Vol. 38 (2007) ACTA PHYSICA POLONICA B No 5

EVOLVING DIRECTED NETWORK WITH INTRINSIC VARIABLES AND LOCAL RULES: A SIMPLE MODEL OF WWW NETWORK*

Andrzej Grabowski, Robert Kosiński

Central Institute for Labour Protection, National Research Institute
Czerniakowska 16, 00-701 Warsaw, Poland
and
Faculty of Physics, Warsaw University of Technology
Pl. Politechniki 1, 00-662 Warsaw, Poland

(Received August 10, 2006)

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 in- and out-degree distributions and a hierarchical structure. Each node is characterised by intrinsic variable S and the number of outgoing k_{out} and incoming k_{in} links. 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. Our model also reproduces other nontrivial properties of real WWW network, *e.g.* a large clustering coefficient and weak correlations between the age of a node and its connectivity.

PACS numbers: 89.75.–k, 89.75.Da, 87.23.Ge

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 connec-

* Presented at the XIX Marian Smoluchowski Symposium on Statistical Physics, Kraków, Poland, May 14–17, 2006.

tion 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 neighbours) rules [10–12].

Taking into account that real networks are often very large, the assumption that a node has knowledge of its local neighbourhood only is much more realistic than the assumption that a node has knowledge of the whole network. Such an 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].

In our model we investigate the evolution of a directed network with local rules and intrinsic variables. Each node is described by intrinsic variable S, drawn from uniform distribution in the range $(0;1)$ [6,14]. This value does not change during time evolution of the system. In time $t=0$ we create a set of m fully connected nodes. Thus, each node has $m-1$ outgoing links and $m-1$ incoming links.

In one time step of simulation we add a new node with randomly generated number of outgoing links $n \in (0;m]$ pointing to randomly chosen nodes (multiple connections are not allowed). In real growing networks nodes can also be removed, thus in each time step we remove randomly chosen node and all its outgoing links from the network (incoming links are rewired to randomly chosen nodes).

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 out-degree of a node increases with probability proportional to its present out-degree k_{out} [15,16]. A new outgoing link to a randomly chosen node is created with probability p_{cc}, which has the following form

$$p_{\mathrm{cc}} = r\frac{k_{\mathrm{out}}}{k_{\mathrm{out}}^{\mathrm{max}}}\,, \tag{1}$$

where $k_{\mathrm{out}}^{\mathrm{max}}$ is the maximal out-degree in the network, and $r \in (0;1)$ is a parameter which controls the process of adding new connections. The greater r the faster increases the number of connections. A scheme of the creation of new connections is shown in Fig. 1(a).

Number of outgoing links grows not only by simple adding new links to other nodes in the network (in our model with probability p_{cc}). In some real networks this process is connected with creating of a new node. In the case of WWW a new page containing more data on a given subject can be added to main page. Such a process is more probable for larger (*i.e.* with greater

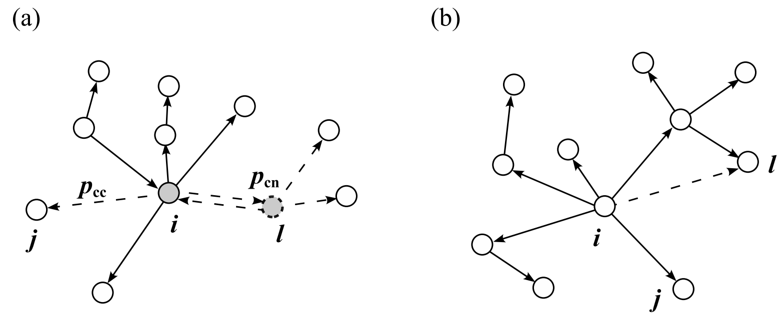

Fig. 1. A scheme of the creation of new connections (a). New connections are denoted by dashed line. Node i creates a new connection to a randomly chosen node j with probability $p_{\rm cc}$ and with probability $p_{\rm cn}$ a new node l is created and attached to the i-th node. The value of intrinsic variable of the new node is $S_{\rm l} = S_{\rm i}$. As a result of rewiring process (b) the connection pointing to the j-th node is rewired to the l-th node. A new neighbour of the i-th individual is chosen from neighbours of the current neighbours of the i-th individual.

