Scale-free network on Euclidean space optimized by rewiring of links

2003 J. Phys. A: Math. Gen. 36 L279

(http://iopscience.iop.org/0305-4470/36/19/101)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) L279–L285 PII: S0305-4470(03)59558-0

LETTER TO THE EDITOR

Scale-free network on Euclidean space optimized by rewiring of links

S S Manna[1,2] **and A Kabakçıoğlu**[1]

[1] INFM—Dipartimento di Fisica, Università di Padova, I-35131 Padova, Italy
[2] Satyendra Nath Bose National Centre for Basic Sciences, Block-JD, Sector-III, Salt Lake, Kolkata-700098, India

Received 12 February 2003
Published 29 April 2003
Online at stacks.iop.org/JPhysA/36/L279

Abstract
A Barabási–Albert scale-free network is constructed whose nodes are the Poisson distributed random points within a unit square and links are the straight line connections among the nodes. The cost function, which is the total wiring length associated with such a network defined on a two-dimensional plane, is optimized. The optimization process consists of random selection of a pair of links and rewiring them to reduce the total length of the pair but with the constraint that the degree as well as the out-degree and in-degree of each node are precisely maintained. The resulting optimized network has a small diameter as well as high clustering and the link-length distribution has a stretched exponential tail.

PACS numbers: 02.50.−r, 89.20.−a, 89.75.−k, 89.75.Hc

The Internet is a very complex network connecting a large number of computers around the world [1, 2]. The nodes of this network may be interpreted as the routers and links as the cables connecting computers. The network can also be described in the inter-domain level where each domain is represented by a single node and each link is an inter-domain connection. Such a network is well described by a graph consisting of a set of vertices and another set of edges among the vertices. Without assigning any weight to the links between the nodes only the topological structure of the Internet is meaningful. Study of the Internet's topological structure may be important for designing efficient routing protocols and modelling Internet traffic.

The Internet is one of the large class of real-world networks that exhibit small-world and/or scale-free properties, e.g., social networks [3], biological networks [4, 5], electronic communication networks [1, 2], etc. Quantities that characterize a network of N nodes are the diameter $\mathcal{D}(N)$ which measures the topological extension of the network, the clustering coefficient $\mathcal{C}$ which measures the local correlations among the links of the network and the nodal degree distribution $\mathcal{P}(k)$. In a small-world network (SWN) [6], the diameter $\mathcal{D}(N)$ of the network scales logarithmically with N whereas for a scale-free network (SFN) the degree distribution has a power law tail: $\mathcal{P}(k) \propto k^{-\gamma}$. Barabási and Albert (BA) showed that a growing network with preferential link attachment probability is a SFN with $\gamma = 3$ [7].

Waxman first studied probabilistic graph models of the Internet where links have weights which are their physical lengths [8]. The link-length distribution in such networks decays exponentially: $D(\ell) \sim \exp(-\ell/\ell_0)$. Faloutsos *et al* observed that the out-degree distribution of the Internet follows a power law tail [1]. Yook *et al* observed that the distribution of routers of North America is a fractal set and the link-length distribution is inversely proportional to the link-lengths [9]. It is suggested that in the growing Internet, when a new node becomes a member of the network, two competing factors control the decision to which node of the already grown Internet the new node will be connected. The factors are the degree k_i of the existing node i and in general the αth power of the length ℓ of the link connecting the new node and the node i. The preferential attachment probability for the ith node is therefore: $\pi_i \propto k_i \ell^{\alpha}$.

Recently it has been argued that such a network is scale-free for all values of $\alpha > \alpha_c = 1 - d$ in d dimensions and the link-length distribution generally follows a power law $D(\ell) \sim \ell^{\delta}$ where $\delta(\alpha) = \alpha + d - 1$ [10, 11]. For $\alpha < \alpha_c$, the degree distribution decays stretched exponentially but $D(\ell)$ still maintains a power law where δ saturates at $-d - 1$. The limit of $\alpha \to -\infty$ is interesting where each node connects only to its nearest earlier node. In a regular network in the form of a linear chain, similar studies have been done [12]. Interplay between the preferential attachment and the link-length selection within an interaction range for the Euclidean networks is studied in [13].

