Modelling weighted networks using connection count

Menghui Li[1]**, Dahui Wang**[1]**, Ying Fan**[1]**, Zengru Di**[1]
and Jinshan Wu[1,2,3]

[1] Department of Systems Science, School of Management,
Beijing Normal University, Beijing 100875, People's Republic of China
[2] Department of Physics & Astronomy, University of British Columbia,
Vancouver, BC V6T 1Z1, Canada
E-mail: jinshanw@physics.ubc.ca

Abstract. Weight-driven and degree-driven evolutionary models for weighted networks are proposed, with a distance-dependent mechanism to increase the clustering coefficient. Compared with the well-known BA model there are two generalizations. Firstly, introducing connection count and converting it into edge weight, both new and old vertices can attempt to build up connections. When a new edge is inserted between already connected nodes, the connection count (and thereby the weight) of the corresponding link is increased. Secondly, the distribution of local path distance is also used as a reference in the preferential attachment. The models show some interesting results including scale-free distributions on degree, vertex weight and edge weight. Even the pure distance-dependent preferential attachment model shows all these typical behaviours and a high cluster coefficient. The concepts of reconnecting and converting count as weight are not new and have been used in empirical studies of weighted networks but are first used to model weighted networks.

---

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

Contents

1. Introduction

Recently more and more empirical and modelling studies on weighted networks have been appearing. In a binary network, the edges only represent the presence or absence of interactions, while in a weighted network the weight on the edges provides a natural way to take into account the interaction strength between vertices. For example, the number of passengers or the number of flights between any two airports in airport networks [1]–[3], the closeness of any two scientists in scientific collaboration networks [4, 5], or the reaction rates in metabolic networks [6] are all crucial in the characterization of the corresponding networks.

Empirical research reveals that usually a weighted network exhibits power-law distributions for the degree, the vertex weight and the link weight, and a high clustering coefficient. Also, some of the empirical study show linear and nonlinear correlation between vertex weight and degree, $w = k^{-\eta}$ [2]. Besides the investigation on the statistical properties, some recent advances have been made to study the evolution of weighted networks. There are several ways to assign the weight to the link in these models. Some previous approaches have transferred some quantities from non-weighted networks into the weight of the edges [7]. They are usually related to the degree or other intrinsic quantities of the nodes. In some models, weights on edges are generated from some *a priori* distribution [8]–[10]. The link weight in all the above models does not change with the evolution of the networks. Recently Barrat *et al* has proposed a weight-driven model [11] with evolving weights. In [12], Bianconi has also presented a model with evolving weights by introducing an additional quantity, the fitness of vertices and edges.

Although the models [11, 12] mentioned above couple the weight with network evolution, they both require some extra mechanism for the evolution of edge weights, for example, a Δ increase and its distribution over linked edges in [11]. From the empirical study, however, connection count, defined as the number of occurrences of events represented by the connection between two vertices, is a natural measurement of the relation between vertices. The dynamical process of weight in Barrat's model [11] justified from the background of airport networks, evolves independently on connection count. However, quite a lot of weighted networks are projections of bipartite networks, such as collaboration networks of actors or scientists and networks of emails. As we can see there, weight usually relates closely to this connection count [2, 4, 5]. In our empirical investigation on scientific collaboration networks the connection count is converted into weight, and it gives a good measure of the relation between scientists. Although different authors may use different expressions [2, 4, 5] to convert connection count into weight, every one is a projection from a bipartite (scientists and papers) graph onto a one-mode

weighted network (of scientists only) [13]. For example, the connection between two scientists via different papers in the original bipartite network can be naturally regarded as a repeated connection between them, therefore it gives us connection counts. From this point of view, for the one-mode network projected from a bipartite network, the idea of converting connection count into weight seems quite general. Furthermore, the models constructed in this way do not require any extra dynamical process of weight.

For binary networks, the preferential attachment introduced in the BA model [14] is believed to be a good mechanism to generate a power-law distribution of degree. It is a natural and simple way to incorporate the above relation between connection count and weight into modelling weighted networks. In order to realize this, we add two new ingredients into the degree-driven BA model. Firstly, there is activation of old vertices, which has been used for example in [17, 20, 21]. Secondly, we can convert connection count into weight, which has already been used in empirical study [2, 4, 5], but is for the first time applied here on network models. Furthermore, if it gives good results in the degree-driven model, it will be also interesting to apply this idea to the weight-driven model. These two models compose the first part of this work. In the second part, we also introduce a distance-dependent preferential attachment (DDPA) mechanism, which helps to increase the clustering coefficient of the networks. The original BA model and Barrat's model result in quite low clustering coefficient. Some evolutionary models do give high clustering coefficient but have additional structures or quantities [15]–[19]. In our DDPA model, new connections are added into the network according to the distribution of local path distances. In the final part, we will show that a pure DDPA model with second-nearest neighbours gives all those scale-free distributions together with a high clustering coefficient.

