Growing networks with geographical attachment preference: Emergence of small worlds

Jonathan Ozik,[1,*] Brian R. Hunt,[2] and Edward Ott[3]
[1]*Department of Physics and Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20742, USA*
[2]*Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*
[3]*Department of Physics, Department of Electrical and Computer Engineering, and Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20742, USA*
(Received 30 July 2003; published 23 February 2004)

We introduce a simple mechanism for the evolution of small world networks. Our model is a growing network in which all connections are made locally to geographically nearby sites. Although connections are made purely locally, network growth leads to stretching of old connections and to high clustering. Our results suggest that the abundance of small world networks in geographically constrained systems is a natural consequence of system growth and local interactions.

DOI: 10.1103/PhysRevE.69.026108 PACS number(s): 02.50.Cw, 05.10.−a, 89.75.−k, 05.45.Pq

I. INTRODUCTION

Recently there has been considerable interest in the classification of physical systems according to the topological properties of the networks to which they map (e.g., [1–10]), where the constituent parts are modeled as nodes and links between nodes denote some type of interaction (for reviews see [11–14]). This method of classification has the potential to shed light on underlying organizational principles. In this spirit, we focus on the ''small world'' network topology introduced by Watts and Strogatz [1].

Here we represent a network as an undirected graph: a collection of N points (*nodes*) with connections (*links*) between some pairs of them. If two nodes are connected, we say that they are *neighbors*. We call the number of connections to node i the *degree* of node i and we denote it k_i.

Small world networks are characterized by two main properties. First, their characteristic path length L grows as $\ln N$ or slower, similar to an Erdős-Rényi (ER) random network. The characteristic path length is the smallest number of links connecting a pair of nodes, averaged over all pairs of nodes. Second, the network has a high average clustering compared to an ER random network of equal size and average node degree. The clustering C_i of node i is defined by $C_i = q_i/[(1/2)k_i(k_i-1)]$, where q_i is the total number of links between the k_i neighbors of node i, and $(1/2)k_i(k_i-1)$ is the maximum number of links that could exist between k_i nodes. Networks exhibiting small world characteristics are found in many and varied fields of research. Some examples of such networks are the neuronal network of the worm *C. elegans*, the electric power grid of southern California, and the friendship network of Madison Junior High School students [2].

The Watts and Strogatz model is the following prescription for creating a small world network. The initial state has a fixed number of nodes equally spaced on the circumference of a circle. Each node is linked to its m nearest neighbor nodes, where m is even and nearest here refers to the distance along the circumference of the circle. In this way a regular network with a large average clustering is created. Next, a proportion p of the links are chosen at random and ''rewired'' such that one end of the link is kept fixed and the other end is linked to a randomly chosen node. These random links can serve as short cuts across the circle, drastically decreasing the characteristic path length of the network. It was found [1] that, for a relatively small rewiring probability p, the characteristic path length of the network becomes comparable to that of an ER random network, while the network still maintains a high average clustering.

The network construction of the Watts and Strogatz model very nicely illustrates the small world property and, furthermore, it is probably a reasonable model for how some networks are formed. However, the small world property is, of course, much more general than their particular example, and it is useful to study other mechanisms for forming small world networks. In particular, we will be interested in networks that grow in time from small size to large size by the successive addition of new nodes (see [5,11–14] for other models of growing networks).

Many networks have their topology influenced by geographical constraints. The nodes are separated by some physical distance and thus their ability to know the complete state of all the network nodes at a given time is restricted. Consequently, in our model we restrict the formation of links between nodes to result from geographically local processes. That is, when a new node appears, it forms links only to those preexisting nodes that are geographically close to it. In spite of the link formation being exclusively local, long-range links will be shown to arise as a result of network growth. This in addition to the clustering induced by local connections yields the small world property.

We say that a growing network model has the small world property if it satisfies the following three criteria as the number of nodes $N\to\infty$: (a) small average node degree, $\langle k\rangle = O(1)$; (b) small characteristic path length, $L\sim\ln N$; and (c) high average clustering, $\langle C\rangle = O(1)$ (i.e., $\langle C\rangle$ does not decay with N).

