Vol 18 No 3, March 2009
1674-1056/2009/18(03)/0933-06

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

Generating weighted community networks based on local events*

Xu Qi-Xin(徐琪欣) and Xu Xin-Jian(许新建)†

Department of Mathematics, College of Science, Shanghai University, Shanghai 200444, *China*

(Received 30 June 2008; revised manuscript received 22 August 2008)

Many realistic networks have community structures, namely, a network consists of groups of nodes within which links are dense but among which links are sparse. This paper proposes a growing network model based on local processes, the addition of new nodes intra-community and new links intra- or inter-community. Also, it utilizes the preferential attachment for building connections determined by nodes' strengths, which evolves dynamically during the growth of the system. The resulting network reflects the intrinsic community structure with generalized power-law distributions of nodes' degrees and strengths.

Keywords: complex networks, community networks, weighted networks
PACC: 0520, 0250

1. Introduction

Since 1998, with the small-world model introduced by Watts and Strogatz,[1] a new science of networks has appeared on the stage. Because of its powerful function of describing structures and properties of many real systems,[2] the complex network is one of the most popular fields studied nowadays. In the past decade, there are two characteristics studied mostly in complex networks. One is the small-world property,[1] which suggests that a network has a highly degree of clustering like regular networks and a small average distance among any two nodes similar to random networks. The small-world phenomenon has been successfully described by network models with some degree of randomness.[1,3] The other is the scale-free behaviour,[4] which means a power-law distribution of connectivity, $P(k) \sim k^{-\gamma}$, where $P(k)$ is the probability that a node in the network has k connections to other nodes and γ is a positive real number determined by the given network. The origin of the scale-free behaviour has been traced back to two mechanisms observed in many systems, which are growing and preference.[4,5]

Recently, with the progress of research in complex networks, many other statistical characteristics of networks have been reported in the current literature. The so-called 'community' (or 'modularity') is of great interest. That is to say, a network is composed of many clusters of nodes, where the nodes in the same cluster are highly connected, while there are few links among the nodes belonging to different clusters. For instance, groups are formed in scientific collaboration networks.[6] Also, the presence of community can dramatically alter the behaviour of dynamical processes on networks.[7−9]

In the study of community networks, most research has been directed in two distinct directions. On the one hand, attention has been paid to designing algorithms for detecting community structures in real networks. A pioneering method was made by Girvan and Newman,[6] they introduced a quantitative measure for the quality of a partition of a network into communities. Later, a number of algorithms have been proposed in order to find a good optimization with the least computational cost. The fastest available procedures use greedy techniques[10] and extremal optimization,[11] which are capable of detecting communities in large networks. On the other hand, research has focused on modelling of networks with community structures. In Ref.[12], a static social network was introduced where individuals belong to groups which in turn belong to groups of groups and so on. In Ref.[13], a networked seceder model was suggested

*Project supported by Institute of Systems Biology, the Innovation Foundation of Shanghai University of Shanghai University of China and the National Natural Science Foundation of China (Grant No 10805033).
†Corresponding author. E-mail: xinjxu@shu.edu.cn

to illustrate group formation in social networks. In Ref.[14], a growing bipartite network for social communities with group structures was proposed. Each of those models is constructed based on one aspect of reality.

Apart from the topological issue, it is important to realize that many complex networks are weighted, i.e., the interaction between two nodes is characterized not only by the existence of a link but a link with a varying weight assigned to it. Weighted networks can be described by a matrix w_{ij} specifying the weight on the edge connecting the vertices i and j, with $i, j = 1, \ldots, N$, where N is the size of the network ($w_{ij} = 0$ if the vertices i and j are not connected). The strength of the vertex i can be defined as $s_i = \sum_{j \in U(i)} w_{ij}$, where the sum runs over the set $U(i)$ of neighbours of i.[15] The strength of a node integrates the information both about its connectivity and the importance of the weights of its links, and can be considered as the natural generalization of the degree.

In this paper, we propose a weighted network model with communities that gives a realistic description of local events.[16,17] The model incorporates three processes, the addition of new nodes intra-community and new links intra- or inter-community. Using growing and preferential attachment mechanisms, we generate the community network with good right-skewed distributions of nodes' degrees and strengths, which have been observed in many social systems.

