International Journal of Modern Physics C
Vol. 18, No. 1 (2007) 43–52

PROPERTIES OF AN EVOLVING DIRECTED NETWORK WITH LOCAL RULES AND INTRINSIC VARIABLES

ANDRZEJ GRABOWSKI* and ROBERT A. KOSIŃSKI†

Central Institute for Labour Protection - National Research Institute
Czerniakowska 16, 00-701 Warsaw, Poland
Faculty of Physics, Warsaw University of Technology
Koszykowa 75, 00-662 Warsaw, Poland
* *angra@ciop.pl*
† *rokos@ciop.pl*

Received 5 December 2006
Revised 2 January 2007

We present a simple model of an evolving directed network based on local rules. It leads to a complex network with the properties of real systems, like scale-free distribution of outgoing and incoming connectivity, and a hierarchical structure. Each node is characterised by an intrinsic variable S, and the number of outgoing links k_{out}. As a result of network evolution the number of nodes and links (as well as their location) changes in time. For critical values of control parameters there is a transition to a scale-free network. Results for connectivity distribution found analytically agree with numerical calculations. Our model also reproduces other nontrivial properties of real networks, e.g. a large clustering coefficient and weak correlations between the age of a node and its connectivity. We have discovered an unexpected phenomenon that noise can increase the value of the clustering coefficient, whose large value is characteristic for a regular network.

Keywords: Complex networks; growing networks; local rules.

1. Introduction

In recent years it has been found that the structure of different biological, technical, economical and social systems has the form of a complex network.[1,2] The high value of the clustering coefficient and scale-free distribution of connectivity are some of the common properties of those networks.[2,3] Their evolution is successfully modelled using different approaches. However, in many models an undirected network is used even if a real network is directed[4] (e.g., models of the World Wide Web network[5,6]).

Different approaches to a generation of graphs with desirable properties, e.g. a degree distribution or correlations between node connectivity, have been presented.[7,8] Most of them assume that a new connection can be created between each pair of nodes with a certain probability (e.g., proportional to node connectivity[9]), but some models are based on local (i.e., involving a node and its neighbors) rules.[10–12]

Taking into account that real networks are often very large, the assumption that a node has knowledge of its local neighborhood only is much more realistic than the assumption that a node has knowledge of the whole network. This latter assumption is used in models of preferential attachment, where the connectivity of all nodes in the network is known to a node the moment a new connection is created. Hence, many papers describe routing strategies based on local information.[13]

A model of the evolution of the network, where the process of rewiring connections is based on local rules, is described in Sec. 2. The form of out-degree distribution found analytically is compared with numerical simulations in Sec. 3. The other properties of the network, like correlation between the clustering coefficient of a node and its connectivity, are presented in that section, too. The results are summarized in Sec. 4.

2. The Model

In our model we investigate the evolution of a directed network with local rules and intrinsic variables. Each node is described by an intrinsic variable S, drawn from an uniform distribution in the range $(0; 1)$.[6,14] This value does not change during time evolution of the system. The intrinsic variable can be referred eg. to the content of a web page (in the case of WWW) or to the cultural features of an individual (in the case of social networks[12]). At time $t = 0$ we create a pair of connected nodes.

Because most real networks grow in time we add in each time step of the simulation a new node with one outgoing link pointing to a randomly chosen node (multiple connections are not allowed). In real growing networks nodes can also be removed, thus we introduce parameter p_r in our model. With probability p_r a randomly chosen node and all its connections (outgoing and incoming links) are removed from the network. Therefore, as a result of adding and removing nodes in each time step the number of nodes increases by $1 - p_r$. Note that for $p_r = 1$ the number of nodes does not change in time.

The process of creating new connections between nodes is observed in many real networks, e.g., when people make new friends or a new link is added to a WWW page. This process often depends on the number of connections (e.g., it is more probable that a gregarious person with many friends will make a new friend). Therefore, in our model the number of outgoing links of a node increases with probability proportional to its present out-degree k_{out}.[15,16] A new connection to a randomly chosen node is created with probability p_{c}, which has following form

$$p_{\mathrm{c}} = \frac{k_{\mathrm{out}}}{N}\,, \tag{1}$$

where N is the number of nodes in the whole network. We use N as a normalization factor because the out-degree of any node cannot be greater than $N - 1 \approx N$.

