NUMBERS OF n-TH NEIGHBORS AND NODE-TO-NODE DISTANCES IN GROWING NETWORKS*

Krzysztof Malarz†

Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Kraków, Poland

(Received July 7, 2005; revised version received December 14, 2005)

Topology of exponential and scale-free trees and simple graphs is investigated numerically. The numbers of the nearest neighbors, the next-nearest neighbors, the next-next-nearest neighbors, the 4-th and the 5-th neighbors are calculated. The functional dependence [A.E. Motter, T. Hishikawa, Y.-Ch. Lai, *Phys. Rev.* **E66**, 065103(R) (2002)] of the node-to-node distance d_{ij} on the product of connectivities $k_i k_j$ has been studied numerically. The results of simulations for exponential networks agree with the existing analytical predictions.

PACS numbers: 02.10.Ox, 05.10.–a, 07.05.Tp

1. Introduction

Complex networks have been attracting great attention for decades. They describe many real-world systems and are of interest for physics, sociology, biology, computer science, telecommunication and other areas of research [1, 2]. Mathematical description of networks is provided by graph theory [3]. Graph is a set of vertexes (nodes) connected by edges (links). The main local characteristic of a graph is the node degree, *i.e.* the number of links incoming to or outgoing from a node. For almost fifty years, the paradigm of "typical node" has been present in the science of networks. Networks of typical nodes were described by Erdős and Rényi [4] (*classical random graphs* — CRG). In their model, N nodes are connected randomly with L links: each inter-node link is realized with the probability $p = 2L/[N(N-1)]$. In this model, the node degree distribution is given

* Dedicated to Professor Andrzej Z. Maksymowicz on the occasion of his 65th birthday.

† homepage `http://home.agh.edu.pl/malarz/`, e-mail: `malarz@agh.edu.pl`

by a Poisson law, *i.e.* $P_k(k) = \exp(-\{k\})\{k\}^k/k!$, where $\{\cdots\}$ denotes the mean over all N nodes, and the node degrees observed on a graph fluctuate around $\{k\}$.

As pointed out by Albert and Barabási in their seminal paper [5], networks in real world more often exhibit a power-like degree distribution, *i.e.* $P_k(k) \propto k^{-\gamma}$. In the Albert–Barabási (A–B) model, the node degrees assume all integer values in the thermodynamic limit and there is no characteristic value of the degree. Thus, with this observation the Hungarian mathematicians' world of networks with typical nodes became a world of scale-free networks.

CRG and A–B networks are two examples belonging to two different families of networks [6]. The first one belongs to the class of *homogeneous* networks, which may be described via a statistical ensemble. The A–B networks have a temporal structure, as they come into being via growth process. The A–B networks are an example of *causal* networks.

For networks, the act of growing means subsequent attachments of new nodes, each with M links, to previously existing nodes. The procedure of selection of those "old" nodes influences the network topology and the resulting degree distribution. When old nodes are selected randomly with the same probability then *exponential* networks are created. This means that the resulting degree distribution is exponential [2]. On the other hand, when the attachment is preferential — *i.e.* the probability of choosing a node is proportional to its degree — the degree distribution is power-like and network can be termed as *scale-free* [1].

The number of edges M also influences the network topology:

- when $M = 1$, the path between any pair of nodes is unique; the growing structure is called *a tree*,
- when $M > 1$, cyclic paths are possible and *graph* looses its tree-like properties,
- when $M > 1$ and all nodes, to which new links are attached, are distinct a *simple graph*, without multiple edges, arises.

Such an attaching procedure produces no self-connecting links.