2. Model

The Barabási–Albert network[4] only describes a particular type of evolving networks, the addition of new nodes preferentially connecting to the nodes already presented in the network. Systems in the real world, however, are much richer. For example, in scientific collaboration networks, a multidisciplinary scientist is not only collaborate with scientists in his research fields but also has a stronger desire to collaborate with scientists in other fields.[18] In friendship networks, a person usually makes friends with people belonging to different communities besides the community he belongs to. To give a realistic description of the network construction like that, we introduce a growing model of community networks based on local events, the addition of new nodes intra-community and new links intra- or inter-community. The proposed model is defined as follows.

We start with M (≥ 2) isolated communities and each community consists of a small number n of connected nodes, with a same assigned weight w_0 to each link. At each time step, we perform one of the following three operations.

(i) With probability p we add a new node l in a randomly chosen community. The new node is only connected to one node which already presents in the selected community (a weight w_0 is assigned to this link). We denote it as the uth community. The probability that node i in community u will be selected is subjected to the following equation

$$\prod(s_i^{\text{intra}}) = \frac{s_{u,i}^{\text{intra}}}{\sum_j s_{u,j}^{\text{intra}}}, \tag{1}$$

where the sum runs over nodes in community u and $s_{u,i}^{\text{intra}}$ is the intra-community strength of node i in community u. Following ideas developed by Barrat *et al*,[19] once node i is selected to be connected, the additional induced increase δ in strength s_i^{intra} is distributed among its nearest neighbours in community u according to the rule

$$w_{ij}^{\text{intra}} \to w_{ij}^{\text{intra}} + \delta \frac{w_{ij}^{\text{intra}}}{s_i^{\text{intra}}}. \tag{2}$$

(ii) With probability q we add a new link (with an assigned weight w_0) in a randomly chosen community. For this we randomly select a node in a randomly chosen community u as the starting point of the new link. The other end of the link is selected in the same community with the probability given by Eq.(1) and the additional induced increase δ in strength s_i^{intra} is distributed among its nearest neighbours in community u (described by Eq.(2)).

(iii) With probability r ($= 1 - p - q$) we add an new link (with an assigned weight w_0) between two communities. For this we randomly select a node in a randomly chosen community u as the starting point of the new link. The other end i of the link selected in the other community v is proportional to its inter-community strength

$$\prod(s_i^{\text{inter}}) = \frac{s_{v,i}^{\text{inter}}}{\sum_{v \neq u;j}(s_{v,j}^{\text{inter}})}, \tag{3}$$

where the sum runs over nodes in all communities except for community u and $s_{v,i}^{\text{inter}}$ is the inter-community strength of node i in community v. For the end point of the new link, the additional induced increase δ in

strength $s_i^{\rm inter}$ is distributed among its nearest neighbours belonging to other communities

$$w_{ij}^{\rm inter} \to w_{ij}^{\rm inter} + \delta \frac{w_{ij}^{\rm inter}}{s_i^{\rm inter}}. \tag{4}$$

After t time steps, this scheme generates a network of $Mn + pt$ nodes and t links. The parameters p, q, and r control the network structure. In the case of small r, the generated network will have a strong community structure. Notice that whatever process is chosen at each time step, only one link is added to the system at each time step (duplicate and self-connected edges are forbidden), however, this is not essential.

3. Strength distribution

In our community network, the strength of a node consists of two parts, the intra-community strength and the inter-community strength. Increases of the node's strength can be divided into two processes, the increases of the intra-community strength and the inter-community strength. For the sake of simplicity, we can set $w_0 = 1$ and the weight depends only on the dimensionless parameter δ. Based on the continuous approximation,[2,19] we assume that $s_i^{\rm intra}$ and $s_i^{\rm inter}$ change continuously, and the probabilities $\prod(s_i^{\rm intra})$ and $\prod(s_i^{\rm inter})$ can be interpreted as the rates at which $s_i^{\rm intra}$ and $s_i^{\rm inter}$ change, respectively. Thus, the operations (i)–(iii) all contribute to s_i, each being incorporated in the continuum theory as follows.