*Electronic address: jozik@umd.edu

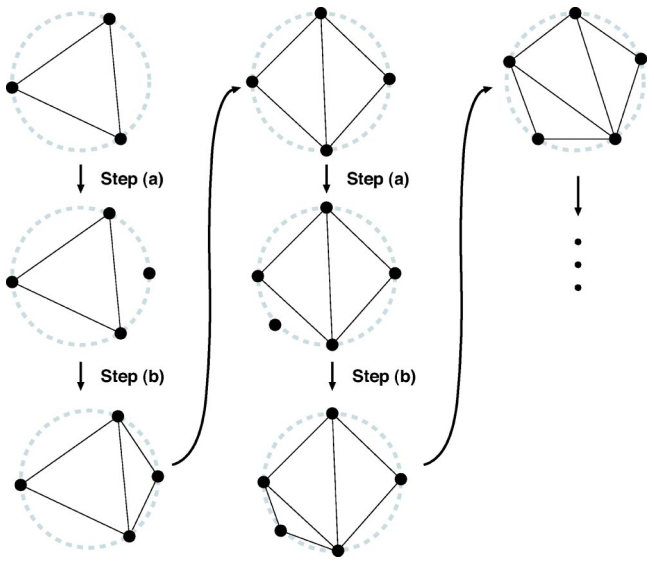

FIG. 1. Our growing network model, illustrated for $m=2$. We begin with $m+1$ completely connected nodes on the circumference of a circle (top left). At each subsequent time step we (a) add a new node in a randomly chosen internode interval along the circle circumference, with every interval having equal probability of being chosen, and (b) connect the new node to its m nearest neighbors, with nearest here referring to distance along the circle circumference. Steps (a) and (b) are repeated until the desired system size is reached.

Part (a) of the definition is included to ensure that highly connected networks that trivially satisfy criteria (b) and (c) are not considered to be small world networks (e.g., if every node is connected to every other node then $L=\langle C\rangle=1$ but $\langle k\rangle=N-1$).

II. GROWING NETWORK MODEL

Our model, as mentioned above, is a growing network. We begin with an initial state of $m+1$ all-to-all connected nodes on the circumference of a circle (Fig. 1). (We take m to be even.) We note that this initial state is chosen solely for convenience and it has no effect on the long-time network properties. At each subsequent discrete time step we grow the network according to the following prescription: (a) a new node is placed in a randomly chosen internode interval along the circle circumference, where all intervals have the same probability of being chosen; (b) the new node makes m links to its m (previously existing) nearest neighbors. Nearest here refers to the distance measured in number of intervals along the circumference of the circle.

These steps are repeated sequentially, creating a network with a temporally growing number of nodes N. We note that, since the network size N is incremented by one with each discrete time step, N can be used interchangeably as a system size or a time variable.

III. DEGREE DISTRIBUTION

We now calculate the degree distribution for our network when N is large. We define $\hat{G}(k,N)$ as the number of nodes with degree k when the system size (or time) is N. Since all new nodes are initially created with $k=m$, and links can only be added to nodes, $\hat{G}(k,N)=0$ for $k<m$. At time N, a node with degree $k=m$ is added to the network, and if it links to a previously existing node i, then $k_i\rightarrow k_i+1$. Each preexisting node is equally likely to be connected to the new node, and therefore the probability that a given preexisting node has its degree increased by 1 is m/N.

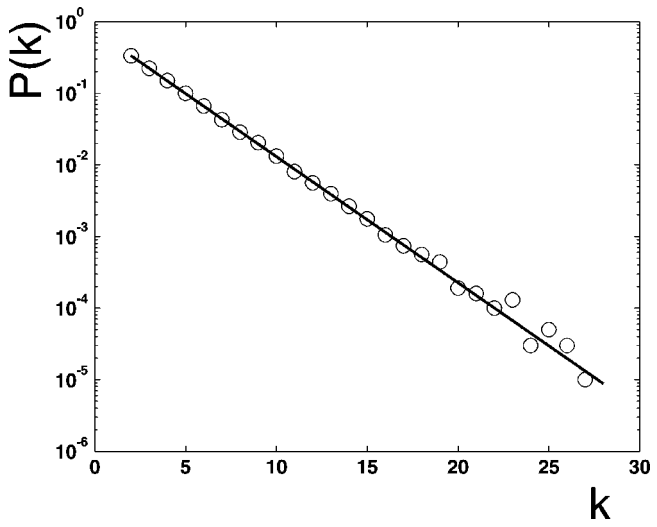

FIG. 2. The open circles represent the degree distribution $P(k)$ for a network grown according to our model with $N=1\times10^5$ and the solid line is the analytically calculated ensemble averaged degree distribution [Eq. (2)], both with $m=2$.

