Emergence of global preferential attachment from local interaction

Menghui Li[1,2], Liang Gao[3], Ying Fan[1], Jinshan Wu[1,4,5] and Zengru Di[1]

[1] Department of Systems Science, School of Management, and Center for Complexity Research, Beijing Normal University, Beijing 100875, People's Republic of China
[2] Temasek Laboratories, National University of Singapore, 117508, Singapore Beijing–Hong Kong–Singapore Joint Centre for Nonlinear and Complex Systems (Singapore), National University of Singapore, Kent Ridge, 119260, Singapore
[3] System Science Institute, School of Traffic and Transportation, MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, People's Republic of China
[4] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada
E-mail: jinshanw@phas.ubc.ca

Abstract. Global degree/strength-based preferential attachment is widely used as an evolution mechanism of networks. But it is hard to believe that any individual can get global information and shape the network architecture based on it. In this paper, it is found that the global preferential attachment emerges from the local interaction models, including the distance-dependent preferential attachment (DDPA) evolving model of weighted networks (Li *et al* 2006 *New J. Phys.* **8** 72), the acquaintance network model (Davidsen *et al* 2002 *Phys. Rev. Lett.* **88** 128701) and the connecting nearest-neighbor (CNN) model (Vázquez 2003 *Phys. Rev.* E **67** 056104). For the DDPA model and the CNN model, the attachment rate depends linearly on the degree or vertex strength, whereas for

[5] Author to whom any correspondence should be addressed.

the acquaintance network model, the dependence follows a sublinear power law. It implies that for the evolution of social networks, local contact could be more fundamental than the presumed global preferential attachment.

Contents

1. Introduction

Power-law distribution commonly exists in complex networks, including planned networks such as the Internet [1] and unplanned social networks, such as scientific collaboration networks [2–5], actor collaboration networks [6], peer-to-peer networks [7], mobile networks [8] and much more. Their degree distributions obey the power law, and if applicable, the vertex strength (summing the weights of links that connect to a vertex) [4, 5] and the link weight [9] also follow the power law. Many models have been proposed to capture the topological evolution of complex networks (for reviews see [10, 11]). Especially a class of models based on the idea of global preferential attachment, first proposed in the Barabási–Albert (BA) model [12], is quite successful in reproducing the power-law distributions of degree/strength [9, 12–14]. Following on from the degree-based preferential attachment in the BA model, Barrat [13] later developed a model of weighted networks with strength-based preferential attachment.

Some networks have hubs or data centers, where global information is collected for future use. For example, WWW has search engines and the Internet has routers and DNS servers. For networks with such centers, the global preferential attachment mechanism may seem reasonable. However, in a social network there is no data centers collecting and providing global information. It is impossible for every individual to know global information about the system, such as degree or strength of every individual. For social networks, the evolution mechanism based on local quantities would be more natural.

Actually, there are already several evolving models for scale-free networks based on local rules, such as the distance-dependent preferential attachment (DDPA) model [15], the acquaintance network model [16], the connecting nearest-neighbor (CNN) model [17], the random walk model [18, 19], the redirection model [20, 21], the optimization model [22, 23] and so on. For instance, in the DDPA model, a measure of closeness relation is defined locally within second neighbors and then links are built up preferentially according to the relation. The acquaintance network model evolves via people introduced to know each other by a common acquaintance [16]. In the CNN model [17], at every time step, a new vertex is

added or a potential edge within second neighbors is converted into an edge. All the above models based on local rules can also reproduce the common topological characters of complex networks.

We note that both of the two potentially conflicting types of models successfully capture the main characters of scale-free networks, but one needs global information, while the other only requires local information. Here, we suggest a wild guess to resolve the conflict and make the whole picture more consistent. We conjecture that global preferential attachment emerges from local contact-based models. In this paper, we investigate whether or not the above conjecture holds in the following local models: the DDPA model [15], the acquaintance networks model [16] and the CNN model [17].

This paper is organized as follows. In section 2, we first briefly describe the method used to check preferential attachment. In section 3, we apply this method to the DDPA model, the acquaintance network model and the CNN model and present the results obtained. In section 4, we give some concluding remarks. It turns out that our conjecture does hold in the above models.

2. Methods for measuring preferential attachment

The preferential attachment hypothesis states that the rate $\Pi(k)$ with which a vertex with k links acquires new links is a monotonically increasing function of k [12], namely

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha} = C(t)k_i^\alpha. \tag{1}$$

For the BA model, $\alpha = 1$ [12].

