Dynamics of epidemics on random networks

M. Marder*

Center for Nonlinear Dynamics and Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA

(Received 9 August 2006; revised manuscript received 16 January 2007; published 13 June 2007)

This paper examines how diseases on random networks spread in time. The disease is described by a probability distribution function for the number of infected and recovered individuals, and the probability distribution is described by a generating function. The time development of the disease is obtained by iterating the generating function. In cases where the disease can expand to an epidemic, the probability distribution function is the sum of two parts; one that is static at long times, and another whose mean grows exponentially. The time development of the mean number of infected individuals is obtained analytically. When epidemics occur, the probability distributions are very broad, and the uncertainty in the number of infected individuals at any given time is typically larger than the mean number of infected individuals.

DOI: 10.1103/PhysRevE.75.066103 PACS number(s): 89.75.Hc, 64.60.Ak, 89.75.Da

I. INTRODUCTION

A series of papers by Watts and Strogatz [1], Pastor-Satorras and Vespignani [2,3], Meyers *et al.* [4–6], Newman and co-workers [7–11], Stanley and co-workers [12], Albert and Barabási [13], Cohen *et al.* [14], Moreno and Vazquez [15,16], and Voltz [17] applies methods from graph theory and percolation theory [18,19] to the spread of disease on random networks. These papers mainly study the final state of a population once the disease has run its course, with all individuals susceptible but uninfected, or recovered. This paper shows how to apply the same analytical techniques to the dynamics of the epidemic and find how the number of infected individuals varies in time.

A starting point for this study was to clear up a curious point arising when an epidemic is possible, but not certain. Newman *et al.* [9] find a probability distribution function P_l that l individuals have been infected, and they show that $u \equiv \Sigma_l P_l < 1$. They determine u from a self-consistent equation, and interpret P_l as describing the probability of a finite outbreak that does not grow to system size. The remaining probability, $1-u$, is contained in an outbreak that fills the whole system. This interpretation is puzzling. Since l can have any size, why does P_l describe only finite outbreaks? How do the self-consistent equations determining u figure out how to find only these finite outbreaks, and discard the larger ones? The authors assert that the system-size outbreaks would contain loops that invalidate the formalism they are employing, but how does the formalism know this? These questions are resolved when one examines the probability distribution after n time steps, $P_l^{(n)}$. One finds that the probability distribution is the sum of two pieces. The first piece $Q_l^{(n)}$ converges to a time-independent function Q_l in the long-time limit, with $\Sigma_l Q_l < 1$. The second piece $R_l^{(n)}$ never stops evolving. Its mean and width grow exponentially. So long as the mean of $R_l^{(n)}$ is much smaller than the total system size, it can be described by standard generating function techniques, and this description is not invalidated by the presence of loops. Thus, the generating function formalism has been finding Q_l and the reason this function emerges is that $P_l^{(n)}$ converges to Q_l pointwise, although at any given time step n a finite fraction of $P_l^{(n)}$ is contained in a very broad tail of the distribution that has formed out in front of Q_l. Techniques essentially identical to those used previously to describe Q_l can be used to analyze $R_l^{(n)}$. In particular, one can find closed-form expressions for the mean number of people infected at time n. When an epidemic is possible, both the mean and width of $R_l^{(n)}$ grow exponentially in time. In general, ones uncertainty about precisely how many people will be infected in the future grows as fast as or faster than the number of diseased individuals.

II. DYNAMICAL EQUATIONS

Consider a random network in which the probability distribution of nodes with k edges is p_k. Following Newman *et al.* [9], the generating function for the distribution of nodes is

$$G_0(x) \equiv \sum_{k=0}^{\infty} p_k x^k. \tag{1}$$

Consider choosing a random edge in the system. The probability that the node reached by this edge will have l new edges in addition to the one chosen to start with is generated by the coefficient of x^l in

$$G_1(x) \equiv \frac{G_0'(x)}{G_0'(1)}. \tag{2}$$

Consider conventional susceptible-infected-recovered dynamics on this network [9]. At each time step, uninfected nodes connected by an edge to infected nodes become infected in turn with unit probability. Let $P_l^{(n)}$ give the probability that a grand total of l individuals has been infected after n time steps, and let the generating function for $P_l^{(n)}$ be