We now take the average over all realizations of the possible random placements of the new node. This yields the following evolution equation for the average of $\hat{G}$, which we denote G,

$$G(k,\ N+1)=\left(1-\frac{m}{N}\right)G(k,N)+\frac{m}{N}G(k-1,\ N)+\delta_{km}, \tag{1}$$

where δ_{km} is the Kronecker delta function. The first term on the right-hand side is the expected number of nodes with degree k at time N whose degree remain the same at time $N+1$. The second term is the expected number of nodes with degree $k-1$ at time N whose degree increase to k at time $N+1$. The third term represents the new node with degree m.

We let $H(k,N)=G(k,N)/N$ be the fraction of nodes with degree k at time N, i.e., the degree distribution. In the appendix, we show that for large N, $H(k,N)$ approaches an asymptotically N invariant form $\overline{H}(k)$, given by

$$\overline{H}(k)=\frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m} \tag{2}$$

for $k\geqslant m$ and $\overline{H}(k)=0$ for $k<m$.

In Fig. 2, the data points represent the degree distribution $P(k)$ for a single network realization randomly grown by our algorithm (illustrated in Fig. 1) for $m=2$ at $N=10^5$. The

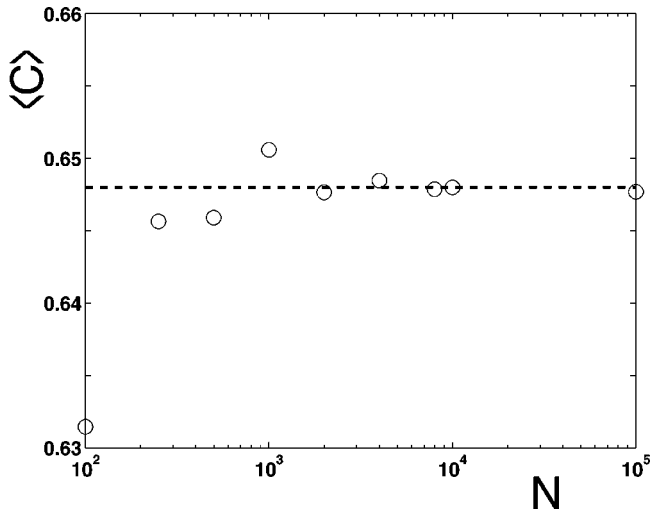

FIG. 3. Average clustering $\langle C\rangle$ vs system size N for simulated networks with $m=2$. As N grows, the average clustering approaches the value (dashed line) predicted in Eq. (4).

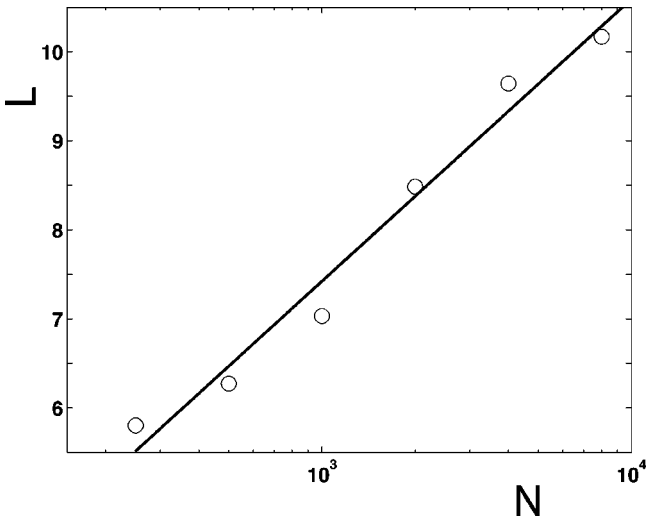

FIG. 4. Semilogarithmic graph of the characteristic path length L vs the system size N. The data shows the small world slow path length growth characteristic, $L\sim\ln N$. The straight line is a fit to the data.

solid line is $\overline{H}(k)$ from Eq. (2), also with $m=2$. We observe good agreement between the analytical calculation for the ensemble average over realizations and the simulation of a single realization, with both showing an exponentially decaying degree distribution. This agreement illustrates that ''self-averaging'' applies for large N.