Jeong *et al* [24] proposed a method to check global preferential attachment from data on the evolution process. Here, we will briefly describe the method. In the evolving process, we record the order of each node and link joining the system within a relatively short time frame after the network evolves steadily in a long time. At a large enough time T_0, consider all vertices existing in the system, called T_0 vertices. Next select a time T_1 ($T_1 > T_0$ and $(T_1 - T_0) \ll T_0$), and add a new vertex at every time step between $[T_0, T_1]$. Firstly, count the number $N(k)$ of such vertices with exactly k degree in the T_0 vertex. Secondly, record all the vertices in the T_0 vertex to which the new links are attached as Ω. Lastly, count the number of vertices with exactly k degree in Ω as $A(k)$. At this condition, $\Pi(k, T_0, T_1)$ will be independent of T_0 and T_1 but depends on k only [24].

A convenient definition of the $\Pi(k, T_0, T_1)$ function could be defined as

$$\Pi(k, T_0, T_1) = \frac{\frac{A(k)}{N(k)}}{\sum_{k'} \frac{A(k')}{N(k')}}. \tag{2}$$

According to mean-field theory and equation (1), $A(k) = M \sum_i (\delta(k_i, k)k_i^\alpha / \sum_j k_j^\alpha)$, where M is the number of new links attaching to vertices Ω. So equation (2) can be written as

$$\Pi(k) = \frac{M \sum_i \delta(k_i, k)\frac{k_i^\alpha}{\sum_l k_l^\alpha}/N(k)}{\sum_{k'} \left(M \sum_j \delta(k_j, k')\frac{k_j^\alpha}{\sum_m k_m^\alpha}/N(k')\right)} = \frac{\frac{k^\alpha}{\sum_l k_l^\alpha}}{\sum_{k'} \frac{k'^\alpha}{\sum_m k_m^\alpha}} = \frac{k^\alpha}{\sum_{k'} k'^\alpha}, \tag{3}$$
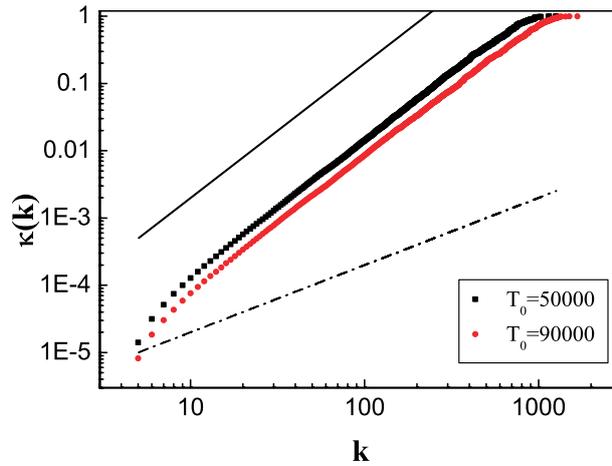
Figure 1. The $\kappa(k)$ function determined numerically for the BA model based on equation (2). The slope of the solid line is 2 (corresponding to $\alpha = 1$ as $\alpha + 1 = 2$), and the slope of the dashed line is 1 ($\alpha = 0$). This indicates that the exponent α is almost 1. If not mentioned in the following figures, we run the model for 100 times with the same parameters to count $A(k, T_0)$ and $N(k, T_0)$, and then we determine $\kappa(k)$ ($\kappa(s)$) according to equation (2) and estimate the value of α, where the power-law exponents are calculated on intermediate degree (vertex strength). We used $T_1 = T_0 + 100$ unless otherwise stated. The parameters are $n_0 = 10$ and $m = 5$.

where $\sum_i \delta(k_i, k) k_i^\alpha = k^\alpha N(k)$ is used. To avoid the effect of noise, we also study the cumulative function instead of $\Pi(k)$

$$\kappa(k) = \int_0^k \Pi(k') \, \mathrm{d}k'. \tag{4}$$

If there is a global preferential attachment, then

$$\kappa(k) \propto k^{\alpha+1}. \tag{5}$$

In order to prove the validity of formula (2), we apply this method to the BA model, starting from a fully connected n_0 initial network, where $\langle k \rangle = 2m$, $\alpha = 1$ and $\gamma = 3$. The exponent in equation (3) should be 1 and the measurement shows that they are approximately 1 (as shown in figure 1). This indicates that formula (2) is fit to measure the form of $\Pi(k)$.