$k_{\rm out}$) and more popular (*i.e.* with greater $k_{\rm in}$) pages. Therefore, in our model each node creates a new node and connection to it with probability $p_{\rm cn}$ proportional to its in- and out-degree

$$p_{\rm cn} = r\frac{k_{\rm out} + k_{\rm in}}{max\,(k_{\rm out} + k_{\rm in})}\,. \tag{2}$$

New node has the same value of variable S and $n \in (0; m]$ connections. One of its links points to main node (the i-th node in Fig. 1(a)) and the rest to randomly chosen nodes (see Fig. 1(a)).

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} which measures the difference between values of intrinsic variables of a pair of nodes (i,j)

$$d_{ij} = |S_{\rm i} - S_{\rm j}|\,. \tag{3}$$

In each time step an outgoing link of the i-th node, pointing to a neighbour for which d_{ij} has the largest value, is chosen. Next, a neighbour of the neighbours of the i-th node is chosen with respect to the smallest distance d_{il}. If $d_{il} < d_{ij}$ the chosen link is rewired to a l-th node (see Fig. 1(b)).

In real networks, rewiring is connected with a certain cost, therefore, this process is not always rewarding. However, in the case of WWW cost of rewiring is very small and in order to simplify our model we do not take it into account.

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: r, which controls the rate of the growth of the network, and m, the maximal number of initial connections of a node.

2. Results

In order to calculate the in- and out-degree distribution we use numerical calculations. It turns out, that for wide range of control parameters ($m > 1$) the network has scale-free properties with power-law form of out-degree distribution $P(k) \sim k^{-\gamma_{\rm out}}$ and in-degree distribution $P(k) \sim k^{-\gamma_{\rm in}}$. The values of $\gamma_{\rm in}$ and $\gamma_{\rm out}$ in the function of control parameters is shown in Fig. 2.

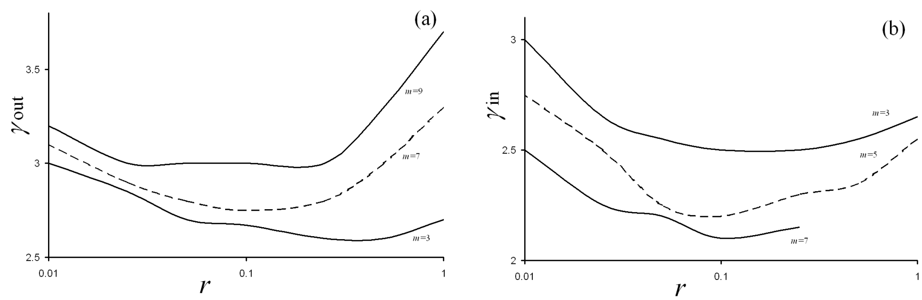

Fig. 2. The influence of value of the parameter r on $\gamma_{\rm out}$ (a) and $\gamma_{\rm in}$ (b) for different values of m. The in- and out-degree distributions have power-law form (scale-free network), $P(k_{\rm out}) \sim k_{\rm out}^{-\gamma_{\rm out}}$ and $P(k_{\rm in}) \sim k_{\rm in}^{-\gamma_{\rm in}}$, respectively.

The in-degree distribution is not scale-free for $m > 5$ and $r < 0.5$, because the process of creating new links is too fast. As a result of fast creation of new links, the number of nodes with different values of S is low and there is a large number of nodes with low connectivity (parallel with process of creating new connections also nodes with small number of links are created with probability $p_{\rm cn}$, see Eq. 2).