Several characteristics of real or simulated networks may be practically useful. For example many papers discuss the networks resistance to possible damage [7], the tolerance with respect to random and/or intentional attack [8] or transport properties in terms of either the percolation theory [9] or the shortest path finding [10, 11]. Newman *et al.* applied the generating function formalism [12] to evaluate the number of nodes

$$z_m = {z_1}^{2-m} {z_2}^{m-1} \tag{1}$$

in subsequent (m-th) layer from a randomly chosen origin [13]. In Eq. (1) z_1 and z_2 are typical values of the number of nodes nearest neighbors and the number of nodes next-nearest neighbors of a node, respectively. The first one (z_1) is obviously equal to the average node degree $z_1 = \{k_1\}$. The latter (z_2) was evaluated lately by Shargel *et al.* [14] as

$$z_2 = \{k^2\} - \{k\}. \tag{2}$$

Using the same technique, Motter *et al.* [15] derived the average node-to-node distance d_{ij} dependence on the product $k_i k_j$ of the node degrees for random networks:

$$\langle d_{ij} \rangle = A - B \ln(k_i k_j), \tag{3}$$

where $\langle \cdots \rangle$ denotes the average over all node pairs, the product of the pair degrees being equal to $k_i k_j$. Recently, Hołyst *et al.* [16] have confirmed this dependence numerically and presented some examples of real-world networks which obey the Motter *et al.* theoretical predictions.

In this paper we check if Motter *et al.* [15], Hołyst *et al.* [16] and Shargel *et al.* [14] predictions apply to the growing exponential networks. Namely, we evaluate number of neighbors in subsequent layers. The node-to-node distance *vs.* product of their degrees is also simulated. For completeness, the calculations and discussion include the growing scale-free A–B networks.

In the next section we explain our numerical approach. In Section 3 we present results of Monte Carlo simulations of the average number of subsequent neighbors (3.1) and the inter-nodes distance dependence on the product of node degrees (3.2). The last section summarizes the results.

2. Numerical approach

Numerical approach is based on an "on-line" construction of the distance matrix $\boldsymbol{D}$ during the network growth [17–20]. An element d_{ij} of the distance matrix gives the length of the shortest path between nodes i and j, *i.e.* the minimal number of edges which connect these vertexes. The numbers d_{ij} in i-th row/column inform how far is the node i from another node j. Then, the number $z_m(i)$ representing the number of occurrences of the m value in the i-th matrix row/column gives the information how many neighbors of the node i are at the distance m [20]. The average number of the matrix elements of a given value in all rows/columns — *i.e.* in the whole matrix — gives a typical number z_m of subsequent neighbors, for example, the nearest neighbors for $m = 1$, the next-nearest neighbors for $m = 2$, the next-next-nearest neighbors for $m = 3$, etc. Additionally, the number of unities ("1") in the i-th row/column gives degree of the i-th node: $k(i) = z_1(i)$.

3. Results of simulations

We construct the distance matrix $\boldsymbol{D}$ for $N = 10^3$ nodes. The results are averaged over $N_{\rm run} = 10^4$ independent simulations.

3.1. Number of nodes in subsequent layers

Fig. 1 shows how the deviation $\delta_m \equiv z_1^{2-m} z_2^{m-1} - z_m$ between z_m calculated from Eq. (1) and from the direct simulations behaves as the function of the system size N for $m = 3, 4, 5$. As one can see, starting with $N \approx 100$ this difference decreases with N for the exponential networks. For the scale-free networks either the number of nodes ($N = 10^3$) is still too small to observe a good agreement between z_m and $z_1^{2-m} z_2^{m-1}$ or Eq. (1) does not hold for the A–B graphs.