In measurement, we only focus on the vertices new links are attached onto, so we only consider ending vertices. Furthermore, measurements on external links from the new vertex and internal links among existing vertices can be done separately when necessary.

3. Measuring preferential attachment of the local contact model

In this section, we check the possibility of such global attachment emerging from the following three local contact models.

3.1. The DDPA model

The evolution of networks starts from a fully connected n_0 initial network. At every time step, one new vertex is added into the network by connecting it randomly to one old vertex. Then other

l old vertices are randomly activated. Every one (denoted as vertex n) of these $1 + l$ vertices can attempt to build up m connections. The probability for every link (except the first link from the new vertex) from a vertex n to a vertex i is given by

$$\Pi_{n \to i} = \frac{L_{ni}}{\sum_{j \in \partial_n^2} L_{nj}}, \qquad (6)$$

where ∂_n^2 means neighbors of vertex n up to the second separation. L_{ni} is the similarity distance [5, 9], which means that the larger the similarity distance, the closer the relation between the two end vertices. If two vertices i and j are connected directly, $L_{ij} = w_{ij}$, where w_{ij} is the similarity weight, meaning the larger, the closer, such as the number of studies in scientific collaboration networks [2–5] and the duration of calls in mobile networks [8]. Otherwise, $L_{ij}^{(\mu)}$ is calculated as follows. If vertex $i(\mu)$ is connected to vertex $\mu(j)$ with similarity link weight $w_{i\mu}$ ($w_{\mu j}$), then $L_{ij}^{(\mu)} = \frac{1}{(1/w_{i\mu} + 1/w_{\mu j})}$. The final similarity distance between vertex i and vertex j is $L_{ij} = \max\{L_{ij}^{(\mu)}\}$, the maximum one via all vertices $\{\mu\}$. In other words, it is sufficient to know the degree of familiarity between second neighbors. In addition, the reconnection between linked vertices is allowed. When building up the connection between vertices i and j represents an event happening between the vertices, the number of occurrences of such an event may be defined as the connection count T_{ij}, which will be transferred into similarity link weight $w_{ij} = f(T_{ij})$, e.g. $w_{ij} = T_{ij}$. This DDPA mechanism does not use vertex degree or vertex strength, i.e. $s_i = \sum_j w_{ij}$, as the reference, not even locally. It only makes use of the information about local similarity distance within second neighbors, but the scale-free phenomenon does appear in the DDPA model for intermediate degrees/strengths (see figure 6 in [15]).

For the DDPA model, we need to do this measure for three kinds of links: the first external links from the new vertex, all other external links from the new vertex and the internal links. The same method can also be applied to check preferential attachment according to vertex strength. Since the ending vertex of the first external link is selected randomly, the probability of a vertex selected should be independent of its degree (k) or strength (s). In the inset of figure 2, we see that the first external link does follow flat distribution. But for all other external links, we find that $\kappa(k)$ (measuring according to degree) and $\kappa(s)$ (measuring according to strength) for intermediate degrees/strengths follow a straight line on a log–log plot, indicating $\Pi(k) \propto k^{\alpha_k}$ or $\Pi(s) \propto s^{\alpha_s}$. The curves of $\kappa(k)$ for the DDPA model are consistent with those for the BA model (figure 1). They are parallel with different T_0 (as shown in figure 2), indicating that $\Pi(k)$ and $\Pi(s)$ are independent of T_0 and T_1, and depend on k or s only. Their power-law exponents are, respectively, $\langle \alpha_k \rangle = 0.85$ and $\langle \alpha_s \rangle = 0.91$.

In the measurement of internal links, we found that global preferential attachment is also valid. The curves of $\kappa(k)$ and $\kappa(s)$ follow a straight line on a log–log plot. Their power-law exponents are, respectively, $\langle \alpha_k \rangle = 1.07$ and $\langle \alpha_s \rangle = 1.1$ (as shown in figure 3). Actually, if checking all the links together, including the first external link, the curves of $\kappa(k)$ and $\kappa(s)$ follow a straight line on a log–log plot, where exponents are, respectively, $\langle \alpha_k \rangle = 0.98$ and $\langle \alpha_s \rangle = 1.04$ when $l = 1$. The parameters have almost no effect on the exponents $\langle \alpha_k \rangle$ and $\langle \alpha_s \rangle$.