2. The model

An N-vertex weighted network is defined by an $N \times N$ matrix, $(w_{ij})_{N \times N}$, which represents the weight on the link between vertices i and j. Weight used here is in the sense of similarity, which means the larger the weight, the closer the two ends nodes are, for example, $w_{lm} = 0$ indicates no relation between vertex l and m. When the link between vertices i, 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}. Therefore such a weighted network can also be characterized by an $N \times N$ matrix $(T_{ij})_{N \times N}$. The link weight w_{ij} is related to the connection count T_{ij} between vertex i and j, by

$$w_{ij} = f(T_{ij}), \tag{1}$$

such as the tanh function $w_{ij} = \tanh(\alpha T_{ij})$ we used in [5], or just the linear relation $w_{ij} = T_{ij}$ used in [2, 4]. The weight defined in this way is a similarity weight, which means the larger the weight the closer the relation between the vertices. Based on this similarity weight, similarity distance will be used to measure the distance between vertices in the sense of similarity. This is different from the usual meaning of distance in a binary network. For example the calculation of such a distance between two vertices 1, 3 along a path connected by two edges (with w_{12} and w_{23} respectively) is $l_{13} = 1/[(1/w_{12}) + (1/w_{23})]$, which is smaller than both w_{12} and w_{23} (see also [5]). Then to find the shortest path is actually to find the path with maximum similarity distance.

The evolution is given as follows. Starting from a fully connected n_0 initial network, with initial times $T_{ij} = 1$ (and initial weight $w_{ij} = f(1)$), at every time step,

1. One new vertex is added into this network, and l old vertices are randomly chosen from the existing network.

2. Every one (denoted as vertex n) of these $(l + 1)$ vertices will build up m connections. The probability for every link from n connecting onto vertex i is given by

$$\prod_{n \to i} = (1 - p) \frac{k_i}{\sum_j k_j} + (p - \delta) \frac{w_i}{\sum_j w_j} + \delta \frac{l_{ni}}{\sum_{j \in \partial_n^d} l_{nj}}, \tag{2}$$

where k_i is the degree of vertex i, $w_i = \sum_j w_{ji}$ is the 'onto' vertex weight of vertex i, l_{ni} is the similarity distance from n to i, and ∂_n^d is the set of nodes within the dth neighbours of vertex n. For example, ∂_n^1 is the set of nearest neighbours, ∂_n^2 includes both the nearest and the second-nearest neighbours, and so on. The first two are degree-based and weight-based preferential attachment as usual. The last term means when the old vertices are activated, they will prefer to connect their neighbours rather than remote vertices.

3. After an end node i^* is chosen from all vertices over the existing network by the probability above in equation (2), the connection count between vertex n and i^* increases by one,

$$T_{ni^*}(t + 1) = T_{ni^*}(t) + 1. \tag{3}$$

4. The weight of the edges changes as

$$w_{ni^*}(t + 1) = f(T_{ni^*}(t + 1)). \tag{4}$$

Although our general model defined above can even be applied onto directed networks, we assume that $w_{ij} = w_{ji}$ in the following analysis. A general form of $f(T_{ij})$ can be used to convert connection count into weight. We mostly consider the special case

$$w_{ij} = T_{ij}, \tag{5}$$

and in one simulation, we use the tanh function,

$$w_{ij} = \tanh(\alpha T_{ij}). \tag{6}$$

3. Pure weight-driven and pure degree-driven models

Let us first consider the pure weight-driven case with $p = 1$ and $\delta = 0$. In some sense, this means the scientists choose their cooperators according to the weight, instead of focusing on degree. In fact, the same idea of this weight-driven mechanism was used in Barrat's paper [11]. The only difference between Barrat's model and this one is the evolution of weight. In Barrat's model, an increment of weight distributes among the neighbours of the chosen vertex. In our model, it evolves through the connection count. Analytical results for the vertex weight distribution can be calculated for this case if we use the linear relation between weight and connection count, equation (5). Vertex weight or strength is

