I. INTRODUCTION

The problem of evolving networks \[1, 2, 3, 4, 5\] belongs to a new area of statistical physics with many interdisciplinary applications, from biology (sexual contacts, food webs, ecological networks) \[6\] via sociophysics (the strength of weak ties, terrorism, scientific collaborations, paper citations networks) \[7\], econophysics (agents' games and interactions, business contacts) \[8\], to computer science (Internet infrastructure and world wide web) \[9\].

In all these cases a central role is played by graphs which allow to describe networks mathematically. A graph is a set \(V\) of vertices and a set \(E\) of edges among vertices. A simple graph is a graph without loops (i.e. self-links) and without multiple edges. A forest is a graph without cycles (i.e. paths which start and end at the same node). Connected forest is called a tree \[1\]. The terminology depends on subject where graph theory is applied: vertices become nodes, actors and agents in computer science, socio- and econophysics, respectively. Edges are called links or interactions as well \[2, 3, 4\]. By evolving we mean adding subsequent nodes to an already existing graph with \(M\) links to \(M\) preexisting nodes. For \(M = 1\) a tree, and for \(M = 2\) a simple graph appears. Here evolution means growth.

Other evolution strategy is to modify an existing network without its growth, i.e. without increasing number of elements of the set \(V\). Formation of a network may take place via the rewiring procedure \[12\]: starting with a ring of \(N\) connected nodes — each of them having \(K\) nearest neighbors — we destroy a randomly selected link \(i \rightarrow m\) and create instead another link \(i \rightarrow n\).

Another strategy is to start with \(N\) nodes and \(L\) edges between them. A structure formed in this way is termed Erdős–Rényi graph \[13\]. Similar approach was proposed by Gilbert \[14\] where number \(N\) of nodes is fixed and a new link between each of \(N(N-1)/2\) pair is realized with given probability \(p\). For \(N \rightarrow \infty\) Gilbert and Erdős–Rényi models give the same results and \(p = 2LN^{-1}(N-1)^{-1}\). Graphs described above are called classical random graphs (CRG) \[14\]. The graph “thermalization” is a generalization of the rewiring strategy accompanied by Metropolis dynamics \[15\].

Let us denote \(P_a(m)\) the probability that a new node will be attached to an existing node \(m\). For the scale-free (preferential, Albert–Barabási) networks \[11\], \(P_a(m)\) is proportional to the node degree \(k(m)\) (i.e. number of edges which leads from/to \(m\)) \[2, 3, 4\]. For the exponential networks, \(P_a(m)\) is uniform \[2, 3\].

In Fig. \[1\] examples of scale-free, exponential and CRG networks are presented.

The aim of this work is a brief recapitulation of our recent numerical results, obtained with a simple algorithm of network growth, and a comparison of these results with some analytical predictions. The algorithm has been applied to the exponential networks, the scale-free networks and the classical random graphs. The algorithm was verified by a comparison of the results with exact iterative equations. In Sec. \[II\] our algorithm is explained. In Sec. \[III\] we present the numerical results on the degree and degree distributions and their correlations, which appear to be relevant for some search algorithms. The distance distributions are compared with theoretical formulas. Final conclusions (Sec. \[IV\]) close the text.

II. MATRIX REPRESENTATION AND THE GROWTH

A network containing \(N\) nodes is fully characterized by its adjacency matrix \(A\), elements of which give number of edges between nodes \(i\) and \(j\). In case of simple graphs — where multiple edges are forbidden — this matrix becomes binary: \(a_{N}(i,j) = 1\) if the nodes \(i, j\) are linked together, and \(a_{N}(i,j) = 0\) elsewhere. Absence of loops gives all diagonal elements equal to zero: \(a_N(i,i) = 0\).

