Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 351 (2005) 671–679
PHYSICA A
www.elsevier.com/locate/physa

Modelling hierarchical and modular complex networks: division and independence

D.-H. Kim[a], G.J. Rodgers[b], B. Kahng[a,*], D. Kim[a]

[a]*School of Physics and Center for Theoretical Physics, Seoul National University NS50, Seoul 151-747, Republic of Korea*
[b]*Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK*

Received 6 June 2004
Available online 8 January 2005

Abstract

We introduce a growing network model which generates both modular and hierarchical structure in a self-organized way. To this end, we modify the Barabási–Albert model into the one evolving under the principles of division and independence as well as growth and preferential attachment (PA). A newly added vertex chooses one of the modules composed of existing vertices, and attaches edges to vertices belonging to that module following the PA rule. When the module size reaches a proper size, the module is divided into two, and a new module is created. The karate club network studied by Zachary is a simple version of the current model. We find that the model can reproduce both modular and hierarchical properties, characterized by the hierarchical clustering function of a vertex with degree k, $C(k)$, being in good agreement with empirical measurements for real-world networks.

PACS: 89.65.−s; 89.75.Hc; 89.75.Da

Keywords: Networks; Social networks; Hierarchical and modular structure

Recently, considerable effort has been made to understand complex systems in terms of random graphs, consisting of vertices and edges [1–5]. Such complex

*Corresponding author. Tel.: +82 2 880 1326; fax: +82 2 884 3002.
E-mail address: kahng@phya.snu.ac.kr (B. Kahng).

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2004.12.026

networks exhibit many interesting emerging patterns as follows: first, the degree distribution follows a power-law, $P(k) \sim k^{-\gamma}$, where the degree is the number of edges connecting to a given vertex [6]. Such networks, called scale-free (SF), are ubiquitous in the real world. To illustrate such SF behavior in the degree distribution, Barabási and Albert (BA) [6] introduced an in silico model: initially, fully-connected m_0 vertices exist in a system. At each time step, a vertex is newly added and connects to m existing vertices, which are chosen with a probability linearly proportional to the degree of target vertex. Such a selection rule is called the preferential attachment (PA) rule.

Second, the degree–degree correlation in real-world networks is nontrivial. The nontrivial behavior is measured in terms of the mixing coefficient r [7], a Pearson correlation coefficient between the degrees of the two vertices on each side of an edge. Complex networks can be classified according to the mixing coefficient r into three types, having $r<0$, $r\approx 0$, and $r>0$, called the disassortative, the neutral, and the assortative network, respectively [7]. Such classifications 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 [8]. For the assortative (disassortative) network, $\langle k_{\rm nn}\rangle(k)$ increases (decreases) with increasing k, i.e., a power law $\langle k_{\rm nn}\rangle(k) \sim k^{-\nu}$ is satisfied where ν is negative (positive) for the assortative (disassortative) network [8].

Third, many real-world networks have modular structures within them. Modular structures form geographically in the Internet [9], functionally in metabolic [10] or protein interaction networks [11], or following social activities in social networks [12,13]. Such modular structures are characterized in terms of the clustering coefficient. Let C_i be the local clustering coefficient of a vertex i, defined as $C_i = 2e_i/k_i(k_i-1)$, where e_i is the number of edges present among the neighbors of vertex i, out of its maximum possible number $k_i(k_i-1)/2$. The clustering coefficient, C, is the average of C_i over all vertices. $C(k)$ means the clustering function, the average of C_i over the vertices with degree k. When a network is modular and hierarchical, $C(k) \sim k^{-\beta}$ and C remains finite for large system size N [10,14]. In the BA model with $\gamma = 3$, however, $C(k)$ is independent of k, but decreases with N [2,14], because the BA model does not contain modules.

In this paper, we are interested in modelling complex networks including both modular and hierarchical structure not in a deterministic way, but in a self-organized way. In real-world networks, modules represent communities which may evolve as time passes and such modules form hierarchical structure. The karate club (KC) network, originally proposed by Zachary [15], is a simple example of a real-world social network containing community structure. Recently, Newman and Girvan [12] studied the KC network to test their new algorithm for clustering communities [12,16,17]. Here we notice that the KC network contains ingredients, division and independence, which forms a modular structure, in addition to growth and PA principles as noticed in the BA model. Thus, we introduce a network model evolving by such rules, and perform numerical simulations for large system size. Indeed, we find that the model exhibits a characteristic feature of both modular and hierarchical structure, $C(k) \sim k^{-1}$, as much as those for empirical data.

