NODE-NODE DISTANCE DISTRIBUTION FOR GROWING NETWORKS*

KRZYSZTOF MALARZ, JUSTYNA KARPIŃSKA,
ALEKSANDER KARDAS AND KRZYSZTOF KUŁAKOWSKI

Faculty of Physics and Nuclear Techniques,
AGH University of Science and Technology,
al. Mickiewicza 30, PL-30059 Cracow, Poland
malarz@agh.edu.pl

(Received 30 September 2003)

Abstract: We present a simulation of the time evolution of the distance matrix. The result is a node-node distance distribution for various kinds of networks. For the exponential trees, analytical formulas are derived for the moments of distance distribution.

Keywords: evolving networks, random networks, scale-free networks, graphs, trees, small-world effect

1. Introduction

A graph is defined as a set of nodes (vertices) and a set of links among nodes (edges) [1–4]. By graph evolution or growth we mean subsequent attaching of new nodes with m edges to previously existing nodes [5]. Such growing graphs may reflect some features of real evolving networks, *e.g.* a network of collaborators, a network of citations of scientific papers, some biological networks (food chains or sexual relations) or Internet and world-wide-web pages with links between them [5–9].

The distance between nodes is the smallest number of edges from one node to the other. The node-node distance (NND) distribution depends on how the subsequent nodes are attached. If each node is connected with *only one* of the preexisting nodes ($m=1$) *a tree* appears. When a new node is attached to several different nodes with $m>1$ edges, the growing structure is called *a simple graph*. We may choose nodes to which new nodes are attached preferentially or randomly. In the latter case we deal with *exponential* networks. If the probability of choosing a node is proportional to its degree (*e.g.* to the number of its nearest neighbors) the growing structure is called a *scale-free* or Albert-Barabási network [10].

In this paper, a numerical algorithm for network growth – based on *distance matrix* evolution – is presented, both for exponential and scale-free networks ($m=$

* Presented at the 37th Polish Physicists' Meeting, Gdansk, Poland, 15–19 September 2003

1,2) [11, 12]. The NND distribution and its characteristics are calculated. Iterative formulas for n^{th} ordinary moments of the NND distribution are derived for exponential trees.

2. Computer simulations

A graph with edges of unit length may be fully characterized by its distance matrix $\mathbf{S}$, an element s_{ij} of which is equal to the shortest path between nodes i and j. This matrix representation is particularly useful when computer simulations for graph evolution are applied.

Attaching a subsequent node with one edge ($m=1$) to a previously existing network of N nodes corresponds to adding a new $(N+1)^{\text{th}}$ row and a new column to $N\times N$ large distance matrix $\mathbf{S}$. The distance from the newly added $(N+1)^{\text{th}}$ node to all others via a selected node labeled as p is larger by one than the distance between the p^{th} node and all others. Thus, the new $(N+1)^{\text{th}}$ row/column is a simple copy of the p^{th} row/column but with all of its elements incremented [11]:

$$\forall\ 1\le i\le N: s_{N+1}(N+1,i)=s_{N+1}(i,N+1)=s_N(p,i)+1. \tag{1}$$

Similarly, when a new node is attached to a network with two edges ($m=2$) to two different nodes labeled as p and q, the distance from all other nodes i to the newly added $(N+1)^{\text{th}}$ node is one plus the smaller distance of the p–i and q–i node pairs [12]:

$$\forall\ 1\le i\le N: s_{N+1}(N+1,i)=s_{N+1}(i,N+1)=\min(s_N(p,i),s_N(q,i))+1. \tag{2}$$

In the above-mentioned case of growth of simple graphs, distances between nodes i and j must also be re-evaluated to check if adding a new node produces a shortcut [12]:

$$\forall\ 1\le i,j\le N: s_{N+1}(i,j)=\min(s_N(i,j),s_N(i,p)+2+s_N(q,j)). \tag{3}$$

In both cases the diagonal elements of the new row/column are zero [11, 12]:

$$s_{N+1}(N+1,N+1)=0. \tag{4}$$

Selection of rows/columns (nodes to which we attempt to add a new node) may be random or preferential. In the latter case an additional evolving vector is introduced, which contains node labels. These labels occur as vector elements with a probability proportional to the degree of the node. Random selection of elements of such a vector corresponds to the Albert-Barabási construction rule. The procedure is known as the Kertész algorithm [13].

3. Analytical calculations

Let us define n^{th} moments of the NND distribution for all distances:

$$\ell_N^n\equiv[\{s^n(i,j)\}]=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[s^n(i,j)], \tag{5}$$

and for non-zero distances only:

$$d_N^n\equiv[\langle s^n(i,j)\rangle]=\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}[s^n(i,j)], \tag{6}$$

where $\{\cdots\}$, $\langle\cdots\rangle$ and $[\cdots]$ denote an average over N^2 matrix elements, an average over $N(N-1)$ non-diagonal matrix elements, and an average over $N_{\rm run}$ independent realizations of the evolution process (matrices), respectively. Moments (5) and (6) for $n=1$ are sometimes called *the network diameter*. Both double sums in r.h.s. of Equations (5) and (6) are equal, due to the obvious fact that $s(i,i)=0$. That allows us to derive a simple dependence between averages $\{\cdots\}$ and $\langle\cdots\rangle$:

$$N\ell_N^n = (N-1)d_N^n. \tag{7}$$

For exponential trees – assuming $s(i,i)=0$ and distance matrix symmetry $s(i,j)=s(j,i)$ – we are able to construct iterative equations for ℓ_{N+1}^n as dependent on ℓ_N^k $(k=1,\ldots,n)$:

$$(N+1)^2\ell_{N+1}^n = \sum_{i=1}^{N+1}\sum_{j=1}^{N+1}[s^n(i,j)] = N^2\ell_N^n + 2\sum_{i=1}^{N}(1+[s(i,q)])^n, \tag{8}$$

where q is the number of the randomly selected row/column of the distance matrix $\mathbf{S}$. A combination of Equations (8) and (7) yields the desired iterative formula:

$$d_{N+1}^n = \frac{(N+2)(N-1)}{(N+1)N}d_N^n + \frac{2}{N+1} + \frac{2(N-1)}{(N+1)N}\sum_{k=1}^{n-1}\binom{n}{k}d_N^k. \tag{9}$$

4. Results and conclusions

For trees, the mean of the NND d_N^1 and its dispersion, $\sigma^2 \equiv d_N^2 - (d_N^1)^2$, grow logarithmically with N (see Tables 1 and 2) [11]. For graphs, only the first cumulant (the average of the NND d_N^1) grows logarithmically (see Table 2) [12]. Such a slow increase of d_N^1 with the number of network nodes is known as the small-world effect [14].

Table 1. Mean distance, $d(N) = a\ln N + b$, for different evolving networks

	exponential	exponential	scale-free	scale-free
m	1	2	1	2
a	2.00	0.71	1.00	0.48
b	−2.84	0.16	−0.08	0.83

Table 2. Dispersion, $\sigma^2(N) = c\ln N + d$, for exponential and scale-free trees $(m=1)$

	exponential	scale-free
c	2.00	1.00
d	−1.44	−1.64

A histogram of NND is presented in Figure 1. As expected, NND's for graphs are more condensed than NND's for trees, while scale-free graphs (trees) are more condensed than exponential graphs (trees).

Knowing the moments d_N^n – the averages of the $n^{\rm th}$ powers of the non-diagonal distance matrix elements (6) – allows us to build all statistical parameters which

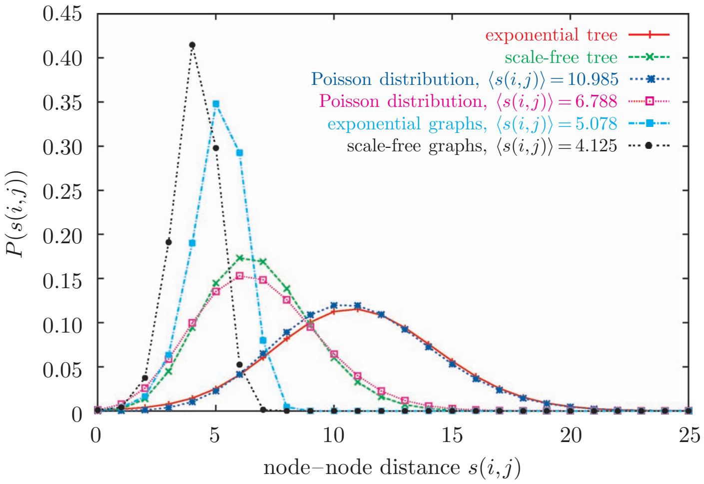

Figure 1. The NND distribution for different types of trees and graphs: $N = 1\,000$, $N_{\text{run}} = 10^5$ $(m=1)$, $N_{\text{run}} = 10^4$ $(m=2)$

characterize the NND distribution, *e.g.* the average distance d, the distance dispersion σ^2, its skewness

$$\nu_3 \equiv \frac{d_N^3 - 3d_N^2 d_N^1 + 2(d_N^1)^3}{\sigma^3}, \tag{10}$$

and kurtosis

$$\kappa_4 \equiv \frac{d_N^4 - 4d_N^3 d_N^1 + 6d_N^2 (d_N^1)^2 - 3(d_N^1)^4}{\sigma^4}. \tag{11}$$

The values of such characteristics of NND for exponential trees obtained from Equation (9) are presented in Figures 2 and 3. For trees, the distributions are similar to the Poisson distribution (see Figure 1). However, the skewness and kurtosis do not vanish even for large N, as one may expect for normal distributions [15].

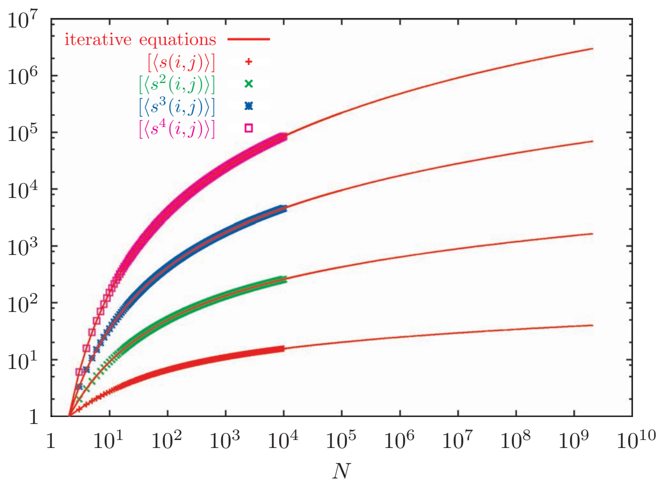

Figure 2. Main moments d_N^k $(k = 1,\ldots,4)$ for exponential trees given by Equation (9) (lines) and from direct simulations (symbols). The latter are averaged over $N_{\text{run}} = 10^4$ independent evolution process realizations

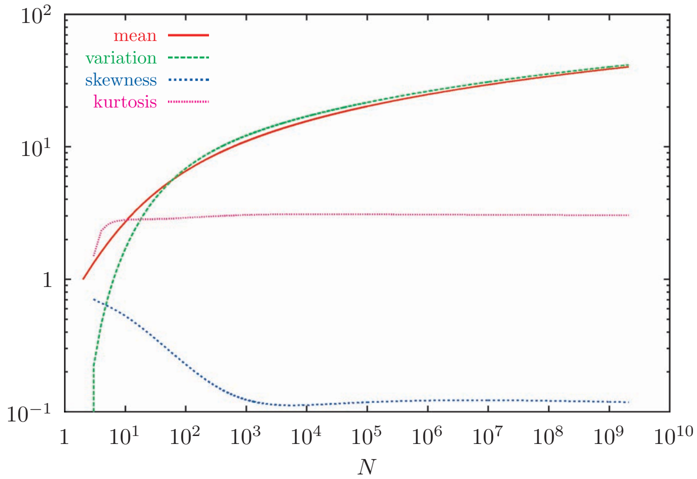

Figure 3. The NND distribution characteristics for exponential trees derived from iterative Equation (9)

Acknowledgements

K M's participation in the 37th Polish Physicists' Meeting in Gdansk was financed by the Cracow Branch of the Polish Physical Society (PTF). The numerical calculations were carried out at ACK-CYFRONET-AGH. The machine time on SGI 2800 was financed by the Ministry of Scientific Research and Information Technology in Poland under grant No. KBN/SGI2800/AGH/018/2003.

References

[1] Clark J and Holton D A 1991 *A First Look at Graph Theory*, World Scientific
[2] Wilson R J 1987 *Introduction to Graph Theory*, Longman Scientific and Technical, New York
[3] Berge C 1985 *Graphs*, North Holland
[4] Harary F 1969 *Graph Theory*, Addison-Wesley
[5] Dorogovtsev S N and Mendes J F F 2002 *Adv. Phys.* **51** 1079
[6] Dorogovtsev S N and Mendes J F F 2002 *From the Genome to the Internet* (Bornholdt S and Schuster H G, Eds), Viley-VCH, Berlin
[7] Dorogovtsev S N, Mendes J F F and Samukhin A N 2003 *Nucl. Phys.* **B 653** 307
[8] Newman M E J 2003 *SIAM Review* **45** 167
[9] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **286** 47
[10] Barabási A-L and Albert R 1999 *Science* **286** 509
[11] Malarz K, Czaplicki J, Kawecka-Magiera B and Kułakowski K 2003 *Int. J. Mod. Phys.* **C 14** (in print); *Preprint* cond-mat/0304636
[12] Malarz K and Kułakowski K 2003 *Preprint* cond-mat/0304693
[13] Stauffer D, private communication
[14] Milgram S 1967 *Psychol. Today* **2** 60
[15] Szabó G, Alava M and Kertész J 2002 *Phys. Rev.* **E 66** 026101