*Electronic address: marder@mail.utexas.edu

$$H^{(n)}(x) \equiv \sum_{k=0}^{\infty} P_l^{(n)} x^l. \tag{3}$$

Imagine starting with a single infected individual. At step 0, one has $H^{(0)}(x)=x$. At the next time step, the generating function for the total number of individuals infected is

$$H^{(1)}(x) = xG_0(x), \tag{4}$$

since $G_0(x)$ gives the probability that a given node has 0, 1, 2,… edges, and one multiplies by x because one began with one infected individual. Each of the edges departing the first one reaches some other node. The probability that it will have l additional edges leaving it is given by $G_1(x)$. Using the *powers* property in Sec. II A of Ref. [9], one has

$$H^{(2)}(x) = xG_0(xG_1(x)). \tag{5}$$

Continuing in this fashion, one has

$$H^{(n)}(x) = H^{(n-1)}(xG_1(x)). \tag{6}$$

This expression is inconvenient for numerical work, so we define instead

$$F^{(0)}(x) = 1, \tag{7a}$$

$$F^{(n)}(x) = G_1(xF^{(n-1)}(x)), \tag{7b}$$

$$H^{(n)}(x) = xG_0(xF^{(n-1)}(x)). \tag{7c}$$

To extract the probability distribution function from a generating function $H(z)$, note that from Cauchy's theorem

$$P_l = \frac{1}{2\pi i}\oint \frac{dz}{z^{l+1}} H(z) = \int_0^1 d\theta\, e^{-2\pi i l\theta} H(e^{2\pi i\theta}). \tag{8}$$

Suppose now that H has been evaluated around the unit circle at M points, with $\theta_m = m/M$, $m \in [0, M-1]$, and let

$$H_m = H(e^{2\pi i\theta_m}). \tag{9}$$

Then one has

$$P_l = \frac{1}{M}\sum_{m=0}^{M-1} e^{-2\pi i l m/M} H_m = \frac{1}{M}\mathcal{F}_{\mathrm{DFT}}(H,-1)[l], \tag{10}$$

where the last expression means that one takes the lth element of the inverse discrete fast Fourier transform. Using Eqs. (7) and employing Eq. (10) to obtain probabilities P_l, one easily obtains hundreds of iterates of the map, for hundreds of thousands of values of l.

III. STATIC AND GROWING DISTRIBUTIONS

Some results of solving Eqs. (7) appear in Fig. 1. Figure 1(a) shows distributions resulting from the polynomial $G_0(x)=0.7x+0.2x^2+0.05x^3+0.04x^4+0.01x^5$. The threshold for an epidemic is determined by $z_2 > z_1$ [2,9,20,21], where $z_1 = G_0'(1)$ is the average number of neighbors of each node, and $z_2 = G_1'(1)z_1$ is the average number of second neighbors. In the present case, $z_1=1.46$ and $z_2=1.38$, so the infection is

