Characterizing the Structure of Small-World Networks

E. Almaas,[1]* R. V. Kulkarni,[2] and D. Stroud[1]

[1]*Department of Physics, The Ohio State University, Columbus, Ohio 43210*

[2]*Department of Physics, University of California, Davis, California 95616*

(Received 12 September 2001; published 14 February 2002)

We give exact relations for small-world networks (SWN's) which are *independent* of the "degree distribution," i.e., the distribution of nearest-neighbor connections. For the original SWN model, we illustrate how these exact relations can be used to obtain approximations for the corresponding basic probability distribution. In the limit of large system sizes and small disorder, we use numerical studies to obtain a functional fit for this distribution. Finally, we obtain the scaling properties for the mean-square displacement of a random walker, which are determined by the scaling behavior of the underlying SWN.

DOI: 10.1103/PhysRevLett.88.098101 PACS numbers: 87.18.Sn, 05.10.–a, 05.40.–a, 05.50.+q

Networks occur in many contexts in the sciences and humanities. Among these are neural networks [1], social networks [1–3], food webs [4,5], and the Internet and the World Wide Web [6,7]. Many properties of such complex systems can be understood by considering the network of interactions which connects their components. For this reason, much recent research has focused on the structure of these networks [8–13].

The study of network models has recently been advanced by the introduction of small-world networks (SWN's) by Watts and Strogatz [1]. These networks have an ordered structure locally, but are random on a global scale. This combination of features suggests that SWN's can be used to describe the behavior of many real-world interacting networks [14]. However, as originally defined, SWN's do not exhibit a property found in some large-scale networks such as the World Wide Web, in which the probability distribution for the number of nearest-neighbor connections to a given site (the "degree distribution") is a power law, rather than Poisson-like, as in the SWN's [9]. Alternative models which do exhibit this power-law distribution have been proposed [15–18].

In this Letter, we study SWN's where shortcuts according to a given degree distribution are superimposed on a regular network. For this class of networks, we show exact results which are *independent* of the specific degree distribution employed. This independence is noteworthy, since the major point of difference between the various network models currently studied is the degree distribution. Our results indicate that there exist "universal" properties of random networks which are independent of the degree distribution. Furthermore, they can also be used to find approximations for the basic probability distribution (as defined below) of the network. On the basis of numerical simulations of the original SWN model (with Poisson-like degree distribution), we propose that the basic probability distribution is characterized by a function of a single variable in the limit of small disorder (as defined below) and large system size: the so-called "small-world" regime. Using this functional form, various structural properties of networks in the small-world regime can be accurately determined without the need for further, expensive computer simulations. We also study how the structural properties of this particular SWN model affect an important example of dynamics on SWN's: the time-dependent mean-square displacement of a random walker.

The SWN's are defined by starting from a one-dimensional regular network with periodic boundary conditions and $L = 2N$ nodes, each node being connected to its $2k$ nearest neighbors. We add shortcut ends to each node according to a given degree distribution by following the prescription of Refs. [17,19]. For the original SWN model (with $k = 1$), the degree distribution is $\mathcal{D}_q = (1 - p)\mathcal{P}_{(q)}(p) + p\mathcal{P}_{(q-1)}(p)$, where $\mathcal{P}_q(\lambda) = \exp(-\lambda)\lambda^q/q!$ is the Poisson distribution. We then select pairs of shortcut ends at random and connect them to each other, thus creating a shortcut. This procedure (with the above $\mathcal{D}_q$) of generating the networks is equivalent to the procedure outlined by Newman and Watts [8], with kp as the probability of having a shortcut at a given site. Thus, on average, there will be $x = kpL$ shortcuts in the network. The present numerical results were calculated for $k = 1$; however, generalization to other k is straightforward and indicated in our equations (which we have verified numerically). The "small disorder" regime corresponds to $p \ll 1$, and we will work in the limit of large system sizes such that terms of $\mathcal{O}(1/L)$ can be omitted.

The SWN is characterized by two types of distances: the "Euclidean distance," defined as the shortest distance between two sites before any shortcuts are introduced into the network; and the "minimal distance," which denotes the shortest distance after shortcuts have been added. We now define $P(m|n)$ (the basic probability distribution) as the probability that two sites have minimal separation m, given that their Euclidean distance is n. For the original SWN model, we have previously proved [13] that $P(m|n)$ can be written in the form

$$P(m \mid n) = \begin{cases} f(m), & m < n\,, \\ 1 - \sum_{i=1}^{n-1} f(i), & m = n\,, \\ 0, & m > n\,. \end{cases} \tag{1}$$

The proof is quite general, however, and the above result is *exact* for SWN's with arbitrary degree distributions. Furthermore, it is also true for two (and higher) dimensional SWNs, which have attracted attention as models for disease spreading among plants [20,21].

