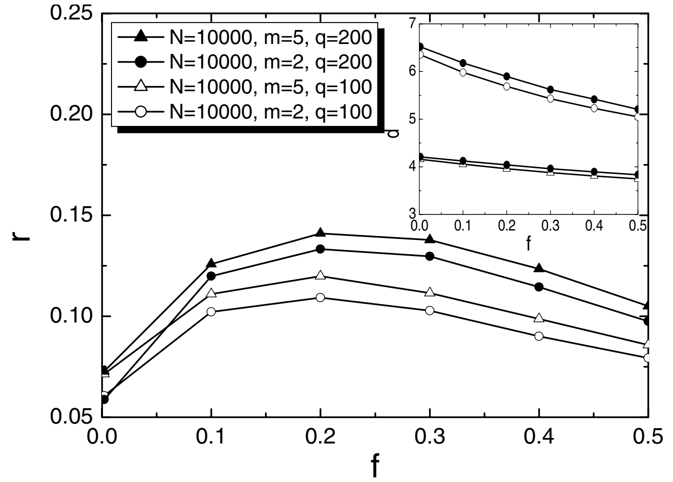

Fig. 3. The assortativity coefficient r vs. the parameter f for various m and q. Inset: f dependence of the diameter d.

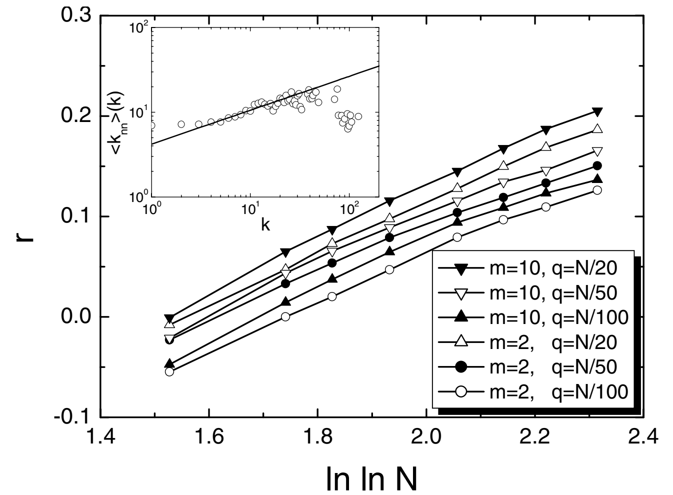

Fig. 4. The assortativity coefficient r vs. $\ln \ln N$ for various m and q values. Inset: Plot of $\langle k_{\rm nn}\rangle(k)$ vs. degree k for $N = 10000$, $m = 2$, $q = 100$ and $f = 0.2$. The slope of the fit line is 0.4.

Table 2. Typical simulation results of the diameter d and the assortative coefficient r obtained under selected conditions of N, m and q with $f = 0.2$.

N	m	q	d	r	similar network
16,000	2	3,200	6.598	0.174	cond-mat
30,000	5	2,000	4.550	0.218	Video movies

with the ones obtained in real social networks. Such an assortative nature can also be checked by positive slope of $\langle k_{\rm nn}\rangle(k)$ [16]. As expected, it increases with increasing k, as shown in Inset of Figure 4.

Fourth, the diameter d is investigated as a function of the number of vertices N for various m and q, as shown in Figure 5. It is found that the diameter is proportional to $d \sim \ln N$ as in the case of random graph, in which $d \sim \ln N/\ln\langle k\rangle$ [23]. Furthermore, the diameter becomes smaller as m increases, which is like the case of random