The in-degree $k_{\rm in}$ increases as a result of creating new connections and rewiring, but the influence of the process of rewiring is greater. When the number of nodes increases too fast (grater r), the rate of the rewiring process decreases and number of nodes with very large $k_{\rm in}$ is low. Thus, an increase in $\gamma_{\rm in}$ is observed for large r. On the other hand for very low r ($r < 0.01$), the stochastic processes of adding and removing nodes are a dominant factor and cause the distribution of links has exponential form $P(k) \sim e^{-\rm k}$, which is typical for growing random networks [9]. Hence, if r is low enough the values of $\gamma_{\rm in}$ and $\gamma_{\rm out}$ increase with a decreasing r.

For values of $r \approx 0.1$ the influence of the process of adding and removing nodes and the process of creating new connections seems to be balanced and therefore, $\gamma_{\rm in}$ and $\gamma_{\rm out}$ takes minimal values. It is interesting that for $r = 0.1$ obtained network has the same values of exponents of in- and out-degree distributions as WWW network (see Fig. 3) [9]. Moreover, in the case of obtained out-degree distribution the values of $P(k_{\rm out})$ for low $k_{\rm out}$ ($k_{\rm out} < m$) are approximately independent on $k_{\rm out}$, and this behaviour is also observed in the case of WWW network [5, 9].

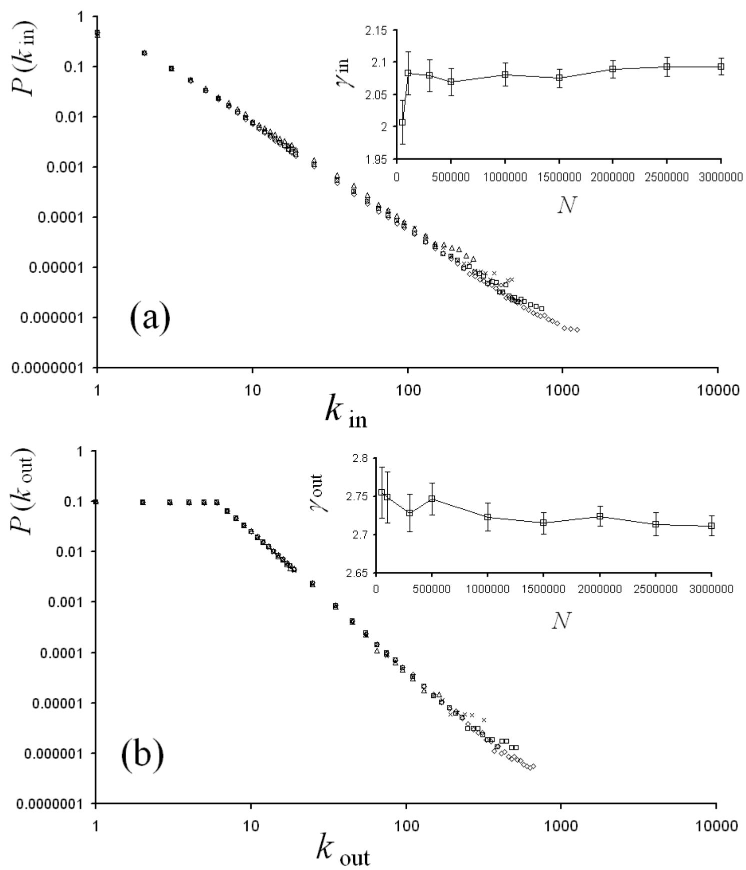

Fig. 3. The in-degree distribution $P(k_{\rm in}) \sim k_{\rm in}^{-\gamma_{\rm in}}$ (a) and out-degree distribution $P(k_{\rm out}) \sim k_{\rm out}^{-\gamma_{\rm out}}$ (b) for $m = 7$, $r = 0.1$ and different sizes of the network ($N = 10^5$ — triangles; $N = 3 \times 10^5$ — crosses; $N = 10^6$ - squares; $N = 3 \times 10^6$ — diamonds). The values of exponents, $\gamma_{\rm in} \approx 2.1$ and $\gamma_{\rm out} \approx 2.7$, are the same as in the case of WWW network. The data were log-binned, to reduce the uneven statistical fluctuations common in heavy-tailed distributions, a procedure that does not alter the slope of the tail.