From Eq. (1)we see that $P(m \mid n)$ is fully characterized by the function $f(m)$ [22], which can be regarded as the probability that two diametrically opposite sites have minimal separation m. Many structural properties, such as the average minimal separation between two randomly chosen points, $\overline{\ell}$, can be expressed in terms of $f(m)$ [13], and these expressions can be used to derive *exact* relations connecting various quantities of interest. An important relation derived in this manner relates the average number of sites within m hops of a randomly selected origin, $V(m)$, to $F(m) = \sum_{k=1}^{m} f(k)$, the cumulative probability distribution, as

$$V(m) = 2km + [L - 2km]F(m). \qquad (2)$$

Through our simulations, we have confirmed the validity of Eqs. (1) and (2) for several different types of degree distributions for 1D networks, including networks with power-law degree distributions. These exact relations can be used to obtain quantities of interest for the network. In a simple model for disease spreading [11], for example, $V(m)$ corresponds to the number of infected sites after m time steps. Furthermore, Eq. (2) tells us how to calculate $F(m)$ [and hence $f(m)$], once we know $V(m)$. Thus, any approximation for $V(m)$ can be converted into an approximation for $f(m)$. In the remainder of this paper, we will illustrate this procedure by focusing on the original SWN model.

We will now use the mean-field result of Newman *et al.* for $V(m)$ to obtain an approximation for $f(m)$, and hence, to motivate a new scaling form for this function. In the asymptotic limit ($x \gg 1$), using Eq. (2) in conjunction with the mean-field results of Ref. [12], we obtain a relation between $f(m)$ and $F(m)$:

$$f_{\rm mft}(m) = 4k^2 pF_{\rm mft}(m)[1 - F_{\rm mft}(m)], \qquad (3)$$

which is the logistic growth equation, as pointed out by Barbour *et al.* [23]. Equation (3) can easily be solved for $f_{\rm mft}(m)$ [24]. We note that the final expression for $f_{\rm mft}(m)$ is consistent with the scaling form $f(m; N, p) = h(m/N, pN)/N$ derived in Ref. [13].

We now consider the central moments of this distribution, defined as

$$\mu_\ell = \sum_{m=1}^{N} [m - \langle m \rangle]^\ell f_{\rm mft}(m). \qquad (4)$$

The scaling form for $f(m)$ implies that the scaled moments μ_ℓ/N^ℓ are functions of x only; however, $f_{\rm mft}(m)$ has an additional property: its central moments in this case take

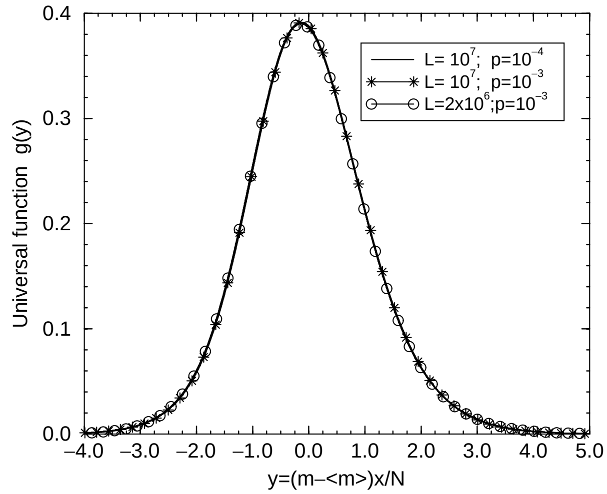

FIG. 1. Plot of the universal function $g(y)$ [cf. Eq. (6)] for three choices of system size L and probability of shortcuts, p (see text): (i) $L = 10^7$, $p = 10^{-4}$ (full line); (ii) $L = 10^7$, $p = 10^{-3}$ (stars and full line); and (iii) $L = 2 \times 10^6$, $p = 10^{-3}$ (circles and full line).

the form

$$\frac{\mu_\ell}{N^\ell} = \frac{c_\ell}{x^\ell}, \qquad (5)$$

where c_ℓ is the ℓth central moment of the logistic distribution. We have confirmed through numerical simulations that Eq. (5) holds for at least the first four central moments of $f(m)$, however, with coefficients c_ℓ which differ significantly from those obtained from the logistic growth equation. For example, from our simulations, we find $c_2 \approx 1.21$, whereas the corresponding mean-field value is $c_2^{\rm mft} \approx 0.82$. Equation (5) also implies that $\mu_\ell \sim p^{-\ell}$, and hence, for fixed p, is independent of N. Consequently, the probability distributions $f(m)$, for fixed p and different values of N, are simply related by a translation of the mean values. We have confirmed this translation property by simulations carried out for a wide range of parameters in the small-world regime.