3.2. The acquaintance model

At first, a random network with N vertices, where every pair of vertices is connected with probability p_l, is generated. Then acquaintance networks evolve according to the following rules: (i) one randomly chosen person introduces two random acquaintances of his to another.
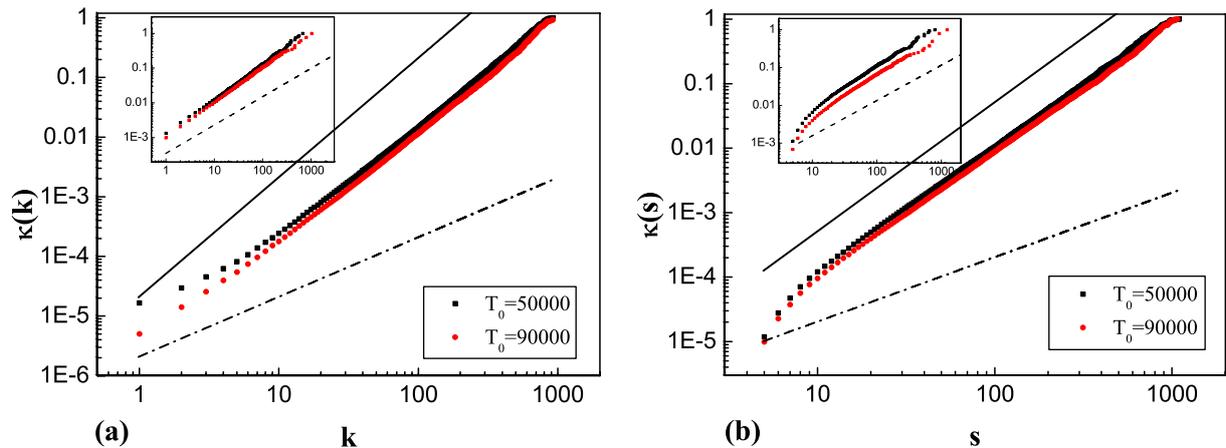
Figure 2. The $\kappa(k)$ (a) and $\kappa(s)$ (b) functions determined numerically for external links of the DDPA model. The insets are the results for the first external links, with almost flat distributions, where $\alpha_k = 0.01$ and $\alpha_s = 0.05$. $\langle \alpha_k \rangle = 0.85$ (a) and $\langle \alpha_s \rangle = 0.91$ (b) for all other external links. The parameters of the DDPA model are $n_0 = 10$ (the size of initial networks), $l = 1$ (the number of activated old vertexes) and $m = 5$ (the number of new connections on every activated vertex).
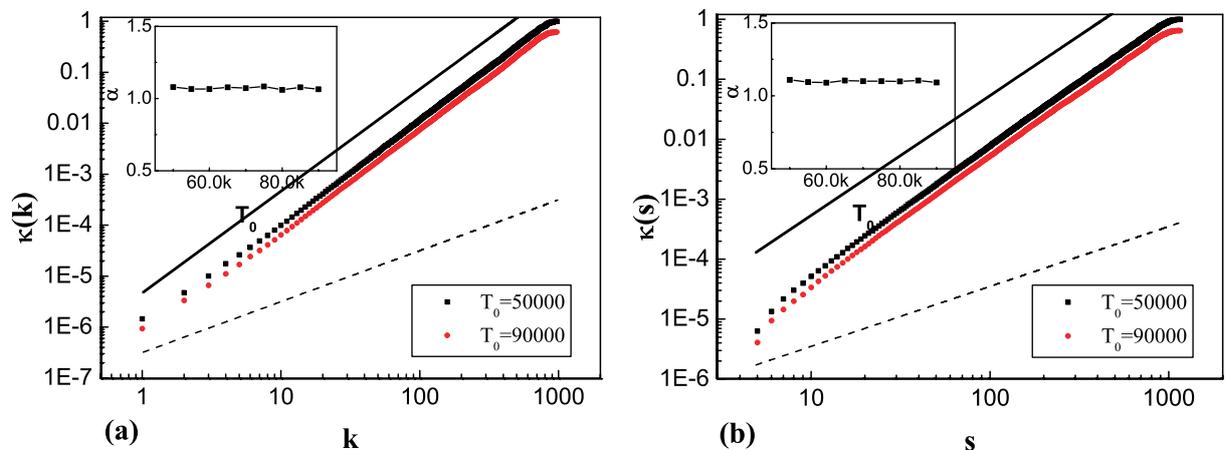

Figure 3. The $\kappa(k)$ (a) and $\kappa(s)$ (b) functions determined numerically for internal links of the DDPA model. The curves increase faster than linear (shown as the dashed lines in the figures), showing that the probability of the vertex linked by an internal link is proportional to the degree and vertex strength with $\langle \alpha_k \rangle = 1.07$ (a) and $\langle \alpha_s \rangle = 1.1$ (b). The values of $\langle \alpha \rangle$ for different choices of T_0 are inserted. It shows very small variation.