Values of exponents $\gamma_{\rm in}$ and $\gamma_{\rm out}$ are slightly influenced by size of the system (see the inset in Fig. 3) and initial conditions (see Fig. 4).

Another characteristic feature of WWW network is a large value of a conditional probability $p \approx 0.5$, that a node x has a connection to a node y if node y has a connection to x [9]. In our model p takes also large values, which are approximately independent on the value of parameter r. In the case of $m = 1$ the probability p has very low values, however for $m > 1$ the probability p is large and decreases with an increasing m (*e.g.* $p \approx 0.46; 0.38; 0.31$ for $m = 3; 5; 7$, respectively).

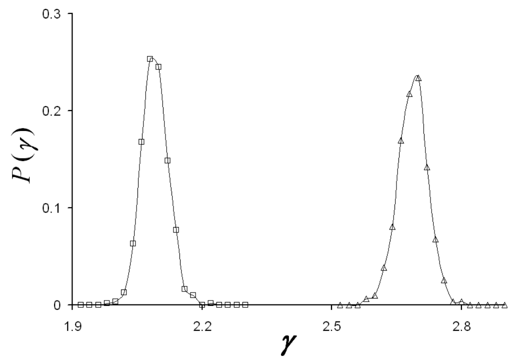

Fig. 4. The influence of the initial conditions on the values of exponents $\gamma_{\rm in}$ (squares) and $\gamma_{\rm out}$ (triangles) for $m = 7$, $r = 0.1$ and $N = 3 \times 10^5$. $P(\gamma)$ is the probability density function and the standard deviation equals $\sigma = 0.03$ for $\gamma_{\rm in}$ and $\sigma = 0.035$ for $\gamma_{\rm out}$. The average values of the exponents are: $\gamma_{\rm in} \approx 2.1$ and $\gamma_{\rm out} \approx 2.7$. Results were averaged over 600 independent simulations.

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

$$C = \left\langle \frac{E_{\rm i}}{k_{\rm i}(k_{\rm i} - 1)} \right\rangle , \tag{4}$$

where E_i is the number of connections between neighbours of the i-th node and $\langle \ldots \rangle$ means averaging over all nodes. The results of numerical simulations are presented in Fig. 5. For a low value of r the clustering coefficient has a large value, as a result of high rate of 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 network (we calculate in the same way the value of the clustering coefficient of a network from the `nd.edu` domain [18]; the result is $C \approx 0.17$). For large r, the process of rewiring is not enough effective as result of fast increase in number of nodes. Therefore, the clustering coefficient decreases with an increasing r.

The behaviour 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 $m > 1$ the power-law relation $C(k) \sim k^{-\beta}$ is visible (Fig. 5). The value of β slightly depends on the values of r and m and equals approximately $b \approx 0.85$. Such a power law is observed in some real networks (*e.g.* in an actor network, the Internet at the level of autonomous system or the WWW) [19, 20]. The power-law relation $C(k)$ obtained in our calculations is similar to the relation observed in hierarchical networks [19]. Such power laws hint at the presence of a hierarchical architecture: when small groups organise themselves into increasingly larger groups in a hierarchical manner, local clustering decreases on different scales according to such a power law.