The evolution of the location of links is based on local rules and depends on the values of intrinsic variables S. In order to describe this evolution we define distance d_{ij} between a pair of nodes (i, j)

$$d_{ij} = |S_i - S_j| \,. \tag{2}$$

In each time step an outgoing link of the ith node, pointing to the neighbor for which d_{ij} has the largest value, is chosen. Next that neighbor l of the neighbors of the ith node is chosen which has the smallest distance d_{il}. If $d_{il} < d_{ij}$, the chosen link can be rewired to node l instead of node j. Therefore, as a result of the rewiring process the average distance d_{ij} between the closest neighbors decreases.

In real networks, however, rewiring is connected with a certain cost; therefore, this process is not always rewarding. We take this into account in our model and we rewire a link with probability p_{rew} proportional to expected benefits, i.e., proportional to a decrease in distance between a node and its neighbor:

$$p_{\mathrm{rew}} = (d_{ij} - d_{il})^r \,, \tag{3}$$

where r is a control parameter regulating the rate of the rewiring process. For $r = 0$ there is no cost of rewiring and $p_{\mathrm{rew}} = 1$. The cost of rewiring increases when r increases. For $r \to \infty$ the probability of rewiring equals zero and the location of connections has its initial, random character (initially randomly chosen nodes are connected). The process of rewiring orders the structure of the network and decreases randomness in the system. Hence, the lack of rewiring in our model can be referred to as some kind of noise in the system; the greater the r, the greater the randomness (noise).

In our model we have three mechanisms of network evolution observed in many real networks:

(a) adding and removing nodes,[17]
(b) creating new connections between nodes and
(c) rewiring connections.
The evolution of the network is controlled by two parameters: p_r, which controls the rate of the growth of the network, and r, which controls the rate of the rewiring process.

3. Results

The process of rewiring does not influence the out-degree of a node. Thus, out-degree distribution is irrespective of parameter r. We can calculate analytically the form of this distribution using the following master equation:

$$\begin{aligned} n(k, t+1) = n(k,t) - \frac{n(k,t)}{N(t)}k - \frac{n(k,t)}{N(t)}p_r - \frac{n(k,t)}{N(t)}kp_r \\ + \frac{n(k-1,t)}{N(t)}(k-1) + \frac{n(k+1,t)}{N(t)}(k+1)p_r + \delta_{k,1} \,, \end{aligned} \tag{4}$$

where $n(k,t)$ denotes the number of nodes with out-degree k in time t. The number $N(t)$ of all nodes in time t equals:

$$N(t) = (1 - p_r)t\,. \tag{5}$$

The number of nodes with out-degree k decreases when a node with out-degree k:

(a) creates a new link — with probability proportional to k,
(b) is removed — with probability p_r and
(c) loses a connection pointing to a node removed in time t — with probability proportional to kp_r.

On the other hand, the number of nodes with out-degree k increases when (a) a node with out-degree $k-1$ creates a new link — with probability proportional to $k-1$, (b) a node with out-degree $k+1$ loses a connection pointing to a node removed in time t — with probability proportional to $(k+1)p_r$, and (c) in each time step we add a new node with one outgoing link ($k = 1$).

The probability $P(k,t)$ that a node has k outgoing links in time t equals $P(k,t) = n(k,t)/N(t)$; thus we can write Eq. (4) in the following form:

$$\begin{aligned}\frac{\partial[N(t)P(k,t)]}{\partial t} &= P(k-1,t)(k-1) + P(k+1,t)(k+1)p_r \\ &\quad - P(k,t)(k + kp_r + p_r) + \delta_{k,1}\,.\end{aligned} \tag{6}$$

The equation for stationary degree distribution $P(k) \equiv P(k, t \to \infty)$ is

$$P(k)(k + kp_r + 1) = P(k-1)(k-1) + P(k+1)(k+1)p_r + \delta_{k,1}\,, \tag{7}$$

because $\partial[N(t)P(k)]/\partial t = (1 - p_r)P(k)$.