(i) Addition of a new node in a randomly chosen community with probability p

$$\frac{\partial s_{u,i}^{\rm intra}}{\partial t} = p\frac{1}{M}\left[\frac{s_{u,i}^{\rm intra}}{\sum_j s_{u,j}^{\rm intra}}(1+\delta) + \sum_{j\in U^{\rm intra}} \frac{s_{u,j}^{\rm intra}}{\sum_l s_{u,l}^{\rm intra}}\delta\frac{w_{ij}^{\rm intra}}{s_{u,j}^{\rm intra}}\right]. \tag{5}$$

(ii) Addition of a new link in a randomly chosen community with probability q

$$\frac{\partial s_{u,i}^{\rm intra}}{\partial t} = q\left[\frac{1}{N} + \frac{s_{u,i}^{\rm intra}}{\sum_j s_{u,j}^{\rm intra}}(1+\delta) + \sum_{j\in U^{\rm intra}} \frac{s_{u,j}^{\rm intra}}{\sum_l s_{u,l}^{\rm intra}}\delta\frac{w_{ij}^{\rm intra}}{s_{u,j}^{\rm intra}}\right]. \tag{6}$$

(iii) Addition of a new links between two communities with probability r

$$\frac{\partial s_{v,i}^{\rm inter}}{\partial t} = r\left[\frac{1}{N} + \left(1-\frac{1}{M}\right)\frac{s_{v,i}^{\rm inter}}{\sum_j s_{v,j}^{\rm inter}}(1+\delta) + \left(1-\frac{1}{M}\right)\sum_{j\in U^{\rm inter}} \frac{s_{v,j}^{\rm inter}}{\sum_l s_{v,l}^{\rm inter}}\delta\frac{w_{ij}^{\rm inter}}{s_{v,j}^{\rm inter}}\right]. \tag{7}$$

Combing the contribution of processes (i) and (ii), we have

$$\frac{\partial s_{u,i}^{\rm intra}}{\partial t} = \frac{p+q}{M}\left[\frac{s_{u,i}^{\rm intra}}{\sum_j s_{u,j}^{\rm intra}}(1+\delta) + \sum_{j\in U^{\rm intra}} \frac{s_{u,j}^{\rm intra}}{\sum_l s_{u,l}^{\rm intra}}\delta\frac{w_{ij}^{\rm intra}}{s_{u,j}^{\rm intra}}\right] + \frac{q}{N}. \tag{8}$$

Notice that $\sum_j s_{u,j}^{\rm intra} = \dfrac{2}{M}(p+q)(1+\delta)t$. Thus we have

$$\frac{\partial s_{u,i}^{\rm intra}}{\partial t} = \frac{s_{u,i}^{\rm intra}}{2(1+\delta)t} + \frac{q}{Mn+pt}, \tag{9}$$

which can be simplified for large t

$$\frac{\partial s_{u,i}^{\rm intra}}{\partial t} \approx \frac{s_{u,i}^{\rm intra}}{2(1+\delta)t} + \frac{q}{pt}. \tag{10}$$

The boundary condition of the intra-community strength at initial time t_s can be estimated in the sense of mathematical expectation, $s_{u,i}^{\rm intra}(t_s) = p + q$. So we write the solution of Eq.(10) as

$$s_{u,i}^{\rm intra} = \frac{[p(p+q)(1+2\delta) + 2q(1+\delta)]\left(\dfrac{t}{t_s}\right)^{\frac{1+2\delta}{2+2\delta}} - 2q(1+\delta)}{p(1+2\delta)}. \tag{11}$$

In random networks, the strength distribution can be calculated by

$$P(s) = \frac{1}{t}\sum_{i=1}^{t}\delta(s_i(t) - s), \tag{12}$$

which gives

$$P(s^{\rm intra}) = \frac{2p(1+\delta)}{p(p+q)(1+2\delta)+2q(1+\delta)}\left[\frac{2q(1+\delta)+p(1+2\delta)s^{\rm intra}}{p(p+q)(1+2\delta)+2q(1+\delta)}\right]^{-\frac{3+4\delta}{1+2\delta}}. \tag{13}$$

For Eq.(7), we notice $\sum_l s_{v,l}^{\rm inter} = 2(1+\delta)r\left(1-\frac{1}{M}\right)t$ and consequently simplify it for large t

$$\frac{\partial s_{v,i}^{\rm inter}}{\partial t} \approx \frac{2(1+\delta)r + p(1+2\delta)s_{v,i}^{\rm inter}}{2(1+\delta)pt}. \tag{14}$$