In this paper, we associate a cost function associated with such networks. Each link of the network has a cost equal to its Euclidean length ℓ and therefore the cost function of the whole network is the total length of all the links of the network. The question we ask is, how can one construct a small-world scale-free network with minimal cost? To study this we start generating an N-node BA SFN on a two-dimensional plane. Links are then interchanged to reduce the cost function keeping the topology, i.e., the degree value of each node, intact. The optimization of the wiring length of networks on lattices has been studied in [14].

We start by constructing a Barabási–Albert SFN embedded in the Euclidean space as follows. Let $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$ be the independent identically and uniformly distributed random variables on the interval [0, 1]. To construct one random configuration of the network let a specific set of values of the N pair variables $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ be the coordinates of the set of N points on the unit square representing the set of nodes of the network with serial numbers $i = 1$ to N assigned to them. We use first $m + 1$ points with serial numbers $i = 1$ to $m + 1$ to construct a $(m + 1)$-clique by connecting each point with rest of the m points. Then following the serial numbers new points are added to the network one after another and each node is connected to randomly selected m distinct previous nodes. The probability of linking the new node with serial number j to a previous node i is linearly proportional to its degree, k_i. The network thus constructed up to N nodes is exactly the BA network [7]. At the same time it is a small-world network, i.e., the diameter $\mathcal{D}(N)$ of the network measured by the maximal distance between an arbitrary pair of nodes grows as $\log(N)$. In this paper we restrict ourselves to $m = 2$.

Let $\mathbf{a}$ denote the symmetric adjacency matrix of size $N \times N$ for our network such that $a_{ij} = 1$ if there is a link between the pair of nodes i and j and 0 otherwise. Let ℓ_{ij} denote the shortest Euclidean distance between the pair of nodes i and j taking into account the periodic boundary condition. Therefore, when $a_{ij} = 1$, ℓ_{ij} is the length of the connecting wire of the link between i and j. The total cost function $\mathcal{L}(N)$ is therefore the sum over all link-lengths of the network, i.e., $\mathcal{L}(N) = \sum_{i>j} a_{ij}\ell_{ij}$.

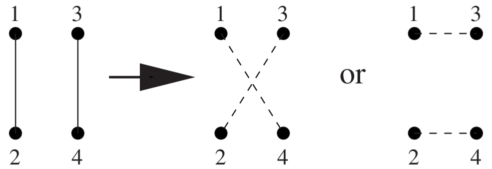

Figure 1. Two possible rewirings of a pair of links to reduce the total length of the two links.

For the convenience of discussion we define a degree vector similar to the contact vector generally used in polymer physics. Our degree vector **c** describes the topological connectivity of the network and has N elements $c_i = k_i$, the degree of the ith node. In the initial BA scale-free network one can associate a notion of time as if nodes are introduced one at each time unit. Therefore, the links of the node i introduced at time i are divided into two groups: 'outgoing' and 'incoming'. Each node has only $k^{\text{out}} = m$ outgoing links connected to m other nodes which are older than this node and it can be connected to $k^{\text{in}} = k - m$ other nodes which are younger than this node. Consequently, the degree vector can be split into two other degree vectors $\mathbf{c}^{\text{out}}$ and $\mathbf{c}^{\text{in}}$ such that $c_i^{\text{out}} = k_i^{\text{out}}$ and $c_i^{\text{in}} = k_i^{\text{in}}$ and $\mathbf{c} = \mathbf{c}^{\text{out}} + \mathbf{c}^{\text{in}}$.