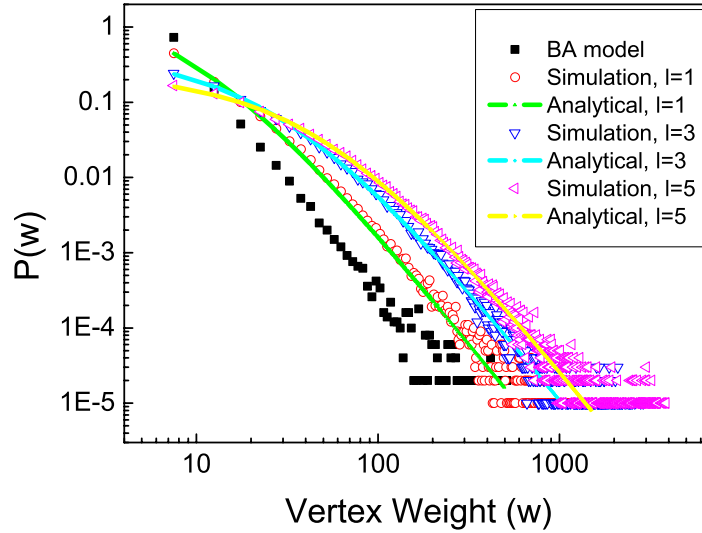$$w_i = \sum_j w_{ji} = \sum_j \alpha T_{ji}. \tag{7}$$

Figure 1. For a pure weight-driven model, analytical results of equation (10) are compared with the computer simulations under different ls. The dashed line is from the BA model, where $\gamma = 3$. Later, unless mentioned otherwise, all the simulation results are the average of 10 runs under the parameters: $m = 5$, $n_0 = 10$, $l = 5$, $N = 10\,000$.

The rate equation for the evolution of the average number of vertices with weight w at time t is

$$N(w, t + 1) = m(1 + l)\frac{(w - 1)N(w - 1, t) - wN(w, t)}{\sum_w wN(w, t)} - \frac{lN(w, t)}{N}$$
$$+ \frac{lN(w - m, t)}{N} + N(w, t) + \delta_{w,m}, \tag{8}$$

where $\sum_w wN(w, t - 1) = 2E_0 + 2m(1 + l)t$ is the total weight and $N = n_0 + t$ is the size of system at time t with initial E_0 edges and n_0 nodes. The first term reflects the preferential attachment to select the other end of the link, while the following two terms correspond to the random selection of l old vertices. When E_0 and n_0 are much smaller than t the size of the network N is approximately the time step t. Then the rate equation (8) can be written as

$$(t + 1)p(w, t + 1) = tp(w, t) + \tfrac{1}{2}[(w - 1)p(w - 1, t) - wp(w, t)]$$
$$- lp(w, t) + lp(w - m, t) + \delta_{w,m}, \tag{9}$$

where $p(w, t) \simeq N(w, t)/t$ is the density of vertices with strength w at time t. When t is large enough, we get the asymptotic distribution of vertex weight,

$$p(w) \propto (w + 2lm)^{-3}. \tag{10}$$

Comparison of the analytical results of equation (10) with computer simulations in figure 1 shows they are very consistent with each other. We can find that the lower end of vertex weight is obviously affected by the parameter l, and departs from a power-law, while the upper end is still distributed as a power-law. From simulations of pure weight-driven model, figure 2(a) shows the typical behaviour of degree and vertex weight distributions. The distribution of link weight also obeys a power law as shown in figure 3. A logarithmic binning will smooth the noisy tail coming from finite size simulation; however such linear frequency counting plots are good enough to see
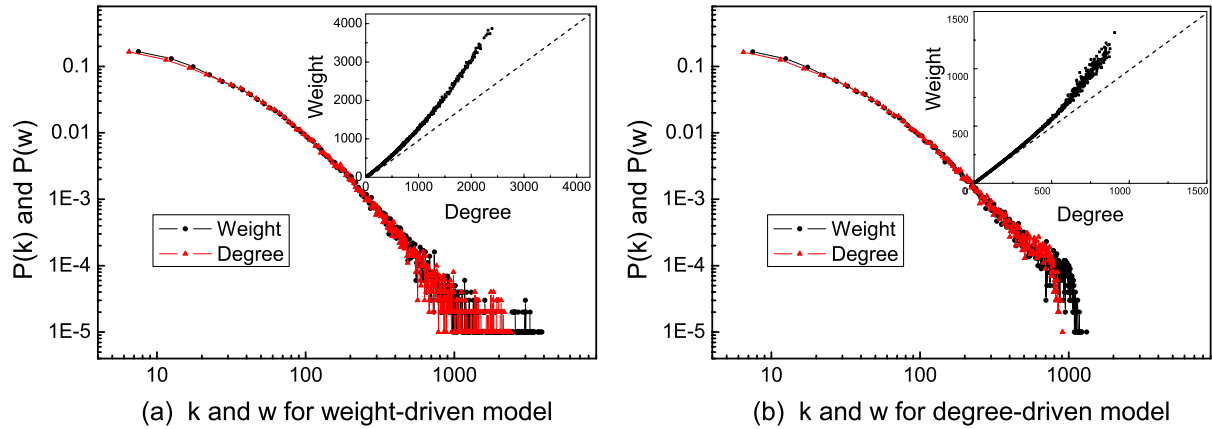
Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