On the basis of these features, we propose the following scaling form for $f(m)$ in the asymptotic limit ($x \gg 1$):

$$f(m; N, p) = \frac{kx}{N}\, g\left[\frac{kx}{N}(m - \langle m \rangle)\right], \qquad (6)$$

where $g(y)$ is a universal function. It is readily seen that the above scaling form is consistent with Eq. (5). In Fig. 1, we show calculations of $f(m)$ for three different combinations of the parameters (p, L): $(10^{-4}, 10^7)$, $(10^{-3}, 10^7)$, and $(10^{-3}, 10^6)$. For all three choices of parameters, $f(m; N, p)$ can be collapsed onto a single curve. Thus, numerical simulations for just one choice of the parameters (p, L) suffice to determine $g(y)$. To fully determine $f(m)$ we need to find functional forms for both $g(y)$ and $\langle m \rangle$, which we now proceed to do.

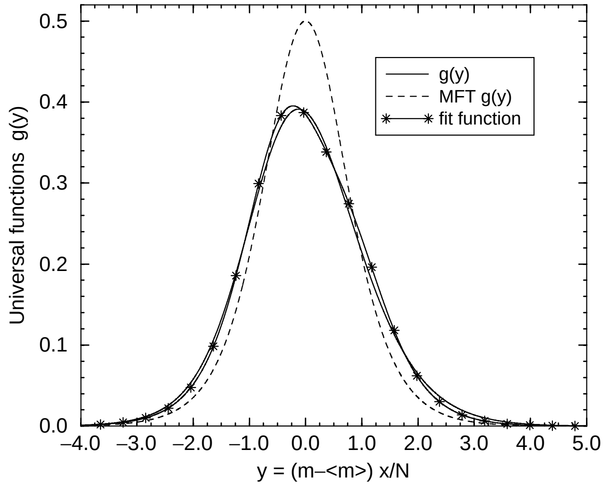

FIG. 2. Comparison of the universal function $g(y)$, as obtained from numerical calculations (full line), mean-field theory (dashed line), and our two-parameter fitting function Eq. (7) (stars and full line).

We can represent $g(y)$ using a two-parameter mixture of logistic distributions:

$$g(y) = \frac{1}{2(1 + \gamma)}[\mathrm{sech}^2(y + \gamma\beta) + \gamma\,\mathrm{sech}^2(y - \beta)]. \tag{7}$$

This form ensures that $g(y)$ is normalized, has zero mean, and has the correct behavior for large $|y|$. In Fig. 2, we show a comparison of the mean-field result, the fitting function, [Eq. (7) with $\gamma = 0.492$ and $\beta = 0.782$], and the numerically obtained $g(y)$. Evidently, this function describes the numerical $g(y)$ very well, while the mean-field $g(y)$ agrees less well.

To determine $\langle m\rangle$, we use the exact relation, which was proved in [13] and is valid for a general degree distribution,

$$\overline{\ell} = \langle m\rangle\left(1 + \frac{k}{L}\right) + \langle m^2\rangle\,\frac{k}{L}, \tag{8}$$

together with the definition $\mu_2 = \langle m^2\rangle - \langle m\rangle^2$. Thus, once we have fitting forms for $\overline{\ell}$ and μ_2, we can determine $\langle m\rangle$. From our simulations, we have already determined μ_2 (see above). As for $\overline{\ell}$, we find numerically that

$$\frac{k\overline{\ell}}{L} = \frac{\log(2x)}{4x} + \frac{\alpha}{x}, \tag{9}$$

with $\alpha = 0.13$. This numerical estimate agrees well with the mathematically rigorous result derived in Ref. [23]. In Fig. 3, we show a comparison between the fitting form (9), the numerically obtained values for $\overline{\ell}$, and the mean-field result $k\overline{\ell}_{\mathrm{mft}}/L \to \log(2x)/(4x)$. As the inset shows, Eq. (9) gives $\overline{\ell}$ with a relative error of less than 1%. Keeping only terms through $\mathcal{O}(1/x)$, and recalling that $\mu_2/L^2 \sim 1/x^2$, we find that $\langle m\rangle \approx \overline{\ell}$. Hence, $\langle m\rangle$ is given by