Next, we perform the optimization dynamics to minimize the total cost function $\mathcal{L}(N)$. The optimization dynamics conserves the number of links in the network; in addition, it not only maintains the same degree vector **c** but also $\mathbf{c}^{\text{out}}$ and $\mathbf{c}^{\text{in}}$ separately and thus ensures that the out and in degree distributions of the network remain exactly the same as they were before the optimization process started. We call this 'time-ordered' rewiring. One trial of rewiring in the optimization scheme consists of selecting four nodes n_1, n_2, n_3 and n_4. The first node n_1 is randomly selected from the set of N nodes. n_2 is selected randomly from the k_1 neighbours of n_1. Similarly n_3 ($\neq n_1 \neq n_2$) is selected randomly from N nodes and n_4 ($\neq n_1 \neq n_2$) is again one of the k_3 neighbours of n_3. The move must maintain the conservation of link numbers as well as degree distribution. We replace the link pair n_1n_2 and n_3n_4 by another pair of links if either of the following two conditions is satisfied:

(i) if $a_{13} = a_{24} = 0$ and $\ell_{12} + \ell_{34} > \ell_{13} + \ell_{24}$, we link n_1n_3 and n_2n_4
(ii) if $a_{14} = a_{23} = 0$ and $\ell_{12} + \ell_{34} > \ell_{14} + \ell_{23}$, we link n_1n_4 and n_2n_3.

If both are satisfied we accept one of them with probability $1/2$. If only one is satisfied we accept that (figure 1). If neither of the two is satisfied we go for a fresh trial. We also study a second type of rewiring process where only the total degree vector **c** is maintained but not individually $\mathbf{c}^{\text{out}}$ and $\mathbf{c}^{\text{in}}$. Here in the final optimized network, a particular node may have all neighbours which are younger than this node. We call this process the 'random' rewiring method. Note that time-ordering always rules out one of the two conditions above, so the case of accepting (i) or (ii) with probability 1/2, in effect, applies to 'random' rewiring only. In figure 2 we show how an initially complicated network becomes less messy with increasing number of rewiring trials.

Since we accept the move only if the total rewired cost is reduced, the trial is similar to the zero-temperature Monte Carlo dynamics. The total cost $\mathcal{L}(N)$ monotonically decreases with the number of successful trials and the number of unsuccessful trials between successive accepted moves increases. We typically try around $10(mN)^2$ such trials so that the plot of $\mathcal{L}(N)$ with logarithm of the number of trials nearly reaches a plateau.

From each point one can measure $N - 1$ distances and if these distances are sorted in an increasing sequence, one has the first-neighbour distance, second-neighbour distance, ... $(N - 1)$th-neighbour distance, etc. It is known that the average nth-neighbour distance $\langle R_N^n \rangle$

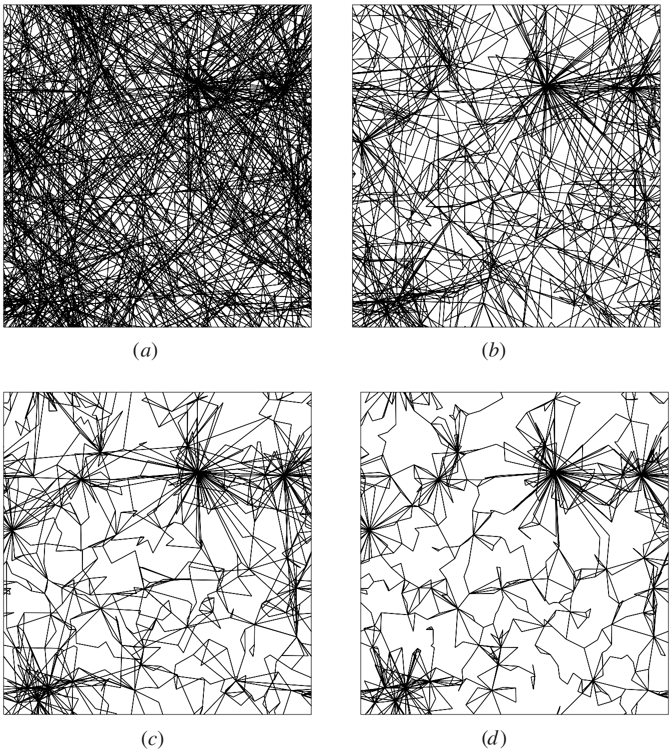

Figure 2. Snapshots of the optimized network generated by the random rewiring process for $N = 512$ and $m = 2$. (*a*) The initial network with $\mathcal{L} \approx 395$, (*b*) $\mathcal{L} \approx 304$ at $t = 1000$, (*c*) $\mathcal{L} \approx 170$ at $t = 10\,000$ and (*d*) $\mathcal{L} \approx 68$ at $t = 10\,000\,000$ where t is the number of rewiring trials.

varies as $N^{-1/2}$ if $n/N \ll 1$ and it is of the order of 1 when $n/N \sim 1$ in the limit of $N \to \infty$ [15]. There is no other variation like N^{-x} when x is neither 0 nor 1/2 but lies in between.