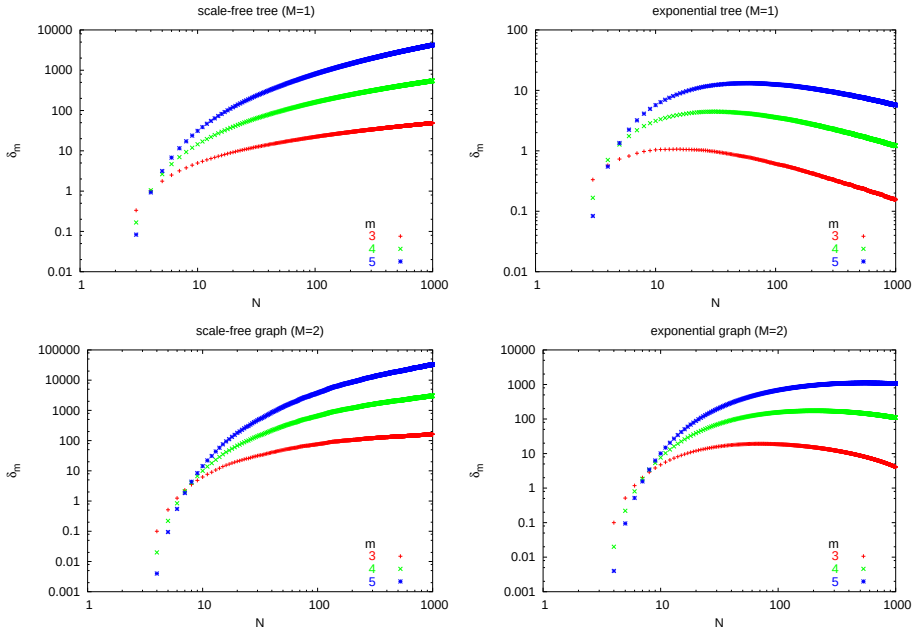

Fig. 1. Dependence of $\delta_m = {z_1}^{2-m} {z_2}^{m-1} - z_m$ $(m = 3, 4, 5)$ on network size N for growing exponential and scale-free trees ($M = 1$) and simple graphs ($M = 2$).

By construction, the average number of the nearest neighbors z_1 is 2 and 4 for trees and simple graphs, respectively. The number of next-nearest neighbors z_2 depends on the applied rules of growth: when the growth is governed by the preferential attachment rule, we have $z_2 \approx 14$ and $z_2 \approx 38$

for $M = 1$ and $M = 2$, respectively. For the exponential networks, these numbers are $z_2 \approx 4$ $(M = 1)$ and $z_2 \approx 17$ $(M = 2)$. As the average number of the nearest neighbors z_1 is exactly equal to the average nodes degree $\{k\}$, it may be evaluated from the degree distribution $P_k(k)$ as $\{k\} = \sum_{k=M}^{\infty} kP_k(k)$, as well. For the exponential network this distribution [2, 20] is given by

$$P_k(k \geq M) = \begin{cases} 2^{-k} & \text{for } M = 1, \\ 3/4 \cdot (3/2)^{-k} & \text{for } M = 2, \end{cases} \tag{4}$$

while for the scale-free networks [21, 22] it is

$$P_k(k \geq M) = \frac{2M(M+1)}{(k+2)(k+1)k}. \tag{5}$$

The mean number of the next-nearest neighbors (z_2) may be evaluated from Eq. (2) which diverges for scale-free networks with $N \to \infty$. For finite but large network this sum

$$\sigma \equiv \sum_{k=M}^{N-1} k(k-1)P_k(k) = 2M(M+1) \sum_{k=M}^{N-1} \frac{k-1}{(k+2)(k+1)} \tag{6}$$

grows logarithmically with N, $\sigma = 3.99 \ln N - 7.55$ $(M = 1)$ and $\sigma = 11.96 \ln N - 22.6$ $(M = 2)$ as presented in Fig. 2.

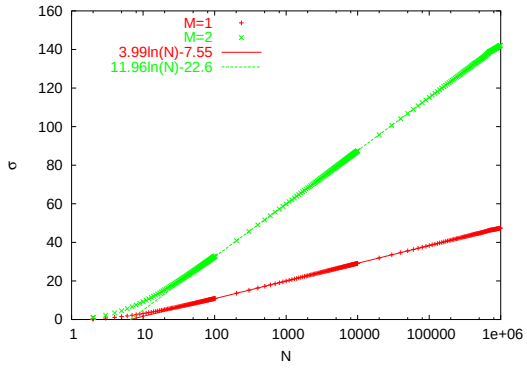

Fig. 2. Dependence $\sigma(N)$ for growing scale-free networks.

The results are collected in Table I.