If they have not met before, a new link between them is formed. If he has less than two acquaintances, then he introduces himself to another random person. (ii) With probability p, one randomly chosen person is removed from the network, including all links connected to this vertex, and replaced by a new person with one randomly chosen acquaintance [16]. The model generates degree distributions spanning scale-free and exponential regimes (see figure 1 in [16]). Although the size of the network is fixed, this model includes the birth and death process.
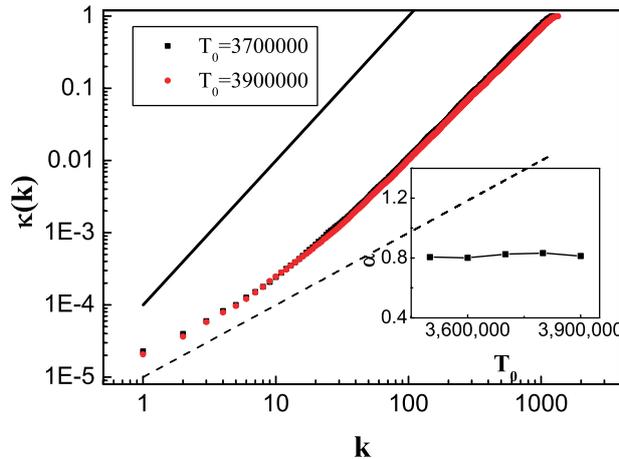
Figure 4. The $\kappa(k)$ function determined numerically for the acquaintance network model. The values of $\langle \alpha \rangle$ for different choices of T_0 are inserted. Results from fitting of α show that $\langle \alpha \rangle = 0.815$. The parameters are $N = 10\,000$, $p_l = 0.01$ and $p = 0.01$.
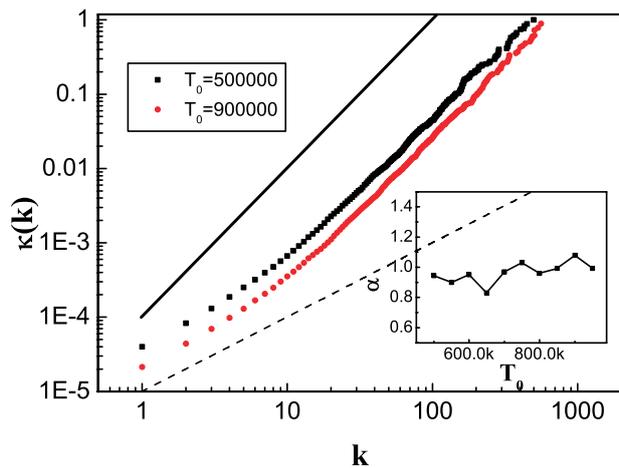

Figure 5. The $\kappa(k)$ function determined numerically for all the CNN model links. The values of $\langle \alpha \rangle$ for different choices of T_0 are inserted. For $u = 0.4$, we find that $\langle \alpha \rangle \approx 1.00$.

New links emerge between individuals frequently. After evolving for T_0 times, again we record and analyze the evolution process between T_0 and T_1. We can see that $\kappa(k)$ also follows a straight line on a log–log plot, where $\alpha \simeq 0.80$ (as shown in figure 4).

3.3. The CNN model

Next we consider a variant of the acquaintance model, the CNN model [17] with increasing network size. The CNN model starts from a single vertex and evolves according to the following rules: (i) with probability $1 - u$, a new vertex (denoted as n) is introduced into the network by connecting it randomly to one old vertex j. The pair of vertices $[n, i]$ between all nearest neighbors i of vertex j and the new vertex n is recorded in $\bar{\partial}([n, i])$. (ii) Then, with

probability u, two vertices $[n, i]$ randomly selected from $\bar{\partial}$ are connected. When the size of the network is sufficiently large, the distribution of intermediate degrees exhibits a power-law decay $P(k) \sim k^{-\gamma}$ (see figure 8 in [17]), and the value of γ depends on the parameter u.