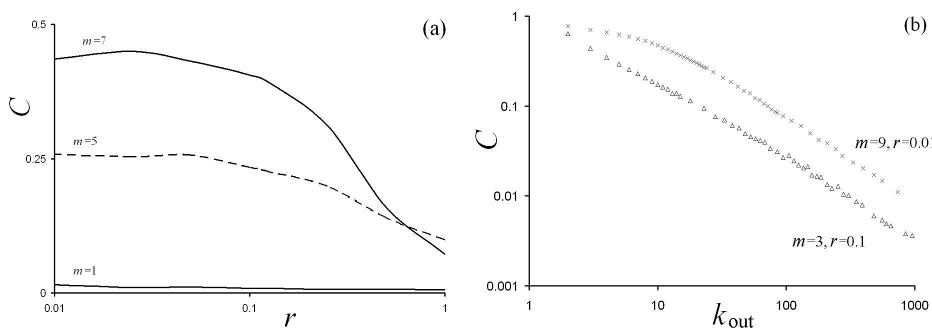

Fig. 5. Relation between the clustering coefficient and r (a) and between the clustering coefficient of a node and its out-degree $k_{\rm out}$ (b). The data were log-binned, to reduce the uneven statistical fluctuations.

Fig. 6. The correlations between the age of a node and its in-degree $k_{\rm in}$, for $p_{\rm r} = 0.5$. The values of other parameters are: $N = 3 \times 10^5$, $r = 0$.

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

3. 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.

A network generated by our model has a large value of the clustering coefficient for a wide range of control parameters. The relation between the clustering coefficient of a node and its outgoing connectivity has a power-law form $C(k) \sim k^{-\beta}$, where $\beta \approx 0.85$. Such power laws hint at the presence of a hierarchical architecture: when small groups organise into increasingly larger groups in a hierarchical manner.

Our model can be treated as a model of an evolution of 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 network, *i.e.* hierarchical structure, large clustering, weak correlation between the age of a node and its connectivity, scale-free in- and out-degree distributions.

The work was supported by the Polish Ministry of Science and Higher Education, Grant No. 1 P03B 145 29.

REFERENCES

[1] R. Albért, A-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
[2] S.H. Strogatz, *Nature* **410**, 268 (2001).
[3] A.E. Motter, T. Nishikawa, Y. Lai, *Phys. Rev.* **E68**, 036105 (2003).
[4] S. Bernhardsson, P. Minnhagen, *Phys. Rev.* **E74**, 026104 (2006).
[5] A-L. Barabási, R. Albért, H. Jeong, *Physica A* **281**, 69 (2000).
[6] A. Capocci, G. Caldarelli, P.D.L. Rios, *Phys. Rev.* **E68**, 047101 (2003).
[7] M. Boguna, R. Pastor-Satorras, *Phys. Rev.* **E68**, 036112 (2003).

[8] A. Grabowski, R.A. Kosiński, *Physica A*, **361**, 651 (2006).

[9] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks*, Oxford Univ. Press, 2004.

[10] A. Vazquez, *Phys. Rev.* **E67**, 056104 (2003).

[11] H.D. Rozenfeld, D. ben-Avraham, *Phys. Rev.* **E70**, 056107 (2004).

[12] A. Grabowski, R.A. Kosiński, *Phys. Rev.* **E73**, 016135 (2006).

[13] Chuan-Yang Yin, Bing-Hong Wang, Wen-Xu Wang, Tao Zhou, Hui-Jie Yang, *Phys. Lett.* **A351**, 220 (2006).

[14] G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Munoz, *Phys. Rev. Lett.* **89**, 258702 (2002).

[15] B.A. Huberman, L.A. Adamic, *Nature* **401**, 131 (1999).

[16] P.L. Krapivsky, G.J. Rodgers, S. Redner, *Phys. Rev. Lett.* **86**, 5401 (2001).

[17] C. Moore, G. Ghoshal, M.E.J. Newman, *Phys. Rev.* **E74**, 036121 (2006).

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

[19] E. Ravasz, A.L. Barabasi, *Phys. Rev.* **E67**, 026112 (2003).

[20] G. Csanyi, B. Szendroi, *Phys. Rev.* **E69**, 036131 (2004).

[21] L.A. Adamic, B.A. Huberman, *Science* **287**, 2115a (2000).