Exact results and scaling properties of small-world networks

R. V. Kulkarni,* E. Almaas, and D. Stroud
Department of Physics, The Ohio State University, Columbus, Ohio 43210
(Received 19 August 1999)

We study the distribution function for minimal paths in small-world networks. We derive its general form, which we verify numerically, and also discuss its scaling properties. Using the general form of this distribution function, we derive exact expressions for the average minimal distance $\overline{\ell}$ and its variance σ^2. Finally, we study the limit of large system sizes and obtain some approximate results.

PACS number(s): 84.35.+i, 05.50.+q, 64.60.Fr, 87.18.Sn

Recently Watts and Strogatz [1] have studied a class of networks that can be ''tuned'' from an ordered to a random state by varying a single parameter. For a range of parameter values, they found that the networks resembled ordered networks locally but random networks globally. They named this class of networks ''small-world'' networks because of their relevance to a well-known problem in sociology [2]. Since their introduction, small-world networks and their properties have received considerable attention [4–18], in part because of their possible application to a broad range of systems, ranging from social networks [2] to coupled oscillators [3].

Much of the work characterizing the properties of small-world networks has focused on the average minimal distance $\overline{\ell}$ separating two randomly chosen points in the network. Previous work has shown that $\overline{\ell}$ has two scaling regimes: for small system sizes L, it is found that $\overline{\ell}\sim L$, whereas for large L, $\overline{\ell}\sim\log(L)$ [19]. A scaling form for $\overline{\ell}$ has been proposed and numerically confirmed; however, the nature of the crossover between the two scaling regimes has been the subject of debate [5,8,10,12,14,15]. In this work, we will focus on the basic probability distributions for small-world networks, and as a consequence obtain exact results for $\overline{\ell}$ and its variance σ^2.

We generate the networks following the prescription of Newman and Watts [8]. We start with a one-dimensional periodic lattice with $L=2N$ sites and nearest neighbor connections. We then add shortcuts uniformly with probability p such that the average number of shortcuts added is $x=pL$. We denote the distance between two sites, counted along the lattice using only nearest neighbor links, as the Euclidean distance. By contrast, the shortest distance between two sites, counted along any bond including shortcut bonds, is called the minimal distance.

Using these definitions, we now introduce the following probability functions: (i) $\overline{P}(n|m)$, the probability that two sites are separated by Euclidean distance n given that their minimal distance is m; (ii) $P(m|n)$, the probability that two sites have minimal separation m given that their Euclidean distance is n; and (iii) $Q(m)$, the probability that two randomly chosen sites have a minimal separation m. Recently, Dorogovtsev and Mendes [18] have introduced two exactly solvable models similar to small-world networks. For these models they drive the explicit form of $P(m|n)$, from which they obtain $\overline{\ell}$ and other properties of their networks. In this paper, we derive the general form of $P(m|n)$ for small-world networks, and we confirm it numerically. Using this form, we derive an exact expression for $\overline{\ell}$ and for the variance of $\overline{\ell}$, $\sigma^2\equiv\overline{\ell^2}-\overline{\ell}^2$. We also study the scaling properties of $P(m|n)$ and obtain some approximate results for it in the limit of large L. Note that in describing the networks, we have considered the case of coordination number of $2k=2$ for each site. However, our arguments for the general form of $P(m|n)$ are valid for arbitrary k. For simplicity we will consider the case $k=1$ in the rest of this paper, and generalizations to arbitrary k will be indicated as appropriate.

We begin by deriving the general form of $\overline{P}(n|m)$. First, since the minimal distance cannot exceed the Euclidean distance, $\overline{P}(n|m)=0$ for $n<m$. For $n>m$, the minimal path must use at least one shortcut. But taking a shortcut is equivalent to randomizing the position along the network, since the shortcuts are uniformly distributed. Hence, $\overline{P}(n|m)$ must be independent of n for all $n>m$. Finally, for $n=m$, it is not necessary to use any shortcuts in the minimal path; so the arguments invoked for $n>m$ do not apply. Instead, $\overline{P}(n|n)$ is determined by the constraint that the probability distribution is normalized.