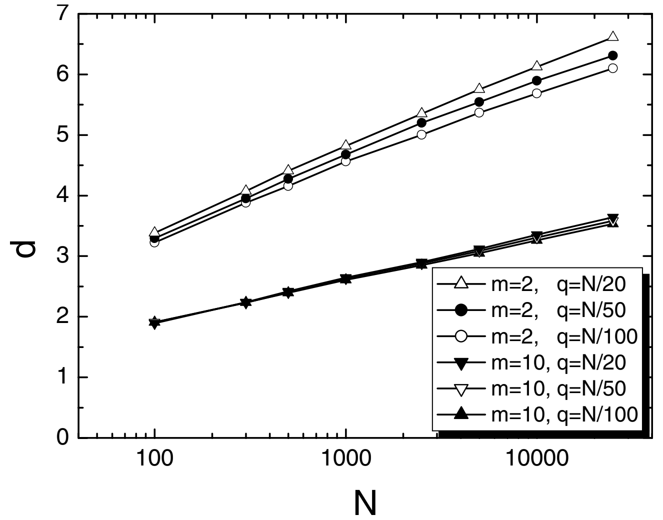

Fig. 5. The diameter d vs. the size N for the same parameters used in Figure 4.

graphs. However, the diameter is almost insensitive to q. To test the so-called "six degrees of separation", we extrapolate the straight line in the semi-logarithmic plot of Figure 5 to large N for $m = 10$. We obtain $d \approx 6.0$ for $N = 10^8$ and $d \approx 6.7$ for $N = 10^9$, in reasonable agreement with the Milgram's "six degrees of separation" [10]. Here the choice of $m = 10$ comes from the facts that a person knows about 20 other people on average (see Chap. 5 of Ref. [24]).

Recently, it was observed that the diameter of a coauthorship network is independent of system size N [19]. Though it is not clear whether such abnormal behavior is intrinsic or not, we can reproduce such behavior easily by modifying our model in the following way. It is known that diameter d is given as $d \sim \ln N/\ln\langle k\rangle$ with mean degree $\langle k\rangle$ in random graph theory [23]. Thus if the number of edges grows much faster than the number of vertices, i.e., in so-called accelerated growth way, then $\langle k\rangle$ depends on N, leading to the result that diameter is independent of system size N. We also confirm this behavior with our model. When we assign the number of edges as proportional to $N^{1+\theta}$ ($\theta > 0$), we find that the diameter d approaches a constant value as N increases.

Fifth, one of the properties well studied for complex networks is the clustering coefficient C, which is defined as the average over all vertices of the ratio of the number of triangles connected to a given vertex to the number of triples centered on that vertex. It is known that for the neutral network, the clustering coefficient behaves as $C(N) \sim N^{(7-3\gamma)/(\gamma-1)}$ [4,11]. Thus when $\gamma = 3$, $C(N) \sim N^{-1}$. For the q-component static model, while r is not close to zero, the rule of attaching edges is such that no explicit degree-degree correlation enters, so that it is natural to expect the behavior $C(N) \sim N^{-1}$. Indeed, the measured behavior is close to the expected one as shown in Figure 6, but the slope in Figure 6 exhibits small deviations for smaller q. For neutral networks, it is known that the clustering coefficient of a vertex with k is

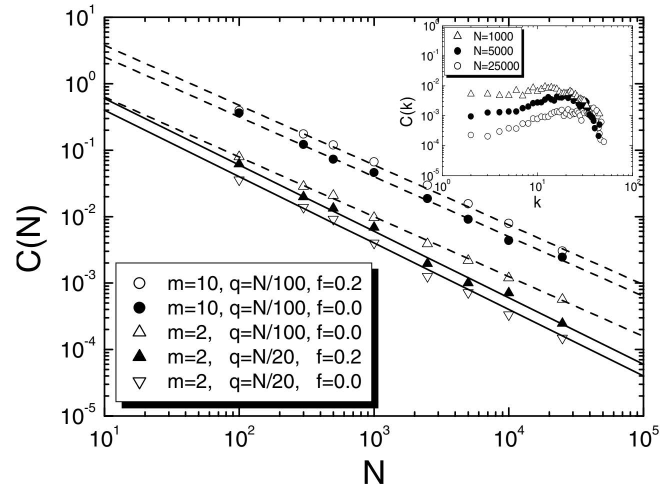

Fig. 6. Plot of $C(N)$ vs. the size N for various m, q and f values. The slopes of the fit lines are -1.0 for solid and –0.9 for the dashed lines, drawn for the eye. Inset: Plot of $C(k)$ vs. degree k for various N with fixed $m = 2$, $q = N/20$ and $f = 0.2$.

almost independent of degree k. Even in our case, we find that indeed $C(k)$ is independent of k for different N as shown in the inset of Figure 6.