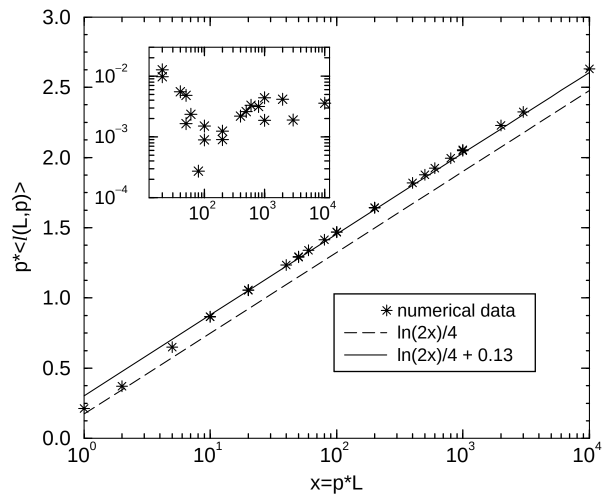

FIG. 3. Numerically computed average minimal distance between two randomly chosen nodes $\overline{\ell}$ (stars) using p values ranging from 10^{-4} to 10^{-3} and system sizes L from 10^4 to 10^7. Dashed line: mean-field result $\overline{\ell}_{\mathrm{mft}}/L = \log(2x)/4x$. Solid line: our fitting form $\overline{\ell}/L = \overline{\ell}_{\mathrm{mft}}/L + 0.13/x$. The inset shows the absolute relative error between the numerical data and our fitting form.

$$\langle m\rangle = \frac{L}{k}\left[\frac{\log(2x)}{4x} + \frac{0.13}{x}\right]. \tag{10}$$

The three equations (6), (7), and (10) provide a full functional form for $f(m)$ in the asymptotic scaling regime for a SWN with degree distribution $\mathcal{D}_q$. Note the striking pattern of successive simplifications during our analysis of the probability functions. Initially, $P(m\,|\,n)$ depended on the *four* variables m, n, p, and N. The exact results obtained in Ref. [13] reduced this dependence to a function of three variables $f(m; N, p)$. Furthermore, the scaling properties discussed in [13] imply that, for $p \ll 1$, $f(m; N, p) = h(m/N, pN)/N$, a function of only two variables. Finally, we have shown in the present work that in the limit $x \gg 1$, the small-world network is characterized by a function of a *single* variable, $g(y)$. This functional form can be used to obtain detailed, accurate, structural information about networks in the small-world regime *without* having to perform computationally expensive simulations.

The scaling forms for $f(m; N, p)$ can also be used to gain insight into *dynamics* on SWN's. As an example, we consider the dynamics of a random walker on SWN's, a problem which has been studied by several authors [25–27]. Here, we focus on the mean-square displacement, $\langle r^2\rangle$ of a random walker on a SWN, as a function of time [28]. At long times, $\langle r^2\rangle$ must saturate to the finite value $\overline{\ell^2}(L, p)$, since, in the long-time limit, each node of the network has equal probability of being occupied by the random walker. The other length scale determining the behavior of the random walker is $\xi = 1/p$, the average

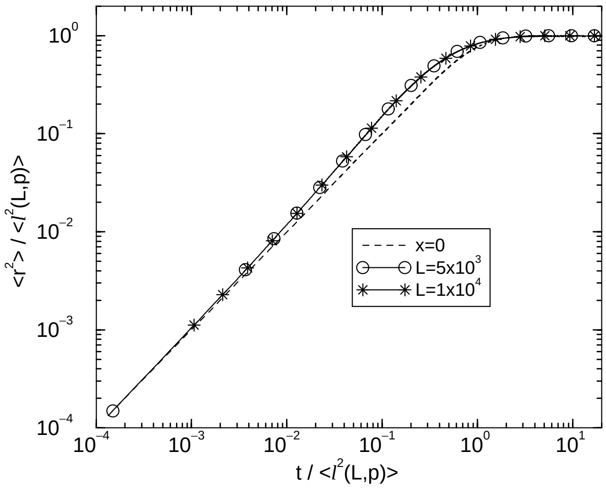

FIG. 4. The mean-square displacement, $\langle r^2\rangle$, for a random walker on SWN's with $x = 100$ (upper curve), using $L = 5 \times 10^3$ (circles and full line), and 1×10^4 (stars and full line). By contrast we also show $\langle r^2\rangle$ for a regular network ($x = 0$; dashed line).