To be specific, the main dynamic process of the evolution of the KC network is as follows. In a KC, a disagreement develops between the administrator of the club and the club's instructor as time goes on, ultimately resulting in the instructor leaving (division) and finding a new club (independence), accompanied by about half of the original club's members. This generic feature of division and independence can be observed in many other social communities such as schools, companies, churches, clubs, parties, etc.

It would be interesting to understand the mechanism for the formation of such modular structures through an in silico model, which is our goal in this paper. Note that Newman introduced simple analytic models generating modular structure with a fixed number of vertices, which is based on bipartite graph [18,19]. Watts et al. [20] introduced static social network models where individuals belong to groups that in turn belong to groups of groups and so on, creating a tree-like hierarchical structure of social organization. Our model is, however, a growth model, where the numbers of vertices and modules increase with time.

To model the evolution of the KC network, we modify the BA model by assigning a color to each vertex. The color assigned to each vertex indicates the group the vertex belongs to. The dynamic rule of our model is as follows:

BA model (*Growth and PA*) Initially, there exist m_0 vertices. They are fully connected. Each vertex i is assigned the same index of color $\mu_i = 1$. Thus the total number of distinct colors $q = 1$. At each time step, a vertex is introduced and connects to m existing vertices following the PA rule. To diversify the model, here m is not fixed, but distributed uniformly among integers in the range $[1, m_c]$, where m_c, a control parameter, is taken as m_0 for convenience. Note that the diversification of m is taken to compare our model with the KC network later. The new vertex j is also assigned the index of color $\mu_j = 1$ and this process is repeated until the number of vertices reaches group size n_1. Here n_1 is not fixed again, but is chosen randomly from the integers in the range $[1, n_c]$, where n_c is another control parameter. This process defines the first group $q = 1$. Thus the model contains two control parameters (m_c, n_c).

Division and independence Then we identify the two vertices i and j among the group with the largest and the second largest degree, respectively, for division and independence. Then the vertex j declares independence and changes its color to a new one, i.e., $\mu_j = q + 1$. Then, each remaining vertex $k(\neq i, j)$ in the group having the same color as vertex i measures the distances $d(k, i)$ and $d(k, j)$ to the vertices i and j, respectively. If $d(k, i) \leqslant d(k, j)$, then the vertex k retains the index of color as it is; otherwise, it changes its color index to that of j. Then the system now comprises of $q + 1$ different groups, and the range of the color index μ is $q' = q + 1$. We then denote q' simply by q. So the newest group has the new color q. The new module is assigned a modular size n_q, which is chosen randomly from the integers in $[1, n_c]$. It is natural to suppose that module sizes are diverse in real-world networks. Also note that if the module size is fixed, then we could not obtain the scale-free behavior of the degree distribution.

Growth and PA again If $q > 1$, then a newly added vertex ℓ chooses one of q colors, μ, with equal probability, and m, the number of outgoing links, also randomly from

the integers in $[1, m_c]$. Then m existing vertices are chosen in the group with the color μ following the PA rule, and m edges are inserted between them and the new node. Note that when m exceeds the number of existing vertices in the selected module, then the rest of the unconnected edges are ignored. When the size of the group μ reaches its own characteristic size n_μ assigned upon its birth in step (ii), we repeat the step of division and independence (ii) in that group only. Otherwise, we repeat step (iii).

We note that through the process of division and independence, hierarchical ordering is made among modules in a self-organized way, because a part of vertices in each module are still connected to vertices in its ancestor module. The hierarchical and modular structure generate the behavior of $C(k) \sim k^{-1}$.

The network constructed in this way is shown in Fig. 1 based on the same number of vertices as the empirical data of the KC network. The structure of the model is different from the BA model due to the presence of modular structure. Note that in our model, one vertex may transfer from one group to another as time goes on, that is, a vertex can change its color as it transfers to a new group. This characteristic is different from that of the q-component static model proposed by the current authors [21], where each individual belongs concurrently to q different groups such as high school alumni, college alumni, company, etc. These two models may reflect different aspects of our social community.