Using the boundary condition of the inter-community strength at initial time t_s which can be estimated in the sense of mathematical expectation, $s_{v,i}^{\rm inter}(t_s) = r$, we write the solution of Eq.(14) as

$$s_{v,i}^{\rm inter} = \frac{r[p(1+2\delta)+2(1+\delta)]\left(\frac{t}{t_s}\right)^{\frac{1+2\delta}{2(1+\delta)}} - 2r(1+\delta)}{p(1+2\delta)}. \tag{15}$$

According to Eq.(12), we obtain the distribution of the inter-community strength

$$\begin{aligned} &P(s^{\rm inter}) \\ &= \frac{2p(1+\delta)}{pr(1+2\delta)+2r(1+\delta)} \\ &\times\left[\frac{2r(1+\delta)+p(1+2\delta)s^{\rm inter}}{pr(1+2\delta)+2r(1+\delta)}\right]^{-\frac{3+4\delta}{1+2\delta}}. \end{aligned} \tag{16}$$

Thus, the strength distribution of our network should have the form of a generalized power-law

$$P(s) \sim [A(p,q) + B(p,q)s]^{-\gamma(\delta)}. \tag{17}$$

In Fig.1 we present numerical results of distributions of the intra-community strength, the inter-community strength, and the total strength of nodes in log–log scale. The distributions of the intra-community strength and the inter-community strength, shown in Figs.1(a) and 1(b), agree with analytical results of Eqs.(13) and (16), respectively. According to the evolving rule of our network, nodes with larger intra- (or inter-) strength have higher probability to gain new links, then the usual strength preferential attachment is reasonably kept. This means that the right-skewed character of the network, such as the node's total strength, will retain. As shown in Fig.1(c), the total strength distribution of nodes is well expected to show a good right-skewed character, which is reasonably in agreement with the condition of many realistic systems.[20]

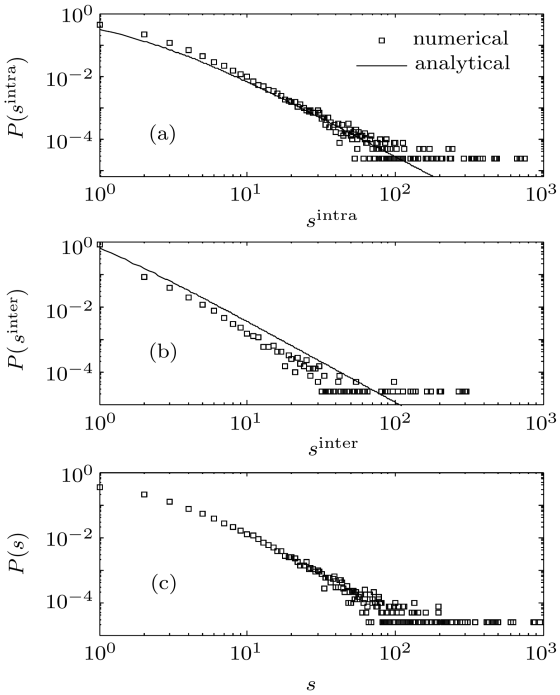


Fig.1. Log–log representation of distributions of the intra-community strength (a), the inter-community strength (b), and the total strength (c) of nodes. All the simulation results (squares) display good right-skewed distributions. Solid lines in (a) and (b) denote analytical results predicted by Eqs.(13) and (16), respectively. The experiment network has a total number of nodes $N = 10^5$ with parameters $M = 10$, $n = 5$, $p = 0.4$, $q = 0.4$, $w_0 = 1$, and $\delta = 0.5$.

4. Degree distribution

According to the evolving rule of our model, one can easily get the relation

$$\frac{\partial k_{u,i}^{\text{intra}}}{\partial t} = \frac{p+q}{M}\frac{s_{u,i}^{\text{intra}}}{\sum_j s_{u,j}^{\text{intra}}} + \frac{q}{N}, \tag{18}$$

$$\frac{\partial k_{v,i}^{\text{inter}}}{\partial t} = \frac{r}{N} + r\left(1-\frac{1}{M}\right)\frac{s_{v,i}^{\text{inter}}}{\sum_j s_{v,j}^{\text{inter}}}, \tag{19}$$

which can be simplified for large t

$$\frac{\partial k_{u,i}^{\text{intra}}}{\partial t} \approx \frac{p s_{u,i}^{\text{intra}} + 2q(1+\delta)}{2p(1+\delta)t}, \tag{20}$$

$$\frac{\partial k_{v,i}^{\text{inter}}}{\partial t} \approx \frac{p s_{v,i}^{\text{inter}} + 2r(1+\delta)}{2p(1+\delta)t}. \tag{21}$$