TABLE I

Average number of the nearest neighbors z_1 and the next-nearest neighbors z_2 for different evolving scale-free and exponential networks with $N = 10^3$ nodes. The results are averaged over $N_{\rm run} = 10^4$ samples. Theoretical predictions of the average nodes degrees $\{k\}(= z_1)$ and $\{k^2\} - \{k\}(= z_2)$ are also included. Four last lines show the least-square fit coefficients A and B in the dependence $\langle d_{ij}\rangle = A - B\ln(k_i k_j)$ and their predictions $A^{\rm th}$ and $B^{\rm th}$ given by Eq. (7).

	scale-free		exponential	
M	1	2	1	2
$\sum_{k=M}^{\infty} kP_k$	2	4	2	4
$z_1 = \{k\}$	1.998	3.994	1.998	3.994
$\sum_{k=M}^{\infty} k(k-1)P_k$	∞	∞	4	18
$\sum_{k=M}^{N-1} k(k-1)P_k$	19.95	59.86		
$\{k^2\} - \{k\}$	13.68	39.66	3.966	17.81
z_2	13.68	38.11	3.966	17.37
z_2^2/z_1	93.6	363.6	7.88	75.6
z_3	44.8	201.5	7.72	71.5
A	7.68	5.09	12.8	6.77
B	0.783	0.438	1.73	0.746
$A^{\rm th}$	4.93	4.68	12.0	6.64
$B^{\rm th}$	0.519	0.443	1.46	0.679

3.2. Node-to-node distances and node degrees

Using the generating function formalism [12,13], Motter *et al.* [15] derived an expression for the length of the shortest path between the nodes for a given value of the product of connectivities $k_i k_j$:

$$\begin{aligned}\langle d_{ij}\rangle &= \left[1 + \frac{\ln(Nz_1)}{\ln(z_2/z_1)}\right] - \left[\frac{1}{\ln(z_2/z_1)}\right]\ln(k_i k_j) \\ &\equiv A^{\rm th} - B^{\rm th}\ln(k_i k_j)\,. \end{aligned} \tag{7}$$

Lately, such a kind of dependence $\langle d_{ij}\rangle$ *vs.* $(k_i k_j)$ has been shown to be valid in few real-world networks, including biological and scientific papers citation networks, public-transportation systems of several Polish towns, and simulated CRG and A–B networks [16].

Here we show that this logarithmic dependence (3) holds for exponential networks with $M = 1, 2$. The results are presented in Fig. 3 and in Table I. The least-square fit was confined to the two first decades of $k_i k_j$ values for the scale-free networks and to the one-and-half decade of $k_i k_j$ values for the exponential ones.

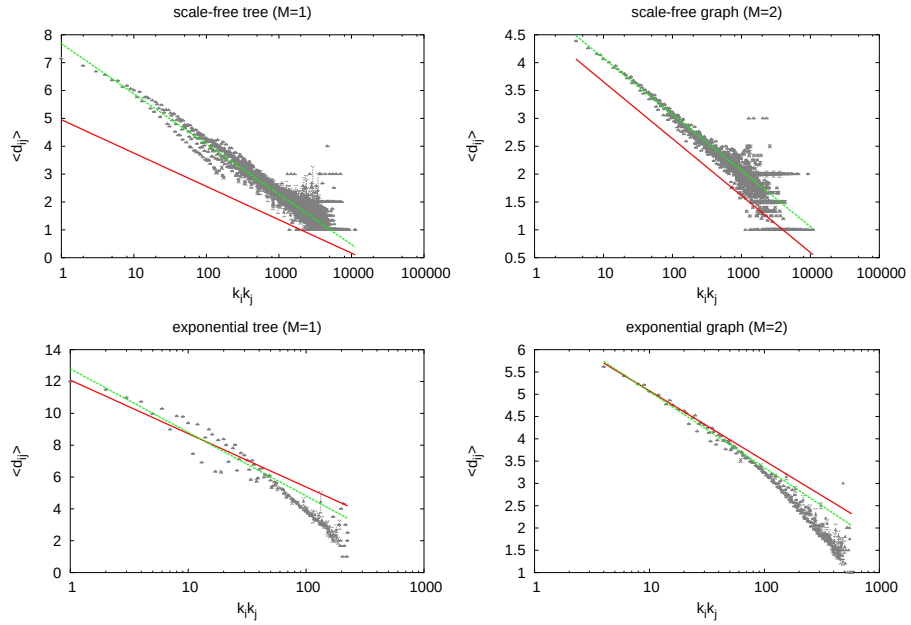

Fig. 3. Dependence $\langle d_{ij} \rangle$ vs $k_i k_j$ for growing exponential and scale-free trees ($M = 1$) and simple graphs ($M = 2$). The solid lines are the plots of Eq. (7) while the dotted lines result from a least-square fit. $N = 10^3$, $N_{\mathrm{run}} = 10^4$.