In the optimized network, the links are not necessarily a fixed (n) neighbour distance but a complex mixture of many neighbour distances. More elaborately, it is expected that many of the links of the optimized network are first-neighbour distances, a smaller number are second-neighbour distances, even fewer are third-neighbour distances, etc. In the optimized network we first calculate the probability density of the link-length distribution $D(\ell)$. This distribution on scaling by the average link-length $\langle \ell(N) \rangle$ is nearly the same for the time-ordered as well as the random rewiring processes. In contrast to expectation, this distribution has a maximum and it fits very well to a functional form $D(\ell)\langle \ell(N) \rangle \sim x^{\alpha}\, \mathrm{e}^{-ax^{\beta}}$ with $x = \ell/\langle \ell(N) \rangle$. The fit on a linear scale gives $\alpha = 1.4$, 1.1 and $\beta = 0.8$, 0.9 approximately for the time-ordered and random rewiring processes, respectively. The network has $N_\ell = 2N - m - 1$ links and therefore $\mathcal{L}(N) = N_\ell \langle \ell(N) \rangle$. We plot in figure 3 $\langle \ell(N) \rangle$ with $N/\log N$ and observe excellent straight lines on a double logarithmic scale. Therefore, $\langle \ell(N) \rangle \sim (N/\log N)^{\mu}$ where $\mu = 0.46$ and 0.52 with an error of 0.05 approximately for the time-ordered and random rewiring processes, respectively.

The topological size of the network is measured by the diameter of the network. The distance d_{ij} between an arbitrary pair of nodes i and j is the number of links on the shortest path connecting the two nodes. The diameter d_m is the maximal distance on a network. The average diameter $\mathcal{D}(N)$ represents the configuration averaged maximal distance $\langle d_m \rangle$. Variations of

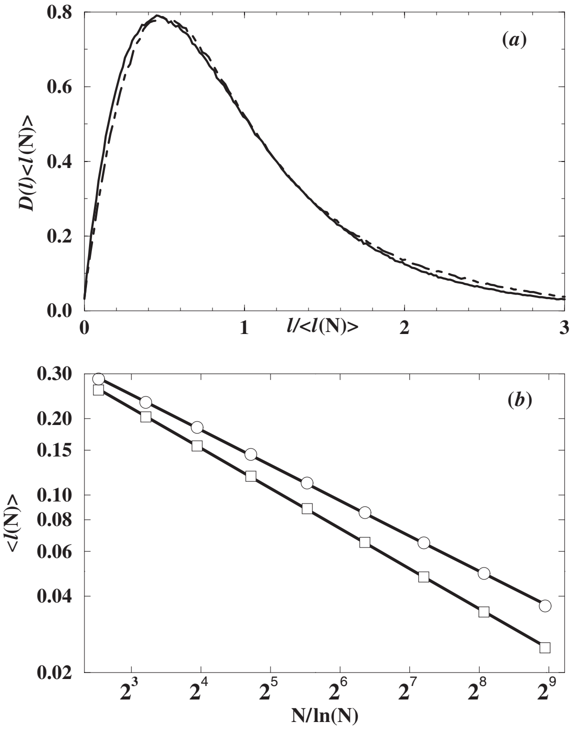

Figure 3. (*a*) The link-length distribution $D(\ell)$ in the optimized network: time-ordered (solid line) and random (dashed line) scaled by the average link-length $\langle \ell(N) \rangle$ for $N = 1024$. (*b*) The average length $\langle \ell(N) \rangle$ varies as $(N/\log N)^{\mu}$ in the optimized network: time-ordered (circles) and random (squares) rewirings.