Substituting Eqs.(11) and (15) into Eqs.(20) and (21), respectively, we have

$$\frac{\partial k_{u,i}^{\text{intra}}}{\partial t} = \frac{[p(p+q)(1+2\delta)+2q(1+\delta)]\left(\frac{t}{t_s}\right)^{\frac{1+2\delta}{2+2\delta}} + 4q(1+\delta)}{2p(1+\delta)(1+2\delta)t}, \tag{22}$$

$$\frac{\partial k_{v,i}^{\text{inter}}}{\partial t} = \frac{[pr(1+2\delta)+2r(1+\delta)]\left(\frac{t}{t_s}\right)^{\frac{1+2\delta}{2+2\delta}} + 4\delta r(1+\delta)}{2p(1+\delta)(1+2\delta)t}. \tag{23}$$

The boundary conditions of the intra-community degree and the inter-community degree at initial time t_s can be estimated in the sense of mathematical expectations, $k_{u,i}^{\text{intra}}(t_s) = p+q$ and $k_{v,i}^{\text{inter}}(t_s) = r$, respectively. So we write solutions of Eqs.(22) and (23) as

$$k_{u,i}^{\text{intra}} = \frac{2+2\delta}{1+2\delta}\left[\frac{p+q}{2(1+\delta)} + \frac{q}{p(1+2\delta)}\right]\left[\left(\frac{t}{t_s}\right)^{\frac{1+2\delta}{2+2\delta}} - 1\right] + \frac{2q\delta}{p(1+2\delta)}\ln\left(\frac{t}{t_s}\right) + q + p, \tag{24}$$

$$k_{v,i}^{\text{inter}} = \frac{2+2\delta}{1+\delta}\left[\frac{r}{2(1+\delta)} + \frac{r}{p(1+2\delta)}\right]\left[\left(\frac{t}{t_s}\right)^{\frac{1+2\delta}{2+2\delta}} - 1\right] + \frac{2r\delta}{p(1+2\delta)}\ln\left(\frac{t}{t_s}\right) + r. \tag{25}$$

In random networks, the degree distribution can be calculated by

$$P(k) = \frac{1}{t}\sum_{i=1}^{t}\delta(k_i(t)-k), \tag{26}$$

which gives

$$P(k^{\text{intra}}) = \frac{2p(1+\delta)(1+2\delta)}{p(p+q)(1+2\delta)+2(1+q)(1+\delta)} \times \left[\frac{(p^2+pq+2q)(1+\delta)-2p(p+q)(1+2\delta)^2+p(1+2\delta)^2k^{\text{intra}}}{p(p+q)(1+2\delta)+(2pq+4q\delta)(1+\delta)}\right]^{-\frac{3+4\delta}{1+2\delta}}, \tag{27}$$

$$P(k^{\text{inter}}) = \frac{4p(1+\delta)^2(1+2\delta)}{r[p(1+2\delta)^2+2(1+\delta)(1+6\delta)]} \times \left[\frac{r(1+2\delta)(2(1+\delta)+p(1+2\delta)^2)+2p(1+\delta)(1+2\delta)^2k^{\text{inter}}}{r(p(1+2\delta)^2+2(1+\delta)(1+6\delta+4\delta^2))}\right]^{-\frac{3+4\delta}{1+2\delta}}. \tag{28}$$

Thus, the degree distribution of our network also has the form of a generalized power-law

$$P(k) \sim [A(p,q)+B(p,q)k]^{-\gamma(\delta)}. \tag{29}$$

In Fig.2 we present numerical results of distributions of the intra-community degree, the inter-community degree, and the total degree of nodes in log–log scale. The experimental network is generated by the proposed scheme with the same parameter values as those in Fig.1. The distributions of the intra-community degree and the inter-community degree, shown in Figs.2(a) and 2(b), agree with analytical