4. Discussion and conclusions

Generating function formalism [13–15,25,26] has a mean-field nature and should work only for homogeneous trees. This mechanism is founded upon the assumption that there are *no* correlations between nodes degrees. But this assumption does not hold for growing (causal) trees, where the oldest nodes — probably well connected — are geometrically close [27].

However, the Motter *et al.* formula (7) works surprisingly well also for growing networks, where triangles and other cyclic paths are possible. Results of numerical simulations agree with the Motter *et al.* [15] predictions particularly well for exponential networks. Also z_m is reproduced quite fairly in the simulation. For scale-free networks this agreement is only qualitative. It seems, that the theoretical predictions given by Eqs. (1) and (7) (and obviously given by Eq. (2)) agree with the results of simulations for only networks for which $\{k^2\}$ is finite.

The average number of vertexes in all generations, z_m, is well known for homogeneous [25] and causal [24] trees. The number z_m of m-th neighbors derived in [13, 15] agree very well for small $m = 3, 4, 5$ with the results of computer simulations for exponential networks where the old nodes, to which the new ones are attached, are chosen equiprobably.

Again, this should be valid for trees, but it works also for $M = 2$ when the random attachment is used. On the other hand the sum $\sum_{k=M}^{\infty} k\,(k-1)P_k$ (k) diverges for power-like distributions $P_k(k)$. For finite but large lattices this sum (σ, Eq. (6)) increases logarithmically with the system size N. In all four investigated cases the numerically computed value of z_2 (given by number of "2" in the distance matrix) agrees with $\{k^2\} - \{k\}$ (averaged over all graph nodes).

For larger m formula (1) fails when applied to real networks, *i.e.* with finite N. Usually, the second layer contains more nodes than the first one, from which follows that $z_2 > z_1$. Then — according to Eq. (1) — z_m increases with $m \in \mathbb{Z}$, but for finite systems it must start to decrease for large m. In particular, any of N nodes which constitute the network has no neighbors in N-th layer and does not posses N-th neighbors ($z_m = 0$ for $m \geq N$). The distribution of the node-to-node distances for the growing networks discussed here were presented in [18,20] and evaluated analytically for trees in Ref. [23].

Still, the method of evaluation of z_m ($m = 3, 4, 5$) based on Eq. (1) may be quite useful. The main effort should be given to a theoretical evaluation of the average number of nodes in the second layer, *i.e.* the number of occurrences of "2" in the distance matrix, basing only on the degree distribution $P_k(k)$. Such an evaluation of z_2 would allow, in principle, to reproduce the whole function z_m.

Although the node-to-node distance $\langle d_{ij} \rangle$ depends logarithmically on the product of the node i and j degrees (Fig. 3, Eq. (3), Ref. [16]), the dependence of the node-to-node distance on the node degrees is not trivial [11]. We have demonstrated, that Eq. (3) can be extended to the case of the growing exponential networks. The Motter *et al.* predictions for the coefficients A and B in Eq. (3) given by Eq. (7) agree for these networks quite fairly.

Author thanks Krzysztof Kułakowski, Zdzisław Burda and Andrzej Lenda for their valuable help and many fruitful discussions. Calculations were carried out in ACK-CYFRONET-AGH. The machine time on SGI 2800 is financed by the Polish Ministry of Education and Science under grant No. KBN/SGI2800/AGH/018/2003.

REFERENCES

[1] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003); S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, *Nucl. Phys.* **B653**, 307 (2003); R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **286**, 47 (2002).