(a) k and w for weight-driven model

(b) k and w for degree-driven model

Figure 2. Degree and vertex weight distributions from (a) the pure weight-driven model and (b) the pure degree-driven model. Both show power-law distributions with almost the same exponent. The insets are the correlation plots between vertex weight and vertex degree, with $w \sim k^\eta$, $\eta = 1.22$ (a) and $\eta = 1.19$ (b).
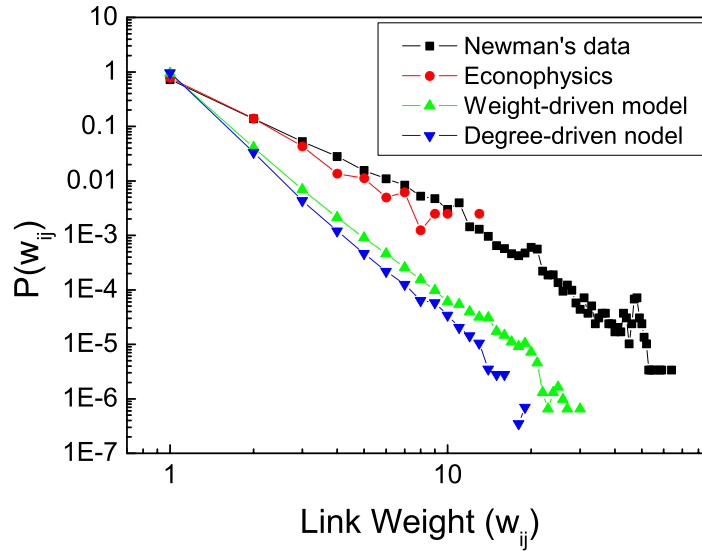

Figure 3. The link weight distribution from the pure weight-driven model and pure degree-driven model, compared with the empirical results of Newman's data and econophysics network.

the general distribution form. From the inset diagrams of the figures, we can see there is a very strong correlation between vertex weight and vertex degree, but not linear. It can be described by $w \sim k^\eta$, as we did in the insets of the figures. This is an indication that the information stored in those weighted networks cannot be fully covered by the corresponding binary networks. Such correlation has been observed by empirical works, for example $\eta = 1.5$ for a world-wide airport network in [2]. For the correlation analysis, an average is taken for the weights of the vertices with the same degree.
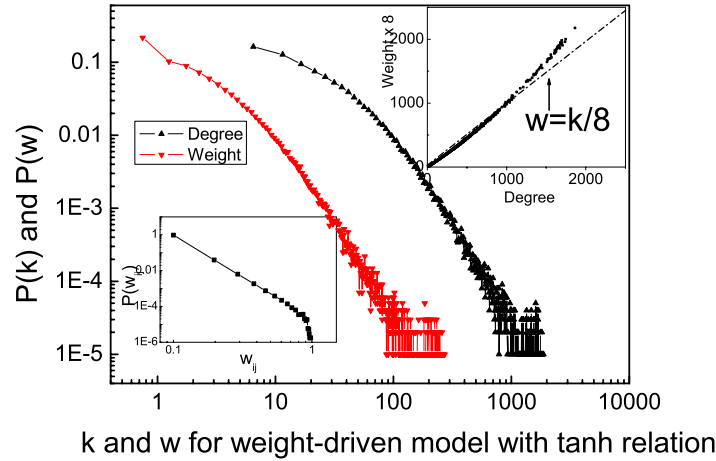
Figure 4. Nonlinear relation of equation (6) with $\alpha = 0.1$ between connection count and weight used in the weight-driven model. Typical behaviours do not show a large difference with the linear weight case. The main plots are degree and weight distributions, the top-right inset is their correlation with $\eta = 1.17$, and the bottom-left inset is the distribution of edge weight.