We now derive the general form of $P(m|n)$. From elementary probability theory, we have

$$\overline{P}(n|m)Q(m)=\begin{cases}\dfrac{2}{L-1}P(m|n), & n<N\\[2ex] \dfrac{1}{L-1}P(m|n), & n=N.\end{cases}\tag{1}$$

From Eq. (1) and the properties discussed in the previous paragraph, $P(m<n|n)\equiv f(m)$ is independent of n, and $P(m>n|n)=0$. Thus the general form of $P(m|n)$ is

$$P(m|n)=\Theta(n-m)f(m)+\left(1-\sum_{m'=1}^{n-1}f(m')\right)\delta_{m,n},\tag{2}$$

where $\Theta(x)$ is defined by $\Theta(x)=0$ for $x\leqslant 0$ and $\Theta(x)=1$ for $x>0$. We have numerically confirmed the validity of this form, as shown in Fig. 1.

*Electronic address: rahul@mps.ohio-state.edu

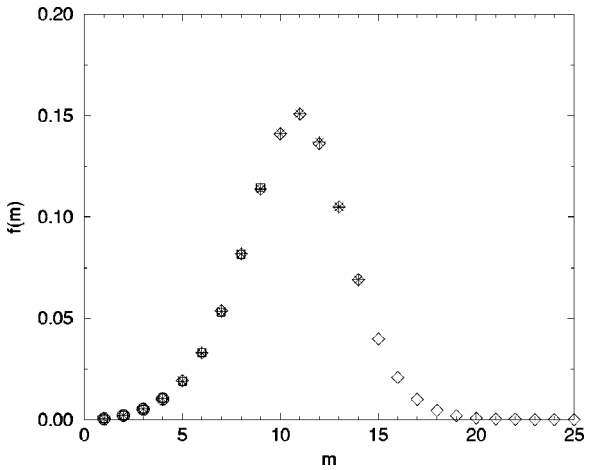

FIG. 1. The values of $f(m)$ as obtained from $P(m|n)$, which is calculated from simulations for the n values $n=5$ (○), $n=10$ (□), $n=15$ (∗), and $n=500$ (◇). This figure confirms the statement that $P(m<n|n)$ is independent of n. The parameters for the figure are $L=1000$, $p=0.25$. Similar confirmation has been obtained from simulations for a wide range of parameter values.

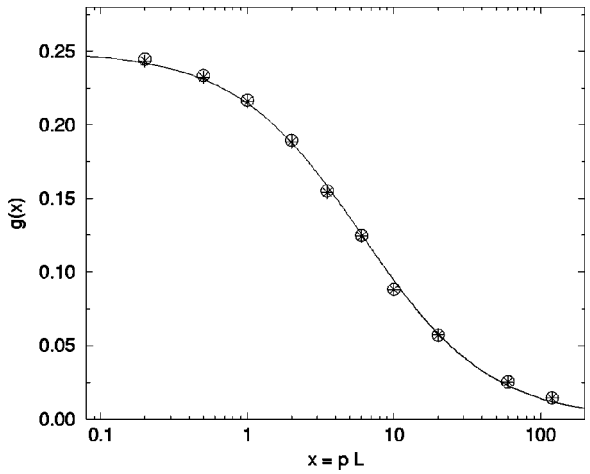

FIG. 2. The average minimal separation $\overline{\ell}=Lg(x)$ vs the average number of shortcuts, $x=pL$, as obtained from numerical simulation by averaging over all pairs (∗), numerical simulation using Eq. (6) (○), and Padé fit as given by Ref. [12] (solid line). This confirms the exact expression Eq. (6).

The fact that $P(m|n)$ is completely determined by $f(m)$ implies that we need to consider only ''diametrically opposite sites'' (d.o.s) in its computation [since $f(m)=P(m|N)$]. It also has some surprising consequences, regardless of the exact form of $f(m)$. To explore these, we examine some other properties of small-world networks. For example, besides $\overline{\ell}$, the quantity $\langle \ell(n)\rangle$, which is the average minimal distance separating two sites with Euclidean distance n, has been discussed in the literature [14,15]. We can express both these quantities in terms of $P(m|n)$ as follows:

$$\langle \ell(n)\rangle = \sum_{m=1}^{n} mP(m|n), \tag{3}$$

$$\overline{\ell} = \frac{1}{L-1}\left(2\sum_{n=1}^{N-1}\sum_{m-1}^{n} mP(m|n) + \sum_{m=1}^{N} mP(m|n)\right). \tag{4}$$