4 Conclusions

We have introduced the q-component static model assigning a q-component weight to each vertex. The weight of a given component of a vertex mimics a weight or fitness of that person in that subgroup. Through this model, we obtained the diameter of the acquaintance network as small as the Milgram's "six degrees of separation" and the assortativity coefficient as positive and large as empirical values for a variety of social networks. The key rule of our model is motivated from the spirit that the social activity among people is engaged following maximum weights each individual has, playing a crucial role in producing an assortativity coefficient. Moreover, we obtain the degree distribution in a skewed form, which is also similar to those of real world social networks. The clustering coefficients $C(N)$ and $C(k)$ behave as those of a neutral network, being due to the absence of intrinsic degree-degree correlation. Such deficiency of the present model may be improved by introducing the hierarchical structure among subgroups, or correlated ranks of a person for different subgroups, or nonuniform subgroup size, etc.

Note that the generalization of the static model can be applied to many other models. For example, the fitness model introduced in reference [25] can also be generalized into a multi-component case. In this case the μ-th component connection is made between vertices i and j with probability $f(x_i^{(\mu)}, x_j^{(\mu)}) = \theta[x_i^{(\mu)} + x_j^{(\mu)} - z^{(\mu)}]$ where $x_i^{(\mu)}$ is the same as defined earlier, $\theta(x)$ is the Heaviside step function, and $z^{(\mu)}$ is a threshold of the μ-th component.

The authors would like to thank K.-I. Goh for sharing his code for the static model and valuable discussions. This work is supported by the KOSEF Grant No. R14-2002-059-01000-0 in the ABRL program and BK21 program of Ministry of Education, Korea.

References

1. S.H. Strogatz, Nature **410**, 268 (2001)
2. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
3. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
4. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
5. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
6. B.A. Huberman, L.A. Adamic, Nature **401**, 131 (1999)
7. K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. **87**, 278701 (2001)
8. For example, E.M. Jin, M. Girvan, M.E.J. Newman Phys. Rev. E **64**, 046132 (2001); J. Davidsen, H. Ebel, S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002); L. López, M.A.F. Sanjuán, Phys. Rev. E **65**, 036107 (2002); K. Klemm, V.M. Eguiluz, R. Toral, M.S. Miguel, Phys. Rev. E **67**, 026120 (2003); A. Vazquez, Phys. Rev. E **67**, 056104 (2003); G. Csányi, B. Szendrői, `cond-mat/0305580`
9. D.J. Watts, P.S. Dodds, M.E.J. Newman, Science **296**, 1302 (2002); See also J.M. Kleinberg, in *Proceedings of the 2001 Neural Information Processing Systems Conference* (MIT Press, Cambridge, 2002); A.E. Motter, T. Nishikawa, Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003)
10. S. Milgram, Psychology Today **1**, 60 (1967); J. Travers, S. Milgram, Sociometry **32**, 425 (1969)
11. M.E.J. Newman, Phys. Rev. E **68**, 026121 (2003)
12. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E **64**, 026118 (2001)
13. M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA **99**, 8271 (2002)
14. M.E.J. Newman, J. Park, Phys. Rev. E **68**, 036122 (2003)
15. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002); Phys. Rev. E **67**, 026126 (2003)
16. R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001)
17. D.S. Callaway, J.E. Hopcropt, J.M. Kleinberg, M.E.J. Newman, S.H. Strogatz, Phys. Rev. E **64**, 041902 (2001)
18. M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001); Phys. Rev. E **64**, 016131 (2001); Phys. Rev. E **64**, 016132 (2001)
19. A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, Physica A **311**, 590 (2002)
20. `http://www.imdb.com`
21. M.S. Granovetter, Am. J. Sociol. **78**, 1360 (1973)
22. M.E.J. Newman, D.J. Watts, S.H. Strogatz, Proc. Natl. Acad. Sci. USA **99**, 2566 (2002)
23. F. Chung, L. Lu, Adv. Appl. Math. **26**, 257 (2001)
24. A.-L. Barabási, *Linked: The New Science of Networks* (Perseus, Cambridge, 2002)
25. G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002)