distance the walker travels to reach a shortcut. Note that in the scaling limit, $\overline{\ell^2}(L,p)/L^2$, and consequently $\xi^2/\overline{\ell^2}(L,p)$, are functions of x only. Correspondingly we can write down the scaling ansatz

$$\langle r^2\rangle = \overline{\ell^2} R\left(\frac{t}{\overline{\ell^2}}; x\right). \tag{11}$$

Because of the diffusive behavior for small y and saturation for large y, we expect $R(y) = y$ for $y \ll \xi^2/\overline{\ell^2}$, and $R(y) = 1$ for $y \gg 1$. We have numerically confirmed the scaling collapse implied by Eq. (11) for a wide range of x values, and an example is shown in Fig. 4. Note that we can readily calculate the length scale $\overline{\ell^2}$ using our functional form for $f(m)$, illustrating how knowledge of the underlying structure helps in understanding the dynamical properties.

In summary, we have given exact results which are valid for SWN's with an arbitrary degree distribution superimposed on a regular lattice. We have also shown that in the small-world limit ($x \gg 1$), the structure of the original SWN's can be characterized by a single function $g(y)$ and the mean value $\langle m\rangle$, and we provide empirical fitting forms for both quantities. In addition, we have shown that the mean-square displacement of a random walker exhibits scaling properties which follow from those of the underlying SWN. The structural aspects of SWN's discussed in this work should be useful in understanding many other properties of dynamics on SWN's.

The authors would like to thank S. V. Barabash for helpful comments. This work has been supported by NASA through Grant No. NCC8-152 and NSF through Grant No. DMR01-04987 (E. A. and D. S.). R. V. K. would like to acknowledge the support of the U.S. Department of Energy, Office of Science, Division of Materials Research. Computational support was provided by the Ohio Supercomputer Center, and the Norwegian University of Science and Technology (NTNU). E. A. also thanks the Santa Fe Institute (SFI) for their hospitality.

*Email address: Almaas.1@osu.edu

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] H. R. Bernard, P. D. Kilworth, M. J. Evans, C. McCarthy, and G. A. Selley, Ethnology **27**, 155 (1988).
[3] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); M. E. J. Newman, *ibid.* **64**, 016132 (2001).
[4] J. E. Cohen, F. Briand, and C. M. Newman, *Community Food Webs: Data and Theory* (Springer-Verlag, Berlin, 1990).
[5] R. J. Williams and N. D. Martinez, Nature (London) **404**, 180 (2000).
[6] B. A. Huberman and L. A. Adamic, Nature (London) **401**, 131 (1999).
[7] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. J. Lukose, Science **280**, 95 (1999).
[8] M. E. J. Newman and D. J. Watts, Phys. Lett. A **263**, 341 (1999).
[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[10] A.-L. Barabási, R. Albert, and H. Jeong, Physica (Amsterdam) **272A**, 173 (1999).
[11] M. E. J. Newman and D. J. Watts, Phys. Rev. E **60**, 7332 (1999).
[12] M. E. J. Newman, C. Moore, and D. J. Watts, Phys. Rev. Lett. **84**, 3201 (2000).
[13] R. V. Kulkarni, E. Almaas, and D. Stroud, Phys. Rev. E **61**, 4268 (2000).
[14] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[15] R. Albert and A.-L. Barabási, cond-mat/0106096.
[16] S. N. Dorogovtsev and J. F. F. Mendes, cond-mat/0106144.
[17] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[18] A. R. Puniyani, R. J. Lukose, and B. A. Huberman, cond-mat/0107212.
[19] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[20] C. P. Warren, L. M. Sander, and I. M. Sokolov, cond-mat/0106450.
[21] M. E. J. Newman, I. Jensen, and R. M. Ziff, cond-mat/0108542.
[22] We do not indicate the dependence on the network parameters L and p, but include it later wherever needed for clarity.
[23] A. D. Barbour and G. Reinert, cond-mat/0006001.
[24] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions* (J. Wiley & Sons, New York, 1995), Vol. 2.
[25] R. Monasson, Eur. Phys. J. B **12**, 555 (1999).
[26] S. Jespersen, I. M. Sokolov, and A. Blumen, Phys. Rev. E **62**, 4405 (2000).
[27] S. A. Pandit and R. E. Amritkar, Phys. Rev. E **63**, 041104 (2001).
[28] Additional properties of the random walk on SWNs will be discussed elsewhere.