In addition, we can calculate the average node degree at time N, $\langle k\rangle$, as $N\to\infty$:

$$\langle k\rangle_{\lim N\to\infty}=\sum_{k=m}^{\infty} k\overline{H}(k)=2m. \tag{3}$$

This can be seen also by observing that each time N increases by 1, m new links are formed, and since each link has two ends, the sum of the degrees of all nodes increases by $2m$ at each time step. Thus, our first criterion for a small world network (that $\langle k\rangle$ remains bounded as $N\to\infty$) is met.

IV. CLUSTERING

For the particular case of $m=2$ we can calculate the average clustering of the network exactly. For this value of m, a new node joins the network with $k=2$ and $q=1$. Each subsequent addition of a link to that node increments both k and q by one. Thus, $q=k-1$ for all nodes. Since, by definition, $C_i=2q_i/k_i(k_i-1)$, the average clustering over all nodes in the $m=2$ case is given by

$$\langle C\rangle=2\left\langle\frac{1}{k}\right\rangle=2\sum_{k=2}^{\infty}\frac{1}{k}\,\overline{H}(k)=\frac{3}{2}\ln 3-1\approx 0.648. \tag{4}$$

The open circles in Fig. 3 are the node averaged clustering for single network realizations randomly grown by our algorithm (illustrated in Fig. 1) versus the network size N for $m=2$. As N grows, these data are observed to approach the ensemble averaged large N result given by Eq. (4) (dashed line). In networks with larger values of m we also observe approach of $\langle C\rangle$ to a constant asymptotic value as N increases (the asymptotic $\langle C\rangle$ grows with m; e.g., for $m=4$, $\langle C\rangle\simeq 0.653$).

The network maintains a high average clustering as $N\to\infty$ and, therefore, the second criterion for a small world network is met. This high clustering is expected due to the local nature of the links made. A new link is inserted in a region that already has high interconnectivity, assuring that the nodes with which the new connections are made have a high probability of having connecting links to each other.

V. CHARACTERISTIC PATH LENGTH

The open circles in Fig. 4 show L, the shortest path length between pairs of nodes averaged over all node pairs of single growing network realizations, on a linear scale versus N on a logarithmic scale. The data shows a linear trend, demonstrating the desired slow growth of geodesic path lengths with system size; i.e., $L\sim\ln N$. Thus, the third, and final, small world network criterion is also satisfied.

To see why L grows more slowly than N, consider the fact that, although the links made by incoming nodes are always local, the network itself is growing. The older nodes that had once been nearest neighbors along the circle (and therefore linked) are pushed apart as newer nodes are inserted into the interval between them. Figure 5 illustrates this for the case of $m=2$. The network begins as three nodes linked to each other. By the time the network reaches $N=100$, we see that the original nodes are not adjacent but, rather, have a large number of newer nodes between them. Thus, growth leads to long links between old nodes, and these long links are the shortcuts responsible for a short characteristic path length.

To see why $L\sim\ln N$, imagine a network of size $N\gg 1$ and characteristic path length L. Now if we grow the network by adding N new nodes, these nodes will be roughly uniformly distributed along the circle circumference. This means that,

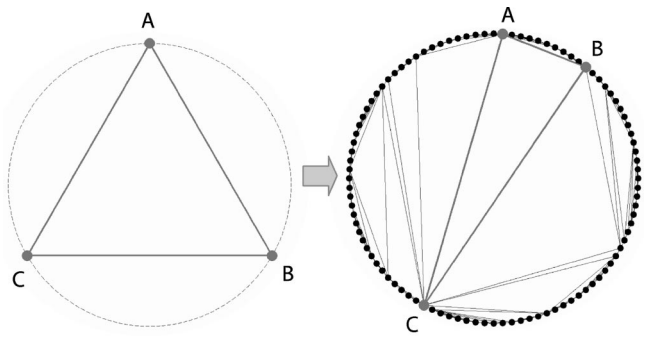

FIG. 5. An illustration of network growth in our model for $m=2$. The network starts off (left) with three adjacent nodes, labeled by *A*, *B*, and *C*, connected to each other via links. When the network reaches a network size of 100 nodes (right), the original three labeled nodes are no longer adjacent, but have been "pushed apart" by the new nodes that were inserted between them. The links connecting *A*, *B*, and *C* serve as shortcuts (similar to the shortcuts in the Watts-Strogatz model [1]), resulting in a small characteristic path length for the network.

on average, a new node would be a distance of $O(1)$ from one of the first N nodes. Thus, as $N\to 2N$ [i.e., $\ln N\to\ln N+O(1)$], we expect L to increase to $L+O(1)$, resulting in $L\sim\ln N$.