In the case $p_r = 0$ we can calculate $P(k)$ without any additional conditions by multiplying both sides of the equation by k and the sum over k, and the solution of the equation is power-law out-degree distribution:[9]

$$P(k) = \frac{1}{k(k+1)} \sim k^{-2}\,. \tag{8}$$

In order to calculate the distribution $P(k)$ for any p_r let us pass to the continuum degree approximation.[9] Therefore, we write Eq. (7) in the following form:

$$\begin{aligned}P(k)(k + kp_r + 1) &= \left(P(k) - \frac{\partial P(k)}{\partial k} + \frac{1}{2}\frac{\partial^2 P(k)}{\partial k^2}\right)(k-1) \\ &\quad + \left(P(k) + \frac{\partial P(k)}{\partial k} + \frac{1}{2}\frac{\partial^2 P(k)}{\partial k^2}\right)(k+1)p_r\,.\end{aligned} \tag{9}$$

After calculations we obtain

$$P(k) \sim (k + \alpha)^{-\gamma_{\text{out}}}\,, \tag{10}$$

where $\alpha = (1 + p_r)/(1 - p_r)$ and

$$\gamma_{\text{out}} = \frac{2 - p_r}{1 - p_r}\,. \tag{11}$$

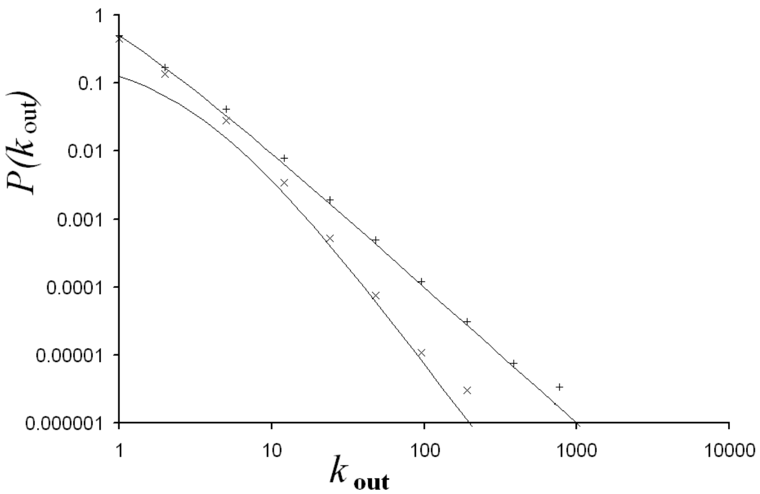

Fig. 1. Comparison of analytical (solid lines) and numerical results for $p_r = 0$ (+) and $p_r = 0.5$ (×). In the case of $p_r = 0.5$, when the continuum degree approximation was used, analytical results of out-degree distribution agree with numerical ones only for large enough k ($k > 10$). For $p_r = 0.4$ the value of $\gamma_{\text{out}} \approx 2.7$, note that in the case of WWW $\gamma_{\text{out}} \approx 2.7$.

Note that for $p_r = 0$ we obtain $\gamma_{\text{out}} = 2$. Thus, using two different approaches we get the same result (see Eq. (8)). A comparison of numerical and analytical results is presented in Fig. 1, and good agreement of both results is visible. Note that for $p_r = 1$ the network does not grow; therefore it cannot have scale-free distribution of connectivity.

In order to calculate the in-degree distribution we use numerical calculations. It turns out that for a wide range of control parameters, the network has scale-free properties with in-degree distribution in the form of a power law $P(k) \sim k^{-\gamma_{\text{in}}}$. The form of in-degree distribution for different values of r and $p_r = 0$ is shown in Fig. 2. The value of γ_{in} as a function of control parameters is shown in the phase diagram in Fig. 3. For low values of r and p_r (see Fig. 2 for $r = 0$) the process of rewiring is a dominant factor; therefore, there is a large number of nodes with high connectivity. For larger values of r and p_r stochastic processes cause a smaller number of nodes with large connectivity. In the range of control parameters for which the network is scale-free the value of γ_{in} starts from 2.1 for $r = 0$ and increases fast with an increasing r (note that in the case of WWW $\gamma_{\text{in}} = 2.1$).[9] When r increases, the rate of the rewiring process decreases. Thus, the location of connections is more random. For large enough r in-degree distribution has an exponential form $P(k) \sim e^{-k}$, which is typical for growing random networks.[9] The value of γ_{in} also increases with increasing probability of the removal of node p_r. Nodes are removed irrespectively of their connectivity; however, with nodes also their connections (incoming and outgoing) are removed and the in-degree of nodes indicated by a removed node decreases. Therefore, the number