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

[3] J. Clark, D.A. Holton, *A First Look at Graph Theory*, World Scientific, 1991; R.J. Wilson, *Introduction to Graph Theory*, Addison-Wesley, 1996.

[4] P. Erdős, A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960); P. Erdős, A. Rényi, *Publications Mathematicae* **6**, 290 (1959).

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

[6] Z. Burda, J.D. Correia, A. Krzywicki, *Phys. Rev.* **E64**, 046118 (2001); P. Białas, Z. Burda, B. Wacław, *AIP Conf. Proc.* **776**, 14 (2005); L. Bogacz, Z. Burda, W. Janke, B. Wacław, *Comp. Phys. Commun.* **173**, 162 (2005).

[7] P. Crucitti, V. Latora, M. Marchiori, A. Rapisarda, *Physica* **A320**, 622 (2003); W.H. Cunningham, *J. Assoc. Comput. Machinery* **32**, 549 (1985).

[8] G.-J. Lin, X. Cheng, Q. Ou-Yang, *Chinese Phys. Lett.* **20**, 22 (2003); L. Zonghua, L. Ying-Cheng, Y. Nong, *Phys. Rev.* **E66**, 36112 (2002); L. Zonghua, L. Ying-Cheng, Y. Nong, P. Dasgupta, *Phys. Lett.* **A303**, 337 (2002); S. N. Dorogovtsev, J.F.F. Mendes, R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **87**, 219802 (2001); S. N. Dorogovtsev, J. F. F. Mendes, R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **87**, 219801 (2001); R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001); R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).

[9] M. Molloy, B. Reed, *Random Structures and Algorithms* **6**, 161 (1995); M. Molloy, B. Reed, *Combinatorics Prob. Comput.* **7**, 295 (1998); M.E.J. Newman, I. Jensen, R.M. Ziff, *Phys. Rev.* **E65**, 021904 (2002); D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000); C. Moore, M.E.J. Newman, *Phys. Rev.* **E61**, 5678 (2000); C. Moore, M.E.J. Newman, *Phys. Rev.* **E62**, 7059 (2000).

[10] B.J. Kim, C.N. Yoo, S.K. Han, H. Jeong, *Phys. Rev.* **E65**, 027103 (2002).

[11] K. Malarz, K. Kułakowski, *Eur. Phys. J.* **B41**, 333 (2004).

[12] H.S. Wilf, *Generatingfunctionology*, Academic Press, London 1994.

[13] M.E.J. Newman, S.H. Strogatz, D.J. Watts, *Phys. Rev.* **E64**, 026118 (2001).

[14] B. Shargel, H. Sayama, I.R. Epstein, Y. Bar-Yam, *Phys. Rev. Lett.* **90**, 068701 (2003).

[15] A.E. Motter, T. Hishikawa, Y.-Ch. Lai, *Phys. Rev.* **E66**, 065103(R) (2002).

[16] J.A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, K. Suchecki, *Physica* **A351**, 167 (2005); J. Sienkiewicz, J.A. Hołyst, *Acta Phys. Pol. B* **36**, 1771 (2005).

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

[18] K. Malarz, J. Karpińska, A. Kardas, K. Kułakowski, *TASK Quarterly* **8**, 115 (2004).

[19] K. Malarz, K. Kułakowski, *Physica* **A345**, 326 (2005).

[20] K. Malarz, K. Kułakowski, *Acta Phys. Pol. B* **36**, 2523 (2005).

[21] P.L. Krapivsky, S. Redner, F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).

[22] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, *Phys. Rev.* **E64**, 066110 (2001).

[23] Z. Burda, J. Erdmann, B. Petersson, M. Wattenberg, *Phys. Rev.* **E67**, 026105 (2003).

[24] P.L. Krapivsky, S. Redner, *Phys. Rev.* **E63**, 066123 (2001).

[25] J. Ambjørn, B. Durhuus, J. Fröhlich, P. Orland, *Nucl. Phys.* **B270**, 457 (1986).

[26] J. Ambjørn, B. Durhuus, T. Jónsson, *Phys. Lett.* **B244**, 403 (1990).

[27] Z. Burda, priv. commun.