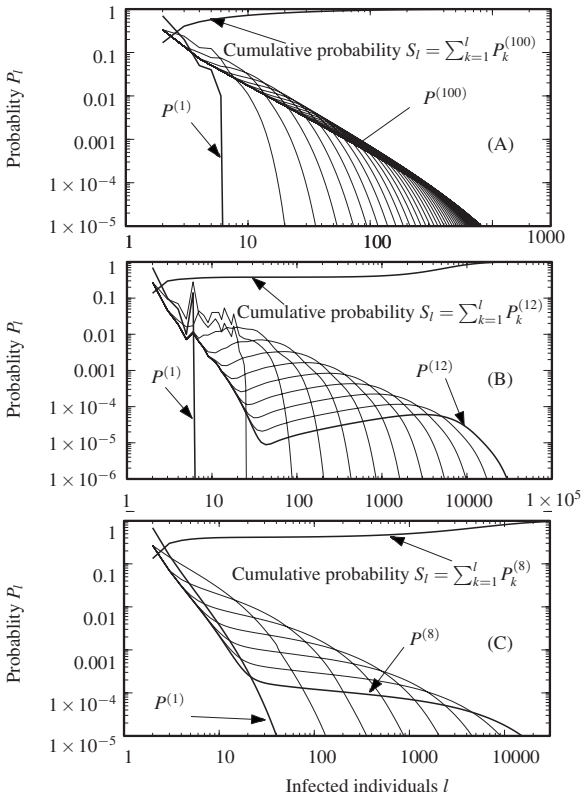

FIG. 1. (a) Dynamical evolution of Eq. (7c) in a case where the average number of second neighbors z_2 is less than the average number of neighbors z_1, and there is no epidemic. The map is iterated 100 times. (b) Dynamical evolution of Eq. (7c) in the case where $z_2 > z_1$, so one expects the existence of a giant component. The map is iterated 12 times. (c) Similar to (b), but now using a broader probability distribution. The epidemic grows quickly and only eight iterations are displayed.

contained, and the probability distribution converges to a definite limit enclosing unit probability. The upper curve shows the cumulative sum $S_l = \Sigma_{k=1}^{l} P_k^{(100)}$. The mean number of people infected after 100 iterations is 27, but the distribution is broad; for example, there is a 1% chance that more than 480 people will be infected. Figure 1(b) shows results from the polynomial $G_0(x)=0.7x+0.1x^2+0.05x^3+0.01x^4+0.14x^5$, which gives $z_1=1.79$, $z_2=3.42$. Since $z_2 > z_1$, an epidemic is possible. One can compute the probability of an epidemic spiraling out of control following [9]; also see Eq. (12). There is a root of $G_1(u)-u$ at $u=0.492$ and $G_0(u)=0.3790$. This computation predicts a 37.9% chance that the disease will run its course without becoming an epidemic. The upper curve in Figure 1(b) shows the cumulative sum $S_l = \Sigma_{k=1}^{l} P_k^{(12)}$, and there is a broad plateau where this sum has reached 0.38. The mean number of infected individuals after 11 iterations is 4650 but there is a 1% chance that more than 26 000 people will be infected. Figure 1(c) uses the probability distribution $p_0=0$, $p_k \propto k^{-\alpha} e^{-k/\kappa}$ with $\alpha=2$ and $\kappa=20$. Now $z_1=1.8$, $z_2=5.3$, and the epidemic grows even more rapidly. There is a 41% chance that the epidemic will be contained. The mean number of infected individuals after seven steps is 5500, but there is a 1% chance that more than 50 000 will be infected.

Inspection of Figs. 1(b) and 1(c) confirms that, when there is the possibility of an epidemic, the probability distribution does indeed split into two components. The first component Q_l is static in the long-time limit and describes the probability that spread of disease terminates with a number of infected individuals much smaller than the total population. The second component $R_l^{(n)}$ continues to evolve forever. From a formal point of view, the definition of Q_l is

$$Q_l = \lim_{n\to\infty} \int d\theta\, e^{-2\pi i l\theta} H^{(n)}(e^{2\pi i\theta}). \tag{11}$$

For any fixed l, this limit converges. Then $R_l^{(n)}$ can be defined as $R_l^{(n)} = P_l^{(n)} - Q_l$. One can similarly decompose the probability distribution resulting from $F^{(n)}$ into static and evolving components. To see now how the probability of not participating in the epidemic emerges from self-consistent equations, define $F^{\infty}(x) \equiv \lim_{n\to\infty} F^{(n)}(x)$. This limit exists for any $x<1$, since large powers of $x<1$ in the power series for $F^{(n)}$ suppress the parts of $F^{(n)}$ that are continuing to evolve. Return to Eq. (7b) and write

$$\lim_{x\to 1}\lim_{n\to\infty} F^{(n)}(x) - G_1(xF^{(n-1)}) = 0 \tag{12}$$

$$\Rightarrow \lim_{x\to 1} F^{\infty}(x) - G_1(xF^{\infty}) = 0$$

$$\Rightarrow u = G_1(u) \text{ with } u \equiv \lim_{x\to 1} F^{\infty}(x).$$

Finally $G_0(u) = \lim_{x\to 1}\lim_{n\to\infty} H^{(n)}(x)$