Similar expressions hold for $\langle \ell^2(n)\rangle$, $\langle \ell^3(n)\rangle$, and $\overline{\ell}^2$. Substituting the form of $P(m|n)$ [Eq. (2)] into the expression for $\overline{\ell}$ [Eq. (4), we obtain

$$\overline{\ell} = \frac{1}{L-1}\left[2\sum_{n=1}^{N-1}\sum_{m=1}^{n-1} mf(m) + 2\sum_{n=1}^{N} n\left(1-\sum_{m=1}^{n-1} f(m)\right)\right.$$
$$\left. + \sum_{m=1}^{N-1} mf(m) + N\left(1-\sum_{m=1}^{N-1} f(m)\right)\right], \tag{5}$$

which can be simplified to give the following exact expression:

$$\overline{\ell} = \langle \ell(N)\rangle\left(1+\frac{1}{L-1}\right) - \frac{\langle \ell^2(N)\rangle}{L-1}. \tag{6}$$

Similarly, we obtain an expression for the variance of the distribution of minimal distances, $\sigma^2=\overline{\ell^2}-\overline{\ell}^2$:

$$\sigma^2 = \langle \ell^2(N)\rangle\left(1+\frac{1}{L-1}\right) + \frac{1}{3(L-1)}[\langle \ell(N)\rangle - 4\langle \ell^3(N)\rangle]$$
$$-\left[\langle \ell(N)\rangle\left(1+\frac{1}{L-1}\right) - \frac{\langle \ell^2(N)\rangle}{L-1}\right]^2. \tag{7}$$

The surprising aspect of the above equations is that $\overline{\ell}$ and σ^2, which are average properties of the entire network, are completely determined by the mean separation of d.o.s. $\langle \ell(N)\rangle$ and its higher moments $\langle \ell^2(N)\rangle$ and $\langle \ell^3(N)\rangle$. Note that Eqs. (6) and (7) can readily be generalized to any k by performing the substitution $L\mapsto[L/k]$.

When the network has exactly one shortcut, we can calculate $\overline{\ell}$ analytically using Eq. (6). In this case, in the limit of large N, we get $\langle \ell(N)\rangle=\frac{2}{3}N$ and $\langle \ell^2(N)\rangle=\frac{1}{2}N^2$ which gives $\overline{\ell}=\frac{5}{12}N$. As expected, this is in perfect agreement with the results obtained by Strang and Eriksson [20]. We have further confirmed Eq. (6) by numerically computing $\overline{\ell}$ using the following two procedures: (i) averaging the minimal distance over all pairs of sites, and (ii) considering only pairs of d.o.s. and using Eq. (6). The results, which are presented in Fig. 2, indicate that the two procedures are equivalent.

The results obtained so far have been independent of the functional form of $f(m)$. To gain further insight, we consider the scaling properties of $f(m)$, following the real-space renormalization group (RG) analysis of Newman and Watts [8]. This procedure consists of blocking pairs of adjacent sites while preserving the total number of shortcuts in the network. This gives for the transformed lattice $N'=N/2$ and $p'=2p$. We note the following features of this transformation [8]: (i) the geometry of the minimal paths is unchanged in almost all cases, and the number of site pairs for which the geometry does change is negligible for large L and small p, and (ii) the distance along the minimal path is halved, i.e., $m'=m/2$ for large L and small p. Furthermore, we note that the RG transformation maps two pairs of d.o.s. into a single

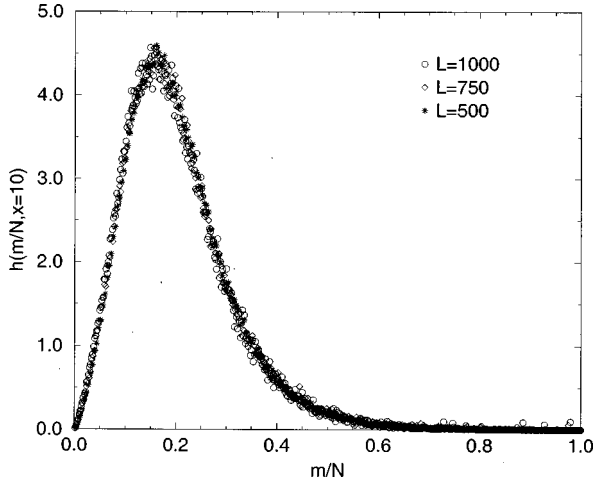

FIG. 3. This figure confirms the proposed scaling form of $f(m,N,p)$ [Eq. (10)] for $x=pL=10$ and system sizes $L=500$ (*), 750 ($\diamond$), and 1000 ($\bigcirc$). We have confirmed this scaling collapse for a wide range of x values.