Based on the empirical data by Zachary, we obtain topological properties of the KC network, which are listed in Table 1 and Fig. 2. Until now, it has been believed that social networks are generally assortative [7,19]. But, in "division and independence" social networks such as the KC network, each element is connected to the other in a hierarchical way, without any mediator, generating a disassortative network, as shown in Table 1 and Fig. 2. Since different colors represent distinct

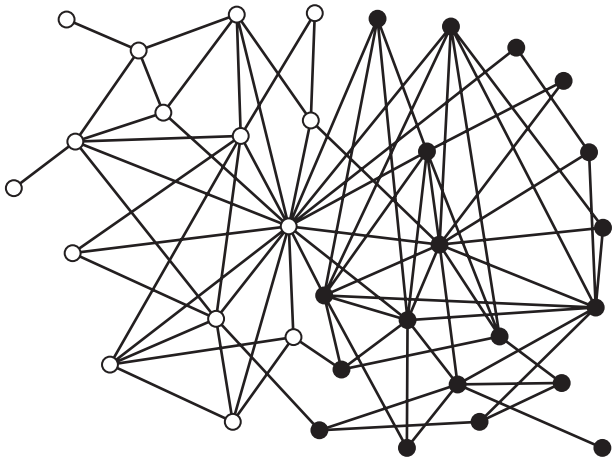

Fig. 1. A snapshot of the model network with parameters $N = 34$, $m_c = 4$ and $n_c = 17$, resembling the karate club network proposed by Zachary. Here two groups are identified by (○) and (●).

Table 1
Mean degree $\langle k\rangle$, the diameter d, the assortativity coefficient r, and the clustering coefficient C obtained from Zachary's KC network and from ours with parameter $N = 34$, $m_c = 4$ and $n_c = 17$. The parameter $m_c = 4$ was chosen to fit $\langle k\rangle$ of our model network to the one of KC network. All the numerical values for the model are averaged over ten configurations

Name	N	$\langle k\rangle$	d	r	C
Zachary's	34	4.59	2.41	−0.48	0.59
Ours	34	4.61	2.54	−0.19 (−0.22[a])	0.56

[a]Note that Zachary presumed that the edge between the administrator and the instructor of the club no longer exists upon division and independence. Following Zachary's way, we obtain $r = -0.22$ in our model.