In the distance matrix \(D\), the matrix element \(d_N(i,j)\) is the number of links along the shortest path from node
FIG. 1: Examples of networks for $N = 200$ and (a) scale-free: tree $(M = 1)$ and (b) simple graph $(M = 2)$, (c) exponential: tree $(M = 1)$ and (d) simple graph $(M = 2)$, (e) classical random graph: $p = 0.02$, (f) $p = 0.05$. (Figures using Pajek [16].)

The conversion $D \rightarrow A$ is trivial, as we need only change all elements larger than one to zero. For building the distance matrix $D$ basing only on the adjacency matrix $A$ is more complicated and several numerical techniques are available [17]. These methods usually base on the list approach with the breadth-first search or depth-first search or subsequent usage the Dijkstra algorithm [18] for all nodes.

Here, we present algorithms for different kind of net-
works which allow for a construction of the distance matrices simultaneously with the network growth, and not afterward \[12, 20, 21, 22\].

A. Numerical approach

For growing networks, the starting point is a matrix $D$ for the tree of two nodes linked together:

$$D_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Selecting a node $m$ to which a new node will be attached is equivalent to select a number $m$ of column/row of the matrix $D$. Approaches presented here base on the fact that the distance $d(n, i) = d(i, n)$ to a new node $n$ from all other preexisting nodes $i$ via node $m$ is $d(m, i) + 1 = d(i, m) + 1$.

1. Growing trees

Let us start with the simplest case, i.e. when $M = 1$. In this case a tree appears. Subsequent stages of distance matrix $D$ evolution $D_N \rightarrow D_{N+1}$ for $N \geq 2$ are

$$\forall 1 \leq i \leq N : d_{N+1}(N + 1, i) = d_{N+1}(i, N + 1) = d_N(m, i) + 1 \quad (1a)$$

and for diagonal element

$$d_{N+1}(N + 1, N + 1) = 0 \quad (1b)$$

since we do not allow for loops $\[14, 20\]$. One step of the distance matrix $D$ evolution for growing trees is presented in Fig. 2.

2. Growing simple graphs

In simple graphs cyclic paths are possible and the distance matrix $D$ is to be rebuilt when adding a new node $\[20, 22\]$. Suppose that a $(N + 1)^{th}$ node is added to existing nodes $m$ and $n \neq m$. Then, we have:

$$\forall 1 \leq i, j \leq N : d_{N+1}(i, j) = \min \{d_N(i, j), d_N(i, m) + 2 + d_N(n, j), d_N(i, n) + 2 + d_N(m, j)\}. \quad (2a)$$

For new, $(N + 1)^{th}$, column/row

$$\forall 1 \leq i \leq N : d_{N+1}(N + 1, i) = d_{N+1}(i, N + 1) = \min \{d_N(m, i), d_N(n, i)\} + 1 \quad (2b)$$

and again for the diagonal element

$$d_{N+1}(N + 1, N + 1) = 0. \quad (2c)$$

Example of the distance matrix $D$ evolution for a simple graph is presented in Fig. 3.

3. Classical random graphs

For CRG the starting point of simulations is an $N \times N$ matrix with all non-diagonal elements equal to $N$. Note that $N$ is larger than any possible distance in connected graphs of $N$ nodes. Now we go through all upper-diagonal elements of $D$ and set $d(i, j < i)$ equal to one with the probability $p$ — basing on the CRG’s definition. Obviously, the matrix $D$ is kept symmetric. Each time, when a new edge is added, we have to rebuild the whole matrix $D$ due to link between nodes $i$ and $j$:

$$\forall 1 \leq m, n \leq N : d(m, n) = \min \{d(m, n), d(m, i) + 1 + d(j, n), d(m, j) + 1 + d(i, n)\}. \quad (3)$$

After this procedure, the matrix $D_N$ contains elements equal to $N$ only if the graph is not connected $\[21\].