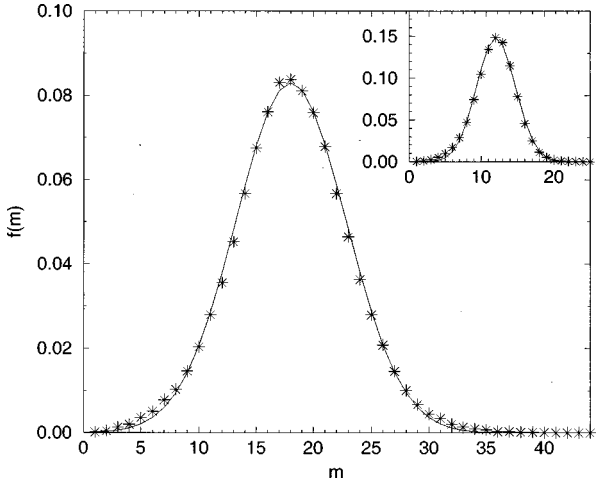

FIG. 4. $f(m)$ vs m for $x=pL=250$ and $L=2000$. The solid line is the Gaussian fit to the calculated data. The inset shows the Gaussian fit for $x=500$ and $L=2000$. Note that with increasing x, the Gaussian becomes more sharply peaked.

pair. This fact, in conjunction with points (i) and (ii) above, give us

$$f'\left(\frac{m}{2},\frac{N}{2},2p\right)=2f(m,N,p), \tag{8}$$

where we have made the dependence on N and p of $f(m)$ explicit. For large N, taking the continuum limit, we can generalize the above expression to

$$f'\left(\frac{m}{\lambda},\frac{N}{\lambda},\lambda p\right)=\lambda f(m,N,p). \tag{9}$$

These observations can now be summarized in the following scaling form:

$$f(m,N,p)=\frac{1}{N}h(y,x) \tag{10}$$

where

$$y=\frac{m}{N},\quad x=2pN.$$

By fixing x, we have observed the scaling collapse of $f(m)$ for different values of N and p. This is demonstrated numerically for $x=10$ in Fig. 3. Our simulations indicate that for any given x, this scaling collapse holds for small p and large enough N.

It is interesting to note that the scaling properties of $\overline{\ell}$ can be derived from the scaling form of $f(m)$. Using the definition of $\overline{\ell}$ [Eq. (4)] and the scaling form for $f(m)$, we get

$$\overline{\ell}=\frac{L}{4}\left(1-\int_0^1 dy(1-y)^2h(y,x)\right) \tag{11}$$

$$=Lg(x), \tag{12}$$

which is consistent with the scaling form proposed in previous works. Similar scaling forms hold for $\overline{\ell^2}$, $\langle\ell(N)\rangle$, and $\langle\ell^2(N)\rangle$.

We now consider the limit of large system sizes such that $x\gg 1$. In this limit, we have observed numerically that we can approximate $f(m)$ by a Gaussian distribution function:

$$f(m)=\frac{1}{\sqrt{2\pi\sigma_g^2}}e^{-(m-\mu_g)^2/2\sigma_g^2}, \tag{13}$$

where μ_g and σ_g^2 are, respectively, the mean and variance of the distribution. The corresponding fit for $x=250$ and $x=500$ is shown in Fig. 4. Our simulations indicate that as x increases, (μ_g/σ_g) also increases, as can be seen from the figure.

Using the Gaussian approximation for $f(m)$, we are now able to calculate the function $\langle\ell(n)\rangle$, which has been discussed elsewhere [14,15]. From Eqs. (3) and (13) in the limit $\mu_g\gg\sigma_g$, we obtain

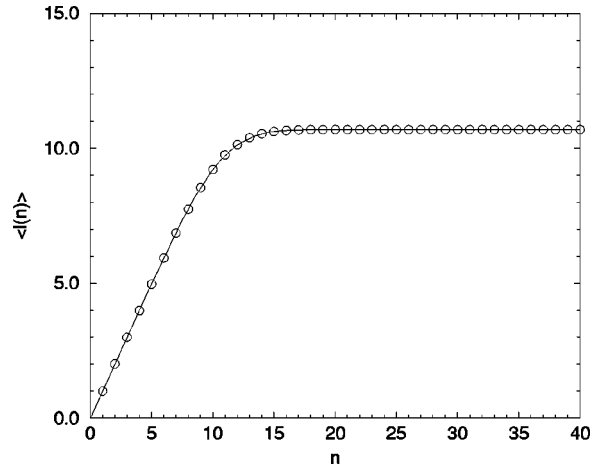

FIG. 5. The mean distance $\langle\ell(n)\rangle$ between two sites having Euclidean separation n for $x=pL=250$. Results are shown for numerical simulation ($\bigcirc$) and analytic expression [Eq. (14)] (solid line). The analytic expression is an excellent fit for $x\gg 1$.