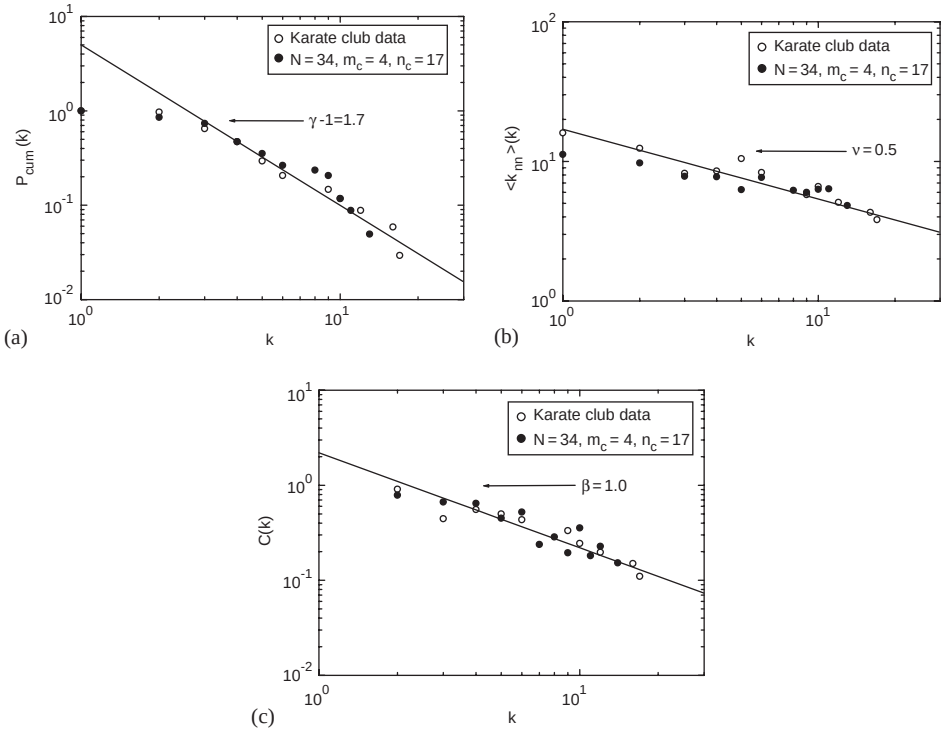

Fig. 2. Plots of the cumulative degree distribution $P_{\rm cum}(k)$ (a), $\langle k_{\rm nn}\rangle(k)$ (b), and the clustering coefficient $C(k)$ (c) versus degree k. In all, the empirical data and the data from the model are denoted by (◯) and (●), respectively. The parameters for the model network are the same as used in Fig. 1. Lines are drawn as a guide to the eye.

modules [10,14] or communities [12], connections are very tight. Thus it is expected that the clustering coefficient C is non-trivially large [18,19]. Table 1 shows the disassortativity and the highly clustered nature of the KC network and our model.

Agreements between the two are excellent except for the mixing coefficient r. Note that the r value of the model is not close to zero although we used the BA-type random attachment rule. Fig. 2 shows that the degree distribution, $P(k) \sim k^{-2.7}$, $\langle k_{\rm nn}\rangle(k) \sim k^{-0.5}$ and $C(k) \sim k^{-1.0}$ of the KC network, which are also in good agreement with those obtained from the present model network.

We investigated the topological properties of our model network for various m_c and n_c with large system size N. For $N = 10^4$, $m_c = 4$, and $n_c = 100$, the degree distribution $P(k)$ follows a power law with the exponent $\gamma \approx 2.4$ in Fig. 3(a) and $\langle k_{\rm nn}\rangle(k)$ shows a disassortative behavior with the exponent $v \approx 0.8$ in Fig. 3(b). The exponents γ and v depend on the parameter n_c as tabulated in Table 2.

The behavior of $C(k)$ is interesting. Fig. 4 shows the n_c-dependence of the hierarchical clustering function $C(k)$. For $n_c = 10, 100, 500$, and 1000 with $N = 10^4$, roughly speaking, $C(k)$ is likely to be fit to $\sim k^{-1}$; however, the data are more scattered with increasing n_c. When $n_c = N$, $C(k)$ is almost independent of k, akin to that of the BA model (Fig. 2(b) of Ref. [14]). Note that the behaviors of $C(k)$ in Fig. 4 are reminiscent of those of the real-world networks studied in previous works. The $C(k)$ of $n_c = 100, 500$, and 1000 resemble those found in the actor network (Fig. 3(a)

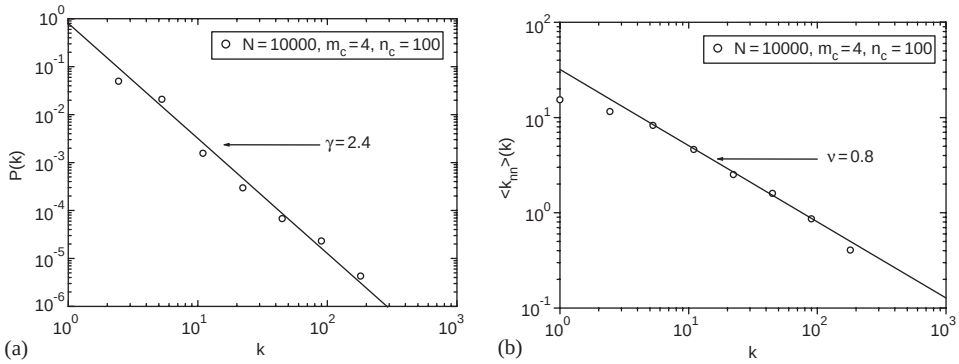

Fig. 3. Plots of the degree distribution $P(k)$ (a) and $\langle k_{\rm nn}\rangle(k)$ (b) versus degree k, both obtained by logarithmic binning method. The data in all figures are obtained with parameters $N = 10\,000$, $m_c = 4$ and $n_c = 100$. In this case we obtain the mean degree $\langle k\rangle = 4.91$, the diameter $d = 7.86$, the assortativity coefficient $r = -0.27$, and the clustering coefficient $C = 0.46$.