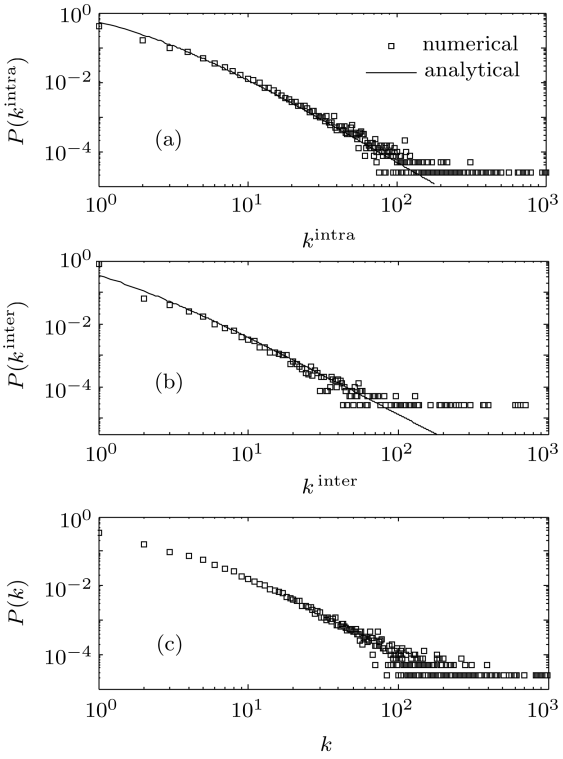


Fig.2. Log–log representation of distributions of the intra-community degree (a), the inter-community degree (b), and the total degree (c) of nodes. All the simulation results (squares) display good right-skewed distributions. Solid lines in (a) and (b) denote analytical results predicted by Eqs.(27) and (28), respectively. Parameters are the same as in Fig.1.

results of Eqs.(27) and (28), respectively. The small deviations between computer simulations and analytical solutions appear to be the mathematical approximation of the boundary conditions and the finite size effect due to the relatively small network sizes used in the simulations. In Fig.1(c), the total degree distribution of nodes also shows a good right-skewed character.

5. Conclusion

Weighted networks with community structures underlie many natural and artificial systems. It is becoming essential to model and study this kind of topological feature. In present work, we suggested a simplified mechanism for networks organized in communities, which corresponds to local events during the system growth. The generated weighted network is highly clustered and has good right-skewed distributions of nodes' strengths and degrees, which have been found very common in most realistic systems. The model we have explored, however, is a simple one in the class of weighted community networks. There exists a series of improvements that can be made further. For instance, rather than remaining a constant, δ is a variable for different vertices depended on their degrees or strengths. This may have a great influence on the structure of community networks.

References

[1] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[2] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[3] Newman M E J and Watts D J 1999 *Phys. Rev.* E **60** 7332
[4] Barabási A L and Albert R 1999 *Science* **286** 509
[5] Krapivsky P L, Redner S and Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629
[6] Girvan M and Newman M E J 2002 *Proc. Natl. Acad. Sci.* USA **99** 7821
[7] Liu Z and Hu B 2005 *Europhys. Lett.* **72** 315
[8] Yan G, Fu Z Q, Ren J and Wang W X 2007 *Phys. Rev.* E **75** 016108
[9] Feng C F, Xu X J, Wu Z X and Wang Y H 2008 *Chin. Phys.* B **17** 1951
[10] Newman M E J 2004 *Phys. Rev.* E **69** 066133
[11] Duch J and Arenas A 2005 *Phys. Rev.* E **72** 027104
[12] Watts D J, Dodds P S and Newman M E J 2002 *Science* **296** 1302
[13] Grönlund A and Holme P 2004 *Phys. Rev.* E **70** 036108
[14] Noh J D, Jeong H C, Ahn Y Y and Jeong H 2005 *Phys. Rev.* E **71** 036131
[15] Yook S H, Jeong H, Barabási A L and Tu Y 2001 *Phys. Rev. Lett.* **86** 5835
[16] Albert R and Barabási A L 2000 *Phys. Rev. Lett.* **85** 5234
[17] Dorogovtsev S N and Mendes J F F 2000 *Europhys. Lett.* **52** 33
[18] Li C and Maini P K 2005 *J. Phys.* A: *Math. Gen.* **38** 9741
[19] Barrat A, Barthélemy M and Vespignani A 2004 *Phys. Rev. Lett.* **92** 228701
[20] Newman M E J 2001 *Proc. Natl. Acad. Sci.* USA **98** 404
Barrat A, Barthélemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl. Acad. Sci.* USA **101** 3747