$$\langle \ell(n) \rangle = n - \tfrac{1}{2}(n-\mu_g)\left[\Phi\left(\frac{n-\mu_g}{\sqrt{2}\sigma_g}\right)+1\right] + \frac{\sigma_g}{\sqrt{2\pi}} e^{-(n-\mu_g)^2/2\sigma_g^2}, \tag{14}$$

where we have assumed $\Phi(\mu_g/\sqrt{2}\sigma_g)=1$. In this limit, substituting the above form of $f(m)$ into the definitions of $\langle \ell(N) \rangle$ and $\langle \ell^2(N) \rangle$, we get

$$\mu_g = \langle \ell(N) \rangle, \tag{15}$$

$$\sigma_g^2 = \langle \ell^2(N) \rangle - \langle \ell(N) \rangle^2. \tag{16}$$

In particular, these equations imply that μ_g and σ_g have the following scaling forms: $\mu_g \sim L g_1(x)$ and $\sigma_g \sim L g_2(x)$. Using these relations, we see that the Gaussian ansatz for $f(m)$ [Eq. (13)] is consistent with the scaling form proposed in Eq. (10). In Fig. 5, we compare Eq. (14) to results from our simulations for $x=250$. In the limit $L \to \infty$, we have $(\sigma_g/\mu_g) \to 0$, which upon substitution into Eq. (14) gives us

$$\langle \ell(n) \rangle = \begin{cases} n, & n < \langle \ell(N) \rangle \\ \langle \ell(N) \rangle, & n \geq \langle \ell(N) \rangle. \end{cases} \tag{17}$$

This expression for $\langle \ell(n) \rangle$ is consistent with that previously derived exactly in Refs. [14,15] in the same limit.

In conclusion, we have studied the probability distribution for minimal path lengths in small-world networks. We have presented arguments for the general analytical form this distribution must take, and have verified this numerically. Using this form, we have also derived some exact relations. We have obtained an approximate scaling form for this probability distribution in the limit of large system sizes. It is our hope that further efforts along these lines will provide a better understanding of the structure of small-world networks.

This work has been supported by NASA through Grant No. NCC8-152 and NSF through Grant No. DMR97-31511. Computational support was provided by the Ohio Supercomputer Center, the San Diego Supercomputer Center, and the Norwegian University of Science and Technology (NTNU).

[1] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[2] S. Milgram, Psychol. Today **2**, 60 (1967).
[3] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer, Berlin, 1984).
[4] S.A. Pandit and R.E. Amritkar, Phys. Rev. E **60**, R1119 (1999).
[5] M. Barthelemy and L.A.N. Amaral, Phys. Rev. Lett. **82**, 3180 (1999); **82**, 5180(E) (1999).
[6] A. Barrat, Phys. Rev. E **57**, 3629 (1998).
[7] R. Monasson, e-print cond-mat/9903347.
[8] M.E.J. Newman and D.J. Watts, Phys. Lett. A (to be published).
[9] A. Barrat and M. Weigt, Eur. Phys. J. B (to be published).
[10] M.A. de Menezes, C. Moukarzel, and T.J.P. Penna, e-print cond-mat/9903426.
[11] R. Kasturirangan, e-print cond-mat/9904055.
[12] M.E.J. Newman and D.J. Watts, Phys. Rev. E **60**, 7332 (1999).
[13] R.V. Kulkarni, E. Almaas, and D. Stroud, e-print cond-mat/9905066.
[14] C.F. Moukarzel and M.A. de Menezes, e-print cond-mat/9005131.
[15] C.F. Moukarzel, Phys. Rev. E **60**, 6263 (1999).
[16] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **401**, 130 (1999).
[17] A.-L. Barabasi, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[18] S.N. Dorogovtsev and J.F.F. Mendes, e-print cond-mat/9907445.
[19] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).
[20] G. Strang and H. Eriksson (unpublished). See [12].