Table 2
The n_c-dependence of the exponents, γ, v, and β for $N = 10^4$ and $m_c = 4$. The exponent values are obtained by using both the cumulative and the logarithmic binning method

n_c	100	500	1000	10000
γ	2.4	2.5	2.7	2.9
v	0.8	1.2	1.6	0.0
β	1.0	1.0	1.0	0.0

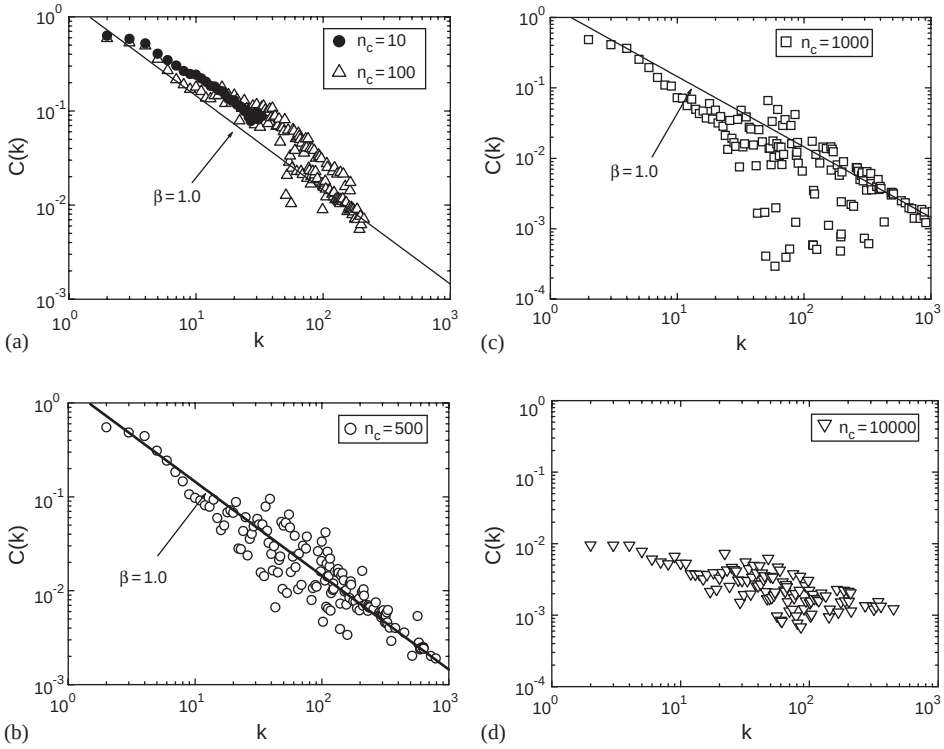

Fig. 4. The clustering coefficient $C(k)$ versus degree k obtained with the parameters $N = 10^4$, $m_c = 4$, and $n_c = 10$ and 100 (a), 500 (b), $n_c = 1000$ (c) and 10 000 (d).

of Ref. [14]), the internet autonomous system (Fig. 3(d) of Ref. [14]), and the world wide web (WWW) (Fig. 3(c) of Ref. [14]), respectively.

Fig. 5 shows the m_c-dependence of the hierarchical clustering function $C(k)$. When m_c is comparable to n_c, $C(k)$ is a constant for small k, while it decays as $C(k) \sim k^{-1.0}$ for large k. The constant value is close to 1, because every vertex within a certain module is almost fully connected. Our model can reproduce the modular and hierarchical clustering structures of real-world networks qualitatively well by controlling the two parameters m_c and n_c properly. As an example, the case of $m_0 = 10$ and $n_c = 200$ of Fig. 5 shows plateau as well as scattered behavior in the middle of the power law regime, which are very similar to the actor network (Fig. 3(a) of Ref. [14]). Note that we present the scattered data as they are in Figs. 4, and 5, to compare them with those of the real networks presented in other previous works.