4. The Kertész list

Additional vector $r$ of nodes’ labels may be useful. There, each node’s label appears as an element of $r$ as many times as it is degree of that node (see Fig. 4). At each time when new $(N + 1)^{th}$ node is added with $M$ links to nodes labeled as $m_1, \ldots, m_M$, these $M$ labels are added to the list, as well as the new label $(N + 1)$, which is added $M$ times. By using a randomly chosen element of vector $r$ as an label, to which new node will be attached, we realize the Albert–Barabási rule $\[11\]$ of preferential growth $\[23\].$

III. RESULTS OF THE SIMULATIONS

When the evolution process is accomplished, the matrix $D_N$ may provide much information on the graph. Let us denote $\{\cdot \}, \{\cdot, \cdot\}$ and $\langle \cdot \rangle$ the averages over $N_\text{run}$ different matrices, $N$ nodes of network and $N^2$ matrix elements, respectively. For example, the average number of elements equal to $d$, denoted as $z_d$, in matrix $D$ gives the average number of $d^{th}$ neighbors for each node ($d = 1$ — nearest neighbors, $d = 2$ — next-nearest neighbors, $d = 3$ — next-next-nearest neighbors, etc.). The $i^{th}$ node degree $k_i$ is $\sum_{j=1}^{N} a_N(i, j)$ is the number of “1” in $i^{th}$ row/column of $D$ (and $z_1 = \{k\}$). Average distance to node $i$ from all other nodes

$$\xi_i = \big[ N^{-1} \sum_{j=1}^{N} d_N(i, j) \big].$$

The network diameter $\ell$ is the mean length of the shortest path between two vertices

$$\ell = \big[ N^{-1} \sum_{i=1}^{N} \xi_i \big] = \{\{\xi_i\}\} = \{(d_N(i, j))\].$$
A. Distribution of node degrees

Three kind of networks presented here derive their names from the distribution of node degrees:

- for the scale-free networks we reproduce $P_k(k) \propto k^{-\gamma}$ with $\gamma \approx 2.72$ ($M = 1$) and $\gamma \approx 2.63$ ($M = 2$), while the theoretical value is 3 and it is independent on $M$ [2, 11]. The numerical reduction of $\gamma$ is known to be caused by the finite-size effect.

- for the exponential trees the node degree distribution is verified to be $P_k(k) \propto w^{-k}$ [3] where $w = 2$ for $M = 1$ and $w = 3/2$ for $M = 2$.

- the degree distribution for CRG follows the Poisson distribution $P_k(k) = \exp(-\{k\}) \cdot \{k\}^k/k!$, with $\{k\} \approx 20$ and $\{k\} \approx 50$ when $N = 10^3$ for $p = 0.02$ and $p = 0.05$, respectively. Here, the average node degree may be evaluated as $\{k\} = p(N - 1)$ [13].

B. Distribution of node-to-node distances

Distribution of numbers $d$ in distance matrix $D$ gives node-to-node distance distribution (NNDD) $P_d(d)$. As expected, NNDD for the simple graphs are more condensed than NNDD for trees. Also, NNDD for the scale-free graphs (trees) are more condensed than NNDD for the exponential graphs (trees) [20] [see Fig. 6(a)]. As a rule, networks with more condensed NNDD are themselves more compact. For large trees, NNDD may be approximated [21] as $P_d(d) \propto d \exp(-Ad^2) \ldots$ — see Fig. 6(b).

For exponential trees — where each node is chosen with the same probability as a potential candidate to which the next attachment will take place — it is possible to derive [19, 20] exact iterative formula for the network diameter:

$$\langle d \rangle = N(N + 1) \ell(N + 1) = N(N + 2) \ell(N) + 2N \quad (4a)$$

and for all higher moments $n \geq 2$ of NNDD

$$(N + 1)^2 \langle [d_N^{k+1}(i, j)] \rangle = N(N + 2)\langle [d_N^{k+1}(i, j)] \rangle +$$

$$+ 2N \sum_{k=1}^{n-1} \binom{n}{k} \langle [d_N^k(i, j)] \rangle + 2N. \quad (4b)$$