VI. CONCLUSION

We presented a small world network model that has only geographically local interactions. This model provides a physically realistic mechanism by which growing physical systems that have geographical constraints, and therefore limited global information available to each individual node, can form networks with small world characteristics. Additionally, our results suggest that small world networks in geographically constrained physical systems may be a natural consequence of system growth and local interactions.

APPENDIX: CALCULATING $\bar{H}(K)$

We substitute $H(k,N)=G(k,N)/N$, the fraction of nodes with degree k at time N, into Eq. (1), obtaining,

$$(N+1)H(k,N+1)=(N-m)H(k,N)+mH(k-1,N)+\delta_{km}. \quad \text{(A1)}$$

We define $\bar{H}(k)$ to be the N independent solution to Eq. (A1). Substituting into Eq. (A1) and rearranging terms we get

$$\bar{H}(k)=\frac{m}{m+1}\bar{H}(k-1)+\frac{\delta_{km}}{m+1}. \quad \text{(A2)}$$

This recursion relation is solved to yield

$$\bar{H}(k)=\frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m} \quad \text{(A3)}$$

for $k\geqslant m$ and $\bar{H}(k)=0$ for $k<m$.

Now we show that as $N\to\infty$, $H(k,N)$ approaches $\bar{H}(k)$. Dividing (A1) by $N+1$ and subtracting from this the same equation but with $\bar{H}(k)$ inserted, we obtain after some algebra

$$H(k,N+1)-\bar{H}(k)=\frac{N-m}{N+1}[H(k,N)-\bar{H}(k)]+\frac{m}{N+1}[H(k-1,N)-\bar{H}(k-1)]. \quad \text{(A4)}$$

Letting $F(k,N)=\binom{N}{m+1}[H(k,N)-\bar{H}(k)]$, we find that (A4) is equivalent to

$$F(k,N+1)=F(k,N)+\frac{m}{N-m}F(k-1,N). \quad \text{(A5)}$$

Our goal is to show that $F(k,N)/\binom{N}{m+1}\to 0$ as $N\to\infty$.

Consider first the case $k=m$. Since $F(m-1,N)=0$, (A5) implies that $F(m,N)=C$ for some constant C independent of N. Thus, in particular,

$$H(m,N)-\bar{H}(m)=\frac{C}{\binom{N}{m+1}}, \quad \text{(A6)}$$

and $H(m,N)\to\bar{H}(m)$ as $N\to\infty$. Then when $k=m+1$, (A5) becomes

$$F(m+1,N+1)=F(m+1,N)+\frac{m}{N-m}C \quad \text{(A7)}$$

and hence

$$F(m+1,N)=F(m+1,m+1)+C\sum_{N'=m+1}^{N-1}\frac{m}{N'-m}\sim\ln N \quad \text{(A8)}$$

for large N.

Furthermore, it is possible to show by induction on k that

$$|F(k,N)|\leqslant C(k)(m\ln N)^{k-m} \quad \text{(A9)}$$

for some $C(k)$ independent of N. In other words,

$$|H(k,N)-\bar{H}(k)|\leqslant C(k)\frac{(m\ln N)^{k-m}}{\binom{N}{m+1}} \quad \text{(A10)}$$

which approaches 0 as $N\to\infty$.

[1] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[2] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[3] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[4] M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[5] R. Albert and A.-L. Barabási, Science **286**, 509 (1999).
[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, ACM SIGCOMM **29**, 251 (1999).
[7] R.F. i Cancho, C. Janssen, and R.V. Solé, Phys. Rev. E **64**, 046119 (2001).
[8] J. Camacho, R. Guimerà, and L.A.N. Amaral, Phys. Rev. Lett. **88**, 228102 (2002).
[9] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[10] H. Jeong, S.P. Mason, A.-L. Barabasi, and Z.N. Oltvai, Nature (London) **411**, 41 (2001).
[11] S.H. Strogatz, Nature (London) **410**, 268 (2001).
[12] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).
[13] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[14] M.E.J. Newman, SIAM Rev. **45**, 167 (2003).