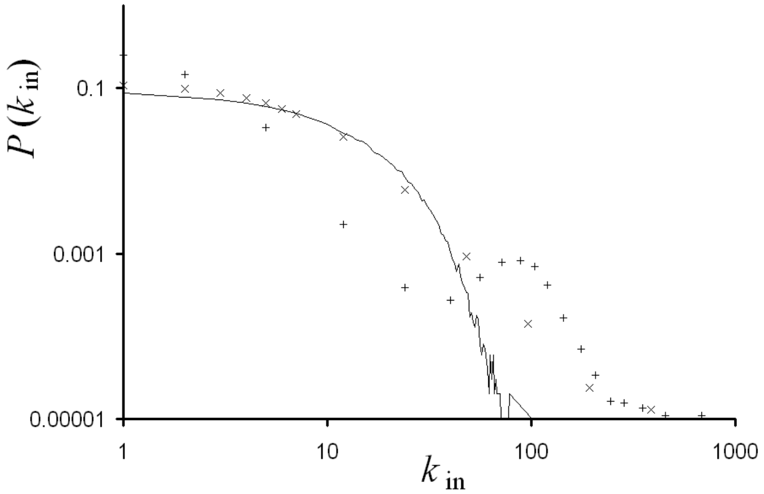

Fig. 2. The in-degree distribution $P(k_{\text{in}})$ for $p_r = 0$ and different values of r ($r = 0$ (+); $r = 5$ ($\times$) and $r = 16$ — solid line). In each case the number of nodes is $N = 10^5$.

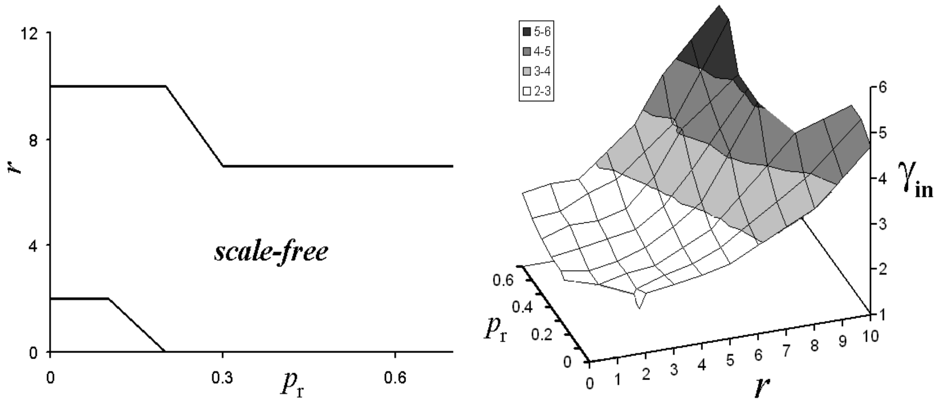

Fig. 3. Phase diagram. The range of control parameters when the in-degree distribution has power-law form $P(k) \sim k^{-\gamma_{\text{in}}}$ (scale-free network), and the relation between values of control parameters and value of γ_{in} is shown, in the left and right column, respectively. The value of γ_{in} increases with r and p_r increasing.

of nodes with large connectivity decreases faster than the number of nodes with low connectivity.

We calculate the clustering coefficient C of the network using the following equation

$$C = \left\langle \frac{E_i}{k_i(k_i - 1)} \right\rangle , \tag{12}$$

where E_i is the number of connections between neighbors of the ith node and $\langle \cdot \rangle$ means averaging over all nodes. The results of numerical simulations are presented

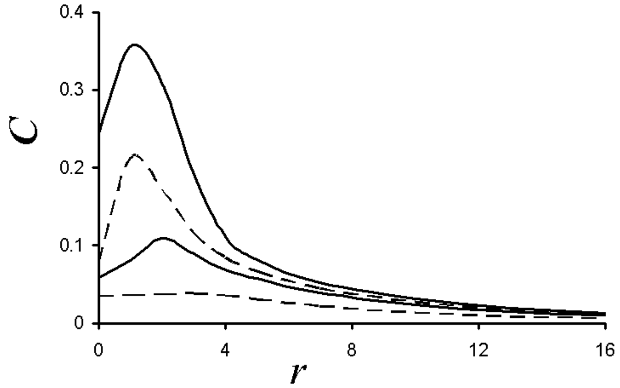

Fig. 4. The relation between the clustering coefficient C and r for different values of the probability of node removal p_r (0; 0.1; 0.2; 0.6 from top to bottom, respectively). The values of other parameters are: $N = 10^5$.