C. Search for a shortest path

Shortest distance between nodes, averaged over the network is termed the network diameter. This diameter appears to be surprisingly short in the growing networks: it increases with the number of nodes $N$ as slowly
as $\log(N)$. This is known as the small world effect \cite{12}. For example, for over 800 million web pages in Internet (in 1999) you need, on average, only nineteen clicks to reach any of them starting with randomly selected home page of your browser \cite{27}. Table II shows the parameters $\alpha$ and $\beta$ in the logarithmic law $\ell(N) = \alpha \ln N + \beta$ for various growing networks \cite{19, 21, 22}, obtained numerically.

The small world effect has its sociological counterpart, discovered by Milgram \cite{26} in 60’s. Several persons had to send letters to a dealer in Boston, unknown to them, using their acquaintances. It was found that in the average, a chain only six links was sufficient to get the target. The essential point in the strategy is the search for the next person. A choice — almost obvious — is to select a person most popular and famous (i.e. the node with the
TABLE I: The mean distance $\ell(N) = \alpha \ln N + \beta$ for different evolving scale-free and exponential networks.

| $M$ | scale-free | exponential |
|-----|------------|-------------|
| $\alpha$ | 1.00 | 2.00 |
| $\beta$ | -0.08 | -2.84 |

This strategy is termed as the most connected neighbor search (MCNS). The dependence of the average distance $\xi(k)$ from the node on given degree $k$ to all other nodes for various networks is presented in Fig. 8a. The slope of the curve $\xi(k)$ brings an information, how this search strategy is effective for a given network. Thus, an effectiveness of MCNS for nodes of given $k$ can be evaluated by an index

$$\eta = -\frac{\partial \xi}{\partial \ln k}.$$  \hspace{1cm} (5)

The dependence $\eta(k)$ is presented in Fig. 8b). The MCNS is more efficient for the exponential trees than for the scale-free trees, and much more efficient for trees ($M = 1$) than for simple graphs ($M = 2$). For the simple graphs, this search strategy is almost as inefficient as for CRG. In the scale-free networks, local fluctuations of degree are enhanced by subsequent linkings. Multiple centers of high degree can be created, and the growing concentrates on these centuries. Then, MCNS can be misleading, as it leads always to a local center; however, sometimes the target is somewhere else. This enhancement is absent in the exponential networks, and that is why MCNS works better there. We note that this argumentation works particularly well for trees.

IV. SUMMARY

We have described the algorithm of construction the distance matrix $D$ of the exponential networks, the scale-free networks and the random networks. The core idea of the algorithm is that it works simultaneously with the network growth. The information coded in the distance matrix is equivalent to the information of the network structure. The algorithm’s complexity is of order of $N^2$ for trees, and of order of $N^3$ for other networks. A next step could be to construct a method of comparison of different networks by a comparison of their distance matrices, to check if the networks are topologically equivalent. However, the order of rows and columns of the distance matrix is set in accordance with the age of nodes, and this information is not preserved in the network topology. Then, to compare two distance matrices, one should shuffle all possible permutations of nodes. Such an algorithm is known to be non-polynomial, and therefore is not useful.

Topological properties of the networks are analyzed by discussing their distance matrices. These matrices are found to be a convenient tool to investigate the degree distributions, the distance distributions, the small-world effect and some search algorithms in the networks.

We believe that an understanding of topological properties of complex network is the first step to understanding complex behavior occurring among actors and agents occupying the networks nodes. It may bring a useful information for modelers of social and economic systems.

Acknowledgments

We thank J. Czaplicki, A. Kardas and J. Karpinska for their valuable help and Z. Burda for fruitful discussions. K.M. acknowledges support from Krakow Branch of the Polish Physical Society (PTF). Part of calculations was carried out in ACK-CYFRONET-AGH. The machine time on SGI 2800 is financed by the Polish Ministry of Science and Information Technology under Grant No. KBN/SGI2800/AGH/018/2003.