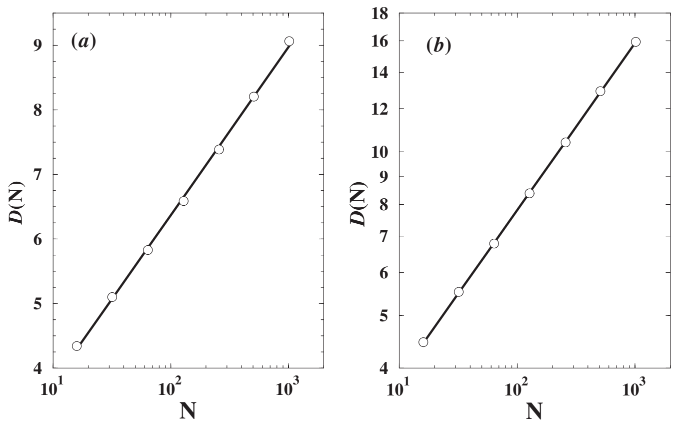

Figure 4. The average diameter of the network $\mathcal{D}(N)$ after optimization, as a function of number of nodes N in the network: (*a*) for the time-ordered exchange $\mathcal{D}(N) = A + B \log N$ where $A = 1.22$ and $B = 1.11$, (*b*) for the random exchange $\mathcal{D}(N) \sim N^{\nu}$ where $\nu = 0.31 \pm 0.04$.

the average diameter of the optimized network with the network size are shown in figure 4. For the the time-ordered exchange $\mathcal{D}(N) = A + B \log N$ with $A \approx 1.22$ and $B \approx 1.11$ whereas for the random exchange $\mathcal{D}(N) \sim N^{\nu}$ with $\nu = 0.31 \pm 0.04$, where the error 0.04

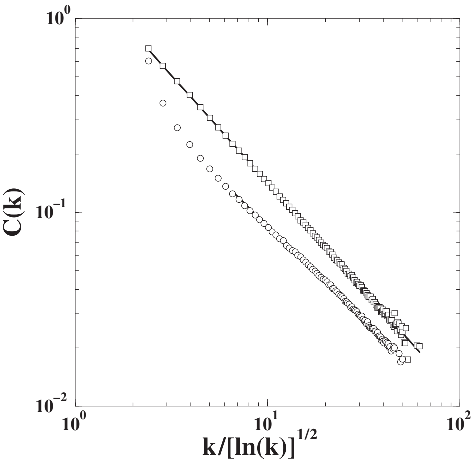

Figure 5. The clustering coefficient $\mathcal{C}(k)$ as a function of the degree k shows a $\{k/\{\ln(k)\}^{1/2}\}^{-b}$ behaviour with $b \approx 0.94$ and 1.1 for the time-ordered (circles) and random (squares) rewiring processes, respectively.

is estimated by the largest difference of the local slopes between successive points and the mean slope. Although the degree distribution remains scale-free in the optimized networks generated by both the time-ordered and random rewiring procedures, the first network retains some long-range connections due to the constraint that both $\mathbf{c}^{\text{out}}$ and $\mathbf{c}^{\text{in}}$ are strictly maintained whereas in the second network by random rewirings essentially all the connections are local, i.e., typically a node has all neighbours within a spatial distance of the order of $\sim N^{-1/2}$.

The local correlation among the links is measured by the clustering coefficient. The clustering coefficient $\mathcal{C}_i$ of the ith node is measured by the ratio of the number of links e_i within the k_i neighbours of the ith node and the number of links $k_i(k_i - 1)/2$ if the k_i nodes have formed an k_i-clique, i.e., $\mathcal{C}_i = 2e_i/\{k_i(k_i - 1)\}$. The clustering coefficient of the whole network $\mathcal{C}(N)$ is $\langle\bar{C}\rangle$. Also the average clustering coefficient for the set of nodes of degree k is defined as $\mathcal{C}(k)$. In general both these clustering coefficients may decrease as power laws: $\mathcal{C}(N) \sim N^{-a}$ and $\mathcal{C}(k) \sim k^{-b}$. In our case we start from the initial BA network where it is known that $a \approx 3/4$ and $b = 0$ [7]. We also calculate these quantities in the final optimized state. The total clustering coefficient is found to be independent of N and therefore $a = 0$ whereas unlike a simple power law for $C(k)$ we get a power law with logarithmic correction. In figure 5 we plot $\mathcal{C}(k)$ with $k/\{\ln(k)\}^{1/2}$ and observe straight lines on a double logarithmic scale implying the variation as