In the case of the pure degree-driven model, our model is similar with BA model, except now the old vertices can also be activated [17, 20, 21] and connection can be repeated. The typical distributions of degree and vertex weight from simulations of the degree-driven model are shown in figure 2(b), and edge weight distribution is shown in figure 3. There is another important reason driving us to consider both degree-referred and weight-referred preferential attachment. We want to investigate the structural role of weight. Should it be a fundamental quantity at the same level of connection, or is it a higher level structure, which can be defined by fundamental ones? The difference between those two models, and the conclusion as to which one behaves better, reflect the answer to the question: whether the weight of the edge is as fundamental as connection of the edge. However, as far as we can see from figure 2, both these models show typical behaviours even with the same exponent. In the weight-driven model, we also tried the tanh nonlinear weight, equation (6), to couple connection count to weight. The results in figure 4 do not show a large difference with the case of linear weight above.

4. Distance-dependent preferential attachment

According to the mechanism represented by the δ term in equation (2), for a given start vertex, the probability to attach to an end vertex is proportional to the similarity distance between them. Therefore, when a connection is built up from a new active vertex, the similarity distance to every existing vertex is 0, so an end vertex is randomly chosen. When a connection is built up from a vertex already connected to the network, the existing connections within the ∂^d neighbours form a distribution of the similarity distance. Then the probability to attach to a vertex is higher if it has a closer relation (larger similarity distance) with the active one. This term can act as an additional mechanism to the other two mechanisms, or it can drive the network evolution alone. First, we keep the degree term and the δ term in our model so that $p = \delta \in (0, 1)$, and let $\partial_n^d = \partial_n^2$
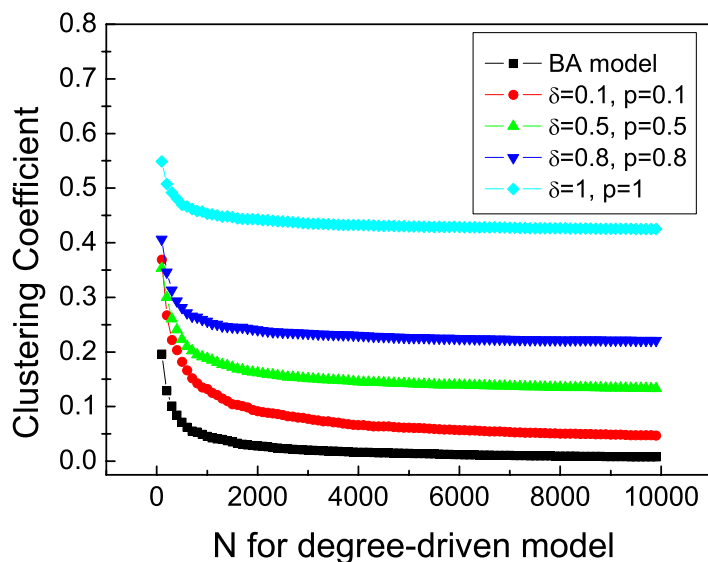
Figure 5. Clustering coefficient when the δ mechanism is added onto the pure degree-driven model, where the first line above is the one from the pure δ mechanism. The plots of weight/degree distributions have been omitted from those mixing models because they are very similar with the above ones from pure models.
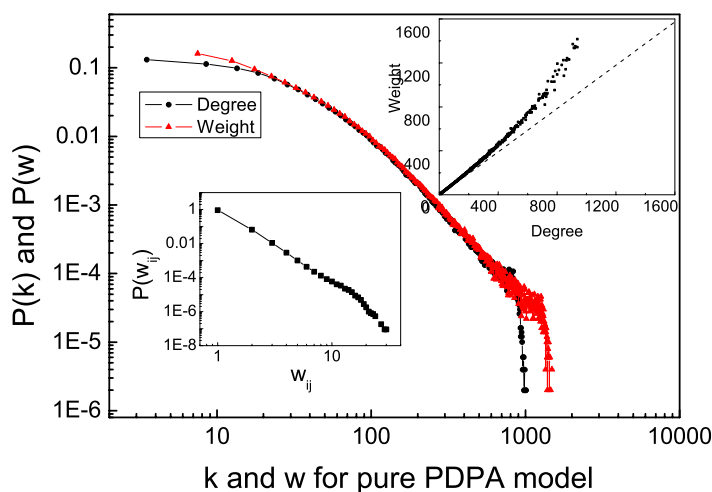