In conclusion, we have generalized the BA model by assigning a color to each vertex for the purpose of modelling modular complex networks in a simple way. The model evolves with time under the rule of division and independence, in a manner reminiscent of the KC network. Through this model, we confirmed the behavior of the hierarchical clustering function, which behaves in accordance with the ones

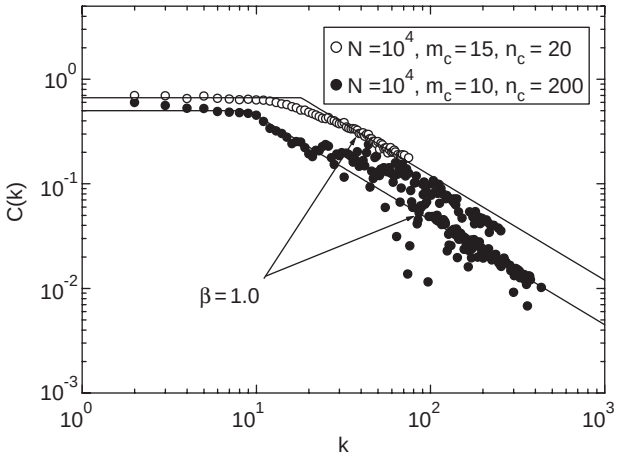

Fig. 5. The clustering coefficient $C(k)$ versus degree k obtained under two selected conditions with large values of m_c.

obtained from the deterministic hierarchical structure and the empirical data such as the internet, the WWW, the actor and language networks [14]. Also it was found that our model exhibits an disassortative mixing behavior as observed in the KC network. Finally, we suggest that the rule of division and independence could be used in constructing modular complex networks in various fields, for example, bio-complex networks, where the strong mutation of a gene may correspond to transferring from one group to another [22].

This work is supported by the KOSEF Grant no. R14-2002-059-01000-0 in the ABRL program, Korea and by the Royal Society, London.

References

[1] S.H. Strogatz, Nature 410 (2001) 268.
[2] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[3] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51 (2002) 1079.
[4] M.E.J. Newman, SIAM Review 45 (2003) 167.
[5] B. Kahng, K.-I. Goh, D.-S. Lee, D. Kim, Saemulli 48 (2004) 115.
[6] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[7] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701;
M.E.J. Newman, Phys. Rev. E 67 (2003) 026126.
[8] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.
[9] K.A. Eriksen, I. Simonsen, S. Maslov, K. Sneppen, Phys. Rev. Lett. 90 (2003) 148701.
[10] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.-L. Barabási, Science 297 (2002) 1551.
[11] A.W. Rives, T. Galitski, Proc. Natl. Acad. Sci. USA 100 (2003) 1124;
V. Spirin, L.A. Mirny, Proc. Natl. Acad. Sci. USA 100 (2003) 12123.

[12] M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA 99 (2002) 8271;
M. Girvan, M.E.J. Newman, (cond-mat/0308217);
M. Girvan, M.E.J. Newman, (cond-mat/0309508).
[13] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, A. Arenas, Phys. Rev. E 68 (2003) 065103(R);
A. Arenas, L. Danon, A. Diaz-Guilera, P.M. Gleiser, R. Guimera, (cond-mat/0312040).
[14] E. Ravasz, A.-L. Barabási, Phys. Rev. E 67 (2003) 026112.
[15] W.W. Zachary, J. Anthropol. Res. 33 (1977) 452.
[16] H. Zhou, Phys. Rev. E 67 (2003) 041908;
H. Zhou, Phys. Rev. E 67 (2003) 061901.
[17] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, (cond-mat/0309488).
[18] M.E.J. Newman, Phys. Rev. E 68 (2003) 026121.
[19] M.E.J. Newman, J. Park, Phys. Rev. E 68 (2003) 036122.
[20] D.J. Watts, P.S. Dodds, M.E.J. Newman, Science 296 (2002) 1302;
See also A.E. Motter, T. Nishikawa, Y.-C. Lai, Phys. Rev. E 68 (2003) 036105;
M. Boguna, R. Pastor-Satorras, A. Diaz-Guilera, A. Arenas, (cond-mat/0309263).
[21] D.-H. Kim, B. Kahng, D. Kim, Eur. Phys. J. B 38 (2004) 305.
[22] K.-I. Goh, B. Kahng, D. Kim, (q-bio.MN/0312009).