$$\mathcal{C}(k) \sim \{k/\{\ln(k)\}^{1/2}\}^{-b} \tag{1}$$

where $b \approx 0.94$ and 1.1 for the time-ordered and random rewiring processes, respectively. We cannot rule out the possibility that b is actually 1 for both processes. Many networks and models show $b = 1$ [7, 16].

To summarize, we have studied a cost optimized network which has three main features of the real-world networks, e.g., it is a small-world network, it is a scale-free network and also it exhibits high clustering properties as well. We studied this network on the two-dimensional Euclidean space which should be relevant in the context of the Internet. While some links in the Internet are cableless (microwave) links, many connections are made by real physical Ethernet cables. Therefore, the question of optimizing the cost of the total wiring length of the network arises naturally, which is the main point of study in this paper. An optimized geographical embedding algorithm for scale-free networks was recently studied independently [14]. Unlike in [14], our time-ordered optimization produces a statistically non-homogeneous network and preserves a significant number of long-distance connections, permitting the network diameter to still scale as $\log(N)$ as $N \to \infty$. We also obtain a stretched-exponentially decaying tail of the link-length distribution in the optimized network which is unlike the power-law tail observed by Yook *et al* [9] and closer to the Waxman result [8].

Acknowledgment

SSM thankfully acknowledges P Sen, P Bhattacharyya and J Kertész for critical reading of the manuscript and the Dipartimento di Fisica, University of Padova for hospitality. AK gratefully acknowledges the financial support of MIUR through COFIN 2001.

References

[1] Faloutsos M, Faloutsos P and Faloutsos C 1999 *Proc. ACM SIGCOMM, Comput. Commun. Rev.* **29** 251
[2] Lawrence S and Giles C L 1998 *Science* **280** 98
Lawrence S and Giles C L 1999 *Nature* **400** 107
Albert R, Jeong H and Barabási A-L 1999 *Nature* **401** 130
[3] Newman M E J 2001 *Proc. Natl Acad. Sci.* **98** 404 (*Preprint* cond-mat/0011155)
[4] Jeong H, Mason S P, Barabási A-L and Oltvai Z N 2001 *Nature* **411** 41
[5] Solé R V and Montoya J M 2001 *Proc. R. Soc.* B **268** 2039
Camacho J, Guimerá R and Amaral L N 2002 *Phys. Rev. Lett.* **88** 228102
[6] Watts D J and Strogatz S H 1998 *Nature* **393** 440
Watts D J 1999 *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton, NJ: Princeton University Press)
[7] Barabási A-L and Albert R 1999 *Science* **286** 509
Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[8] Waxman B 1988 *IEEE J. Sel. Areas Commun.* **6** 1617
[9] Yook S-H, Jeong H and Barabási A-L 2001 *Preprint* cond-mat/0107417
[10] Manna S S and Sen P 2002 *Phys. Rev.* E **66** 066114
[11] Sen P and Manna S S 2003 *Preprint* cond-mat/0301617
[12] Xulvi-Brunet R and Sokolov I M 2002 *Preprint* cond-mat/0205136
[13] Barthélemy M 2002 *Preprint* cond-mat/0212086
[14] Rozenfeld A F, Cohen R, b-Avraham D and Havlin S 2002 *Phys. Rev. Lett.* **89** 218701
b-Avraham D, Rozenfeld A F, Cohen R and Havlin S 2003 *Preprint* cond-mat/0301504
[15] Bhattacharyya P and Chakrabarti B K 2002 *Preprint* math.PR/0212230
[16] Szabó G, Alava M and Kertész J 2002 *Preprint* cond-mat/0208551