in Fig. 4. For a low value of r the clustering coefficient has a large value, as a result of the high rate of the rewiring process. During this process, nodes create connections with nodes which have similar values of intrinsic variable S. As a result, groups of highly interconnected nodes emerge in the network. A high value of C is observed in many real networks, e.g., in social networks or in the WWW (we calculate in the same way the value of the clustering coefficient of a network from the nd.edu domain;[a] the result is $C \approx 0.17$). For large r, the clustering coefficient of the network decreases with an increasing r, because the structure of the network is more similar to the structure of random growing networks (which have a very low value of C). However, the clustering coefficient initially increases with an increasing r, which is an unexpected phenomenon. It turns out that a less ordered network has a greater value of C (note that a high value of the clustering coefficient is typical for regular networks, not for random ones[1,2,9]). Similar behavior is observed in the case of stochastic resonance, where the optimal level of noise in the system leads to a maximum value of the signal-to-noise ratio (SNR).[18] The value of the clustering coefficient decreases with an increasing p_r, as a result of destroyed clusters of highly interconnected nodes.

The behavior of the clustering coefficient of a node for our model and for real networks is an interesting problem. The clustering coefficient of a node decreases with its number of connections, and for a high rate of the rewiring process (low r) the power-law relation $C(k) \sim k^{-\beta}$ is visible (Fig. 5). The value of β slightly depends on the values of p_r and for low r ($r < 4$) equals approximately 1. Such a power law is observed in some real networks (e.g., in an actors' network, the Internet at the level of the autonomous system or the WWW).[19,20] The power-law relation $C(k)$ obtained in our calculations is similar to the relation observed

[a]We would like to thank A.-L. Barabási that data from nd.edu domain are publicly available

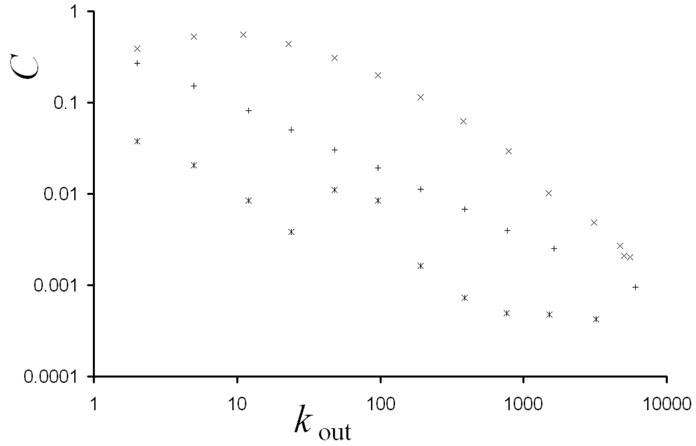

Fig. 5. Relation between the clustering coefficient of a node and its outgoing connectivity k_{out} for $p_r = 0.1$ and different values of r (1; 4; 16 from top to bottom, respectively). The values of other parameters are: $N = 10^5$.

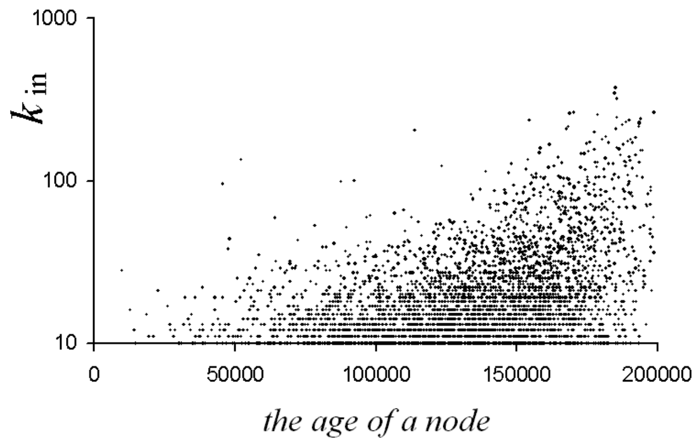

Fig. 6. The correlations between age of a node and its ingoing connectivity k_{in} for $p_r = 0.5$. The values of other parameters are: $N = 10^5$, $r = 0$.

in hierarchical networks.[19] Such power laws hint at the presence of a hierarchical architecture: when small groups organize themselves into increasingly larger groups in a hierarchical manner, local clustering decreases on different scales according to such a power law.