Figure 6. The model with pure δ mechanism demonstrates all the typical power-law distributions with a high clustering coefficient as shown in the top curve in figure 5. The top-right inset is the correlation between vertex weight and vertex degree, with $\eta = 1.19$. The bottom-left inset is the distribution of edge weight.

in equation (2). In figure 5, simulation shows that this δ mechanism significantly increases the clustering coefficient. The distribution of degree and weight still follow a power law. Similar results are obtained when this δ term is added into the weight-driven model. In figure 6, we set $p = \delta = 1$ to keep the δ term only and with also $d = 2$. The model is now a pure DDPA model. The simulation shows that even this case demonstrates all the typical properties and a high clustering coefficient.

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

5. Conclusion

In this paper, we presented evolutionary models for weighted networks which integrate the contributions from both new and old vertices. The three mechanisms: degree-driven, weight-driven and distance-dependent preferential attachment are discussed and compared. Weight has been assigned to each link according to the connection count of the link, which increases when the connection is repeated, so the weight of the link changes as the network evolves. Including the behaviours from the old vertices and recording the connection counts and converting them into weight, are the essential ideas in this work. The way to incorporate local distance into network evolution through the δ-mechanism is also one highlight of this work. Although it is not a new idea to convert connection counts into edge weight, it is the first time it has been the key point in network modelling. Because of this link between weight and connection count, we do not need another mechanism for the evolution of weights. However, the emergence of typical results from our models in an agreement with empirical studies makes this simple idea valuable.

Acknowledgments

References

[1] Li W and Cai X 2004 *Phys. Rev.* E **69** 046106
[2] Barrat A, Barthélemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747–52
[3] Bagler G 2004 *Preprint* cond-mat/0409773
[4] Newman M E J 2001 *Proc. Natl. Acad. Sci. USA* **98** 404
    Newman M E J 2001 *Phys. Rev.* E **64** 016132
[5] Fan Y *et al* 2004 *Int. J. Mod. Phys.* B **18** 2505–12
    M Li *et al* 2005 *Physica* A **350** 643–56
[6] Almaas E, Kovacs B, Vicsek T and Barabási A-L 2004 *Nature* **427** 839
[7] Yook S H, Jeong H, Barabási A-L and Tu Y 2001 *Phys. Rev. Lett.* **86** 5835
[8] Antal T and Krapivsky P L 2004 *Preprint* cond-mat/0408285
[9] Goh K-I, Noh J D, Kahng B and Kim D 2004 *Preprint* cond-mat/0410317
[10] Park K, Lai Y-C and Ye N 2004 *Phys. Rev.* E **70** 026109
[11] Barrat A, Barthélemy M and Vespignani A 2004 *Phys. Rev. Lett.* **92** 228701
    Barrat A, Barthélemy M and Vespignani A 2004 *Phys. Rev.* E **70** 066149
[12] Bianconi G 2004 *Preprint* cond-mat/0412399
[13] Dorogovtsev S N and Mendes1 J F F 2004 *Preprint* cond-mat/0408343
[14] Barabási A-L and Albert R 1999 *Science* **286** 509–12
[15] Holme P and Kim B J 2002 *Phys. Rev.* E **65** 026107
[16] Davidsen J, Ebel H and Bornholdt S 2002 *Phys. Rev. Lett.* **88** 128701
[17] Barabási A-L, Jeong H, Néda Z, Ravasz E, Schubert A and Vicsek T 2002 *Physica* A **311** 590
[18] Jost J and Joy M P 2002 *Phys. Rev.* E **66** 036126
[19] Vázquez A 2003 *Phys. Rev.* E **67** 056104
[20] Albert R and Barabási A-L 2000 *Phys. Rev. Lett.* **85** 5234
[21] Catanzaro M, Caldarelli G and Pietronero L 2004 *Physica* A **338** 119