In the measurement process, we record the degree of vertex j and the other two selected end vertices. We see that $\kappa(k)$ follows a straight line on a log–log plot, and they are parallel with different T_0 (as shown in figure 5). For different values of the parameter u, the values of exponent α are different, for example $\alpha \approx 1.11$ for $u = 0.3$ and $\alpha \approx 0.95$ for $u = 0.7$. In figure 5, we show the measurement on $u = 0.4$. There we can see that $\alpha \approx 1.00$.

Although α is close to 1 for the DDPA and CNN models and $\alpha = 0.80$ for the acquaintance model, we find that $\kappa(k) \sim k^{\alpha+1}$ holds in all the models. This indicates that although links are created by local information, it appears as if the network evolves according to global preferential attachment.

4. Concluding remarks

In this paper, we have measured global degree/strength preferential attachment of the DDPA model, the acquaintance network model and the CNN model. These models make use of local quantities instead of the global degree/strength preferential attachment mechanism. However, our measurement shows that they can still be seen as if the networks are evolving according to global preferential attachment. From this point of view, preferential attachment is an emergent phenomenon from the more fundamental local rules. The empirical study of e-mail networks in [25] also suggests that cyclic closure, especially triadic closure, plays an important role in building up new links. In modeling social networks, this observation may suggest that local rules are preferred rather than the global degree/strength preferential attachment. For people who believe in the latter, our discovery is especially meaningful. Our results indicate that one does not need to worry about the widely believed and used global preferential attachment because it can emerge from properly designed local rules.

Acknowledgments

References

[1] Pastor-Satorras R, Vázquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701
[2] Newman M E J 2001 *Phys. Rev.* E **64** 016131
   Newman M E J 2001 *Phys. Rev.* E **64** 016132
[3] Barabási A-L, Jeong H, Néda Z, Ravasza E, Schubert A and Vicsek T 2002 *Physica* A **311** 590
[4] Barrat A, Barthélemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl Acad. Sci. USA* **101** 3747
[5] Fan Y, Li M, Chen J, Gao L, Di Z and Wu J 2004 *Int. J. Mod. Phys.* B **18** 2505
   Li M, Fan Y, Chen J, Gao L, Di Z and Wu J 2005 *Physica* A **350** 643
[6] Albert R and Barabási A-L 2000 *Phys. Rev. Lett.* **85** 5234

 [7] Wang F, Moreno Y and Sun Y 2006 *Phys. Rev.* E **73** 036123
 [8] Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J and Barabási A-L 2007 *Proc. Natl Acad. Sci. USA* **104** 7332
      Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Argollo de Menezes M, Kaski K, Barabási A-L and Kertész J 2007 *New J. Phys.* **9** 179
 [9] Li M, Wu J, Wang D, Zhou T, Di Z and Fan Y 2007 *Physica* A **375** 355
[10] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[11] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D-U 2006 *Phys. Rep.* **424** 175
[12] Barabási A-L and Albert R 1999 *Science* **286** 509
[13] Barrat A, Barthélemy M and Vespignani A 2004 *Phys. Rev. Lett.* **92** 228701
[14] Wang W, Wang B, Hu B, Yan G and Ou Q 2005 *Phys. Rev. Lett.* **94** 188702
[15] Li M, Wang D, Fan Y, Di Z and Wu J 2006 *New. J. Phys.* **8** 72
[16] Davidsen J, Ebel H and Bornholdt S 2002 *Phys. Rev. Lett.* **88** 128701
[17] Vázquez A 2003 *Phys. Rev.* E **67** 056104
[18] Saramäki J and Kaski K 2004 *Physica* A **341** 80
[19] Evans T S and Saramäki J P 2005 *Phys. Rev.* E **72** 026138
[20] Krapivsky P L and Redner S 2001 *Phys. Rev.* E **63** 066123
      Krapivsky P L and Redner S 2002 *J. Phys. A: Math. Gen.* **35** 9517
[21] Rozenfeld H D and Ben–Avraham D 2004 *Phys. Rev.* E **70** 056107
[22] Pujol J M, Flache A, Delgado J and Sangüesa R 2005 *J. Artif. Soci. Soc. Simul.* **8** 12
[23] D'Souza R M, Borgs C, Chayes J T, Berger N and Kleinberg R D 2007 *Proc. Natl Acad. Sci. USA* **104** 15
[24] Jeong H, Néda Z and Barabási A-L 2003 *Europhys. Lett.* **61** 567
[25] Kossinets G and Watts D J 2006 *Science* **311** 88