In a number of models of growing networks a strong correlation between the age of a node and its connectivity is observed (e.g., the Barabási–Albert model of preferential attachment[5]). However, in many real networks, like the WWW, an opposite situation is observed.[21] Figure 6 illustrates that in our model, a weak correlation between the age of a node and its in-degree k_{in} is also observed. The greater the p_r and the smaller the r, the weaker the correlations that are visible.

4. Conclusions

In our model we investigate the properties of an evolving directed network. It turns out that a simple rewiring process based on local rules leads to a network which has nontrivial properties of real networks, e.g., scale-free distribution of connectivity, a large value of the clustering coefficient and a hierarchical structure.

The form of distribution of outgoing connectivity obtained analytically agrees very well with the numerical results. We calculate the phase diagram for the case when the network has scale-free properties and the in- and out-degree distributions follow the power law.

A network generated by our model has a large value of the clustering coefficient for a wide range of control parameters. We have discovered that noise (randomness) can increase the value of the clustering coefficient. This is an unexpected phenomenon, because a large value of the clustering coefficient is characteristic for regular networks, not for random ones. The relation between the clustering coefficient of a node and its outgoing connectivity has a power-law form $C(k) \sim k^{-\beta}$, where $\beta \approx 1$. Such power laws hint at the presence of a hierarchical architecture: when small groups organize themselves into increasingly larger groups in a hierarchical manner.

Our model can be treated as a model of an evolution of the WWW network, where the intrinsic variable can be referred to the content of a page. For a wide range of control parameters the generated network has similar properties to a real WWW, i.e. a hierarchical structure, large clustering, weak correlation between the age of a node and its connectivity, and scale-free in- and out-degree distributions.

Acknowledgments

This work as a project of the COST Action P10 “Physics of Risk” was supported by Polish Ministry of Science and Higher Education, Grant No. 134/E-365/SPB/COST/KN/DWM 105/2005–2007, and by a special Grant of Warsaw University of Technology.

References

1. R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
2. S. H. Strogatz, *Nature* **410**, 268 (2001).
3. A. E. Motter, T. Nishikawa and Y. Lai, *Phys. Rev. E* **68**, 036105 (2003).
4. S. Bernhardsson and P. Minnhagen, *Phys. Rev. E* **74**, 026104 (2006).
5. A.-L. Barabási, R. Albert and H. Jeong, *Physica A* **281**, 69 (2000).
6. A. Capocci, G. Caldarelli and P. D. L. Rios, *Phys. Rev. E* **68**, 047101 (2003).
7. M. Boguna and R. Pastor-Satorras, *Phys. Rev. E* **68**, 036112 (2003).
8. A. Grabowski and R. A. Kosiński, *Physica A* **361**, 651 (2006).
9. S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford Univ. Press, 2004).
10. A. Vazquez, *Phys. Rev. E* **67**, 056104 (2003).
11. H. D. Rozenfeld and D. ben-Avraham, *Phys. Rev. E* **70**, 056107 (2004).

12. A. Grabowski and R. A. Kosiński, *Phys. Rev. E* **73**, 016135 (2006).
13. C.-Y. Yin, B.-H. Wang, W.-X. Wang, T. Zhou and H.-J. Yang, *Phys. Lett. A* **351**, 220 (2006).
14. G. Caldarelli, A. Capocci, P. De Los Rios and M. A. Munoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
15. B. A. Huberman and L. A. Adamic, *Nature* **401**, 131 (1999).
16. P. L. Krapivsky, G. J. Rodgers and S. Redner, *Phys. Rev. Lett.* **86**, 5401 (2001).
17. C. Moore, G. Ghoshal and M. E. J. Newman, *Phys. Rev. E* **74**, 036121 (2006).
18. A. Krawiecki, A. Sukiennicki and R. A. Kosinski, *Phys. Rev. E* **62**, 7683 (2000).
19. E. Ravasz and A. L. Barabási, *Phys. Rev. E* **67**, 026112 (2003).
20. G. Csanyi and B. Szendroi, *Phys. Rev. E* **69**, 036131 (2004).
21. L. A. Adamic and B. A. Huberman, *Science* **287**, 2115a (2000).