Node-node distance distribution for growing networks[1]

K.Malarz*, J.Karpińska, A.Kardas, and K.Kułakowski†

Faculty of Physics and Nuclear Techniques
AGH University of Science and Technology
al. Mickiewicza 30, PL-30059 Kraków, Poland

E-mail: *malarz@agh.edu.pl, †kulakowski@novell.ftj.agh.edu.pl

November 21, 2018

Abstract

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

1 Introduction

A graph is defined as a set of nodes (vertices) and a set of links between nodes (edges) [1, 2, 3, 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, 6, 7, 8, 9].

The distance between nodes is the shortest number of edges which leads from one node to the other. The node-node distance (NND) distribution depends on how subsequent nodes are attached. If each node is connected with *only one* of 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 in preferential way 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 *scale-free* or Albert–Barabási networks [10].

In this paper, the numerical algorithm for the network growth — basing on *distance matrix* evolution — is presented both, for exponential and scale-free networks ($m = 1, 2$) [11, 12]. The NND distribution and its characteristics are calculated. For the exponential trees the iterative formulas for n-th ordinary moments are derived.

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

[1] presented at the 37-th Polish Physicists' Meeting, Gdańsk, Poland, 15-19 Sep. 2003

i and j. This matrix representation is also particularly useful when computer simulations for graph evolution are applied.

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

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

Similarly, when new node is attached to the network with two edges ($m = 2$) to two different nodes — which are labeled as p and q — but the distance from all other nodes i to the newly added $(N+1)$-th is one plus the smaller distance between p–i or q–i nodes pairs [12]:

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

In the case mentioned above of the growth of the simple graphs also the reevaluation of distances between nodes i and j must be done to check if adding a new node provides the shortcut [12]:

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

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

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

Selecting 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 the 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 correspond to 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 only for non-zero distances

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

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_{run}

Table 1: The 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	0.672	1.00	0.462
b	-2.84	0.296	-0.08	0.889

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

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

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 Eqs. (5) and (6) are equal, due to obvious fact, that $s(i, i) = 0$. That allows to derive simple dependence between averages $\{\cdots\}$ and $\langle\cdots\rangle$:

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

For the exponential trees — basing on $s(i, i) = 0$ 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 ($k = 1, \cdots, n$)

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

where q is the number of the randomly selected row/column of the distance matrix $\mathbf{S}$. Combining Eq. (8) with Eq. (7) gives 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 the 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 Tabs. 1, 2) [11]. For the graphs only the first cumulant (the average of the NND d_N^1) grows logarithmically (see Tab. 2) [12]. Such a slow increase of d_N^1 with number of network nodes is known as the small-world effect [14].

The histogram of NND is presented in Fig. 1. As we expected the NND for the graphs are more condensed than the NND for the trees, as well as, the scale-free graphs (trees) are more condensed than the exponential graphs (trees).

Knowing the moments d_N^n — the averages of the n-th powers of the non-diagonal distance matrix elements (6) — allows to build all statistical parameters which characterize the NND distribution, e.g. the average distance d, the

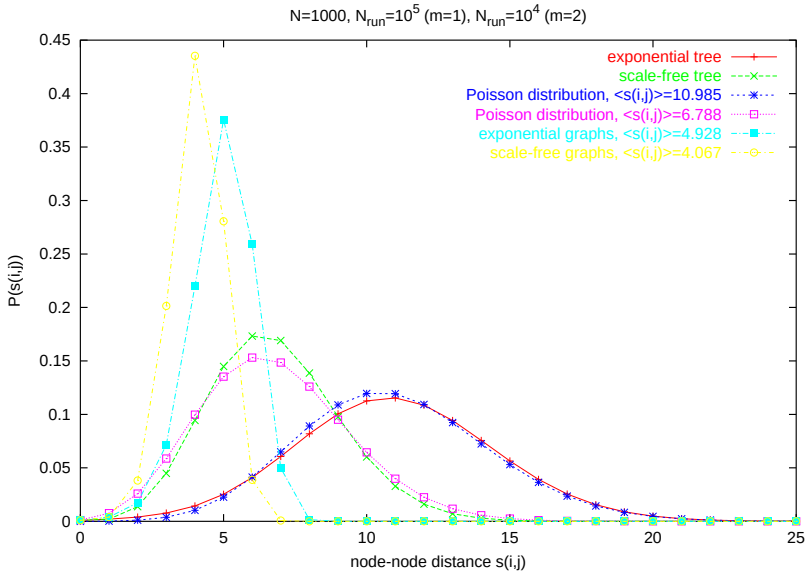

Figure 1: The NND distribution for different types trees and graphs.

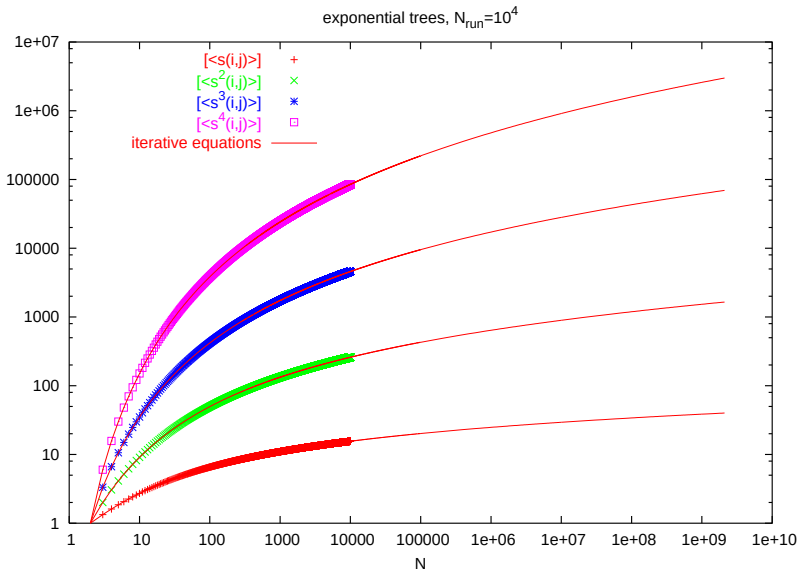

Figure 2: The main moments d_N^k $(k = 1, \cdots, 4)$ for the exponential trees given by Eq. (9) (lines) and from the direct simulations (symbols). The latter are averaged over $N_{\rm run} = 10^4$ independent evolution process realizations.

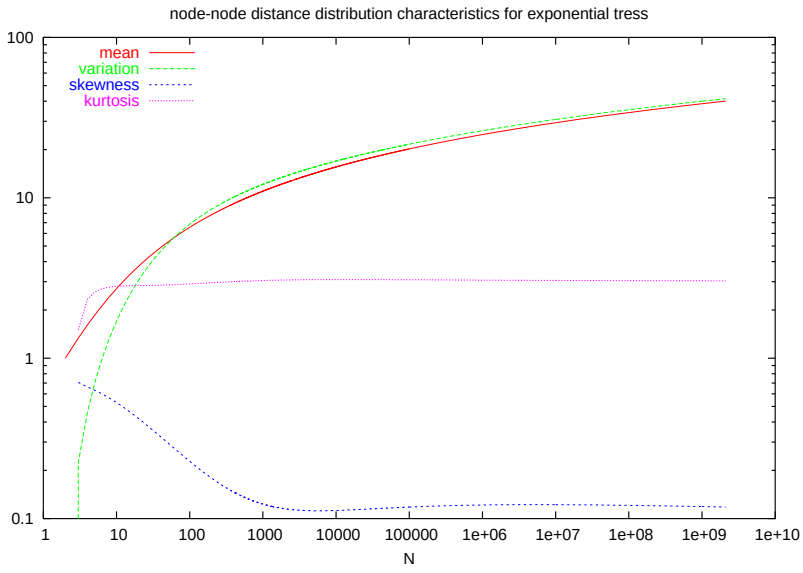

Figure 3: The NND distribution characteristics for the exponential trees as derived from iterative Eq. (9).

distance dispersion σ^2, its skewness

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

and kurtosis

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

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

Acknowledgments

K.M.'s participation in the 37-th Polish Physicists' Meeting in Gdańsk was financed by Kraków Branch of the Polish Physical Society (PTF). The numerical calculations were carried out in ACK–CYFRONET–AGH. The machine time on SGI 2800 is financed by the State Committee for Scientific Research (KBN) in Poland with grant No. KBN/SGI2800/AGH/018/2003.

References

[1] J.Clark, D.A.Holton, *A First Look at Graph Theory*, World Scientific, 1991.

[2] R.J.Wilson, *Introduction to Graph Theory*, Longman Scientific and Technical, New York, 1987.

[3] C.Berge, *Graphs*, North Holland, 1985.

[4] F.Harary, *Graph Theory*, Addison-Wesley, 1969.

[5] S.N.Dorogovtsev, J.F.F.Mendes, Adv. Phys. **51** (2002) 1079.

[6] S.N.Dorogovtsev, J.F.F.Mendes, in *From the Genome to the Internet*, eds. S.Bornholdt and H.G.Schuster, Viley-VCH, Berlin, 2002.

[7] S.N.Dorogovtsev, J.F.F.Mendes, A.N.Samukhin, Nucl. Phys. **B653** (2003) 307.

[8] M.E.J.Newman, SIAM Review **45** (2003) 167.

[9] R.Albert, A.-L.Barabási, Rev. Mod. Phys. **286** (2002) 47.

[10] A.-L.Barabási, R.Albert, Science **286** (1999) 509.

[11] K.Malarz, J.Czaplicki, B.Kawecka-Magiera, K.Kułakowski, Int. J. Mod. Phys. **C14** (2003) 1201.

[12] K.Malarz, K.Kułakowski, `cond-mat/0304693`.

[13] D.Stauffer, private communication.

[14] S.Milgram, Psychol. Today **2** (1967) 60.

[15] G.Szabó, M.Alava, J.Kertész, Phys. Rev. **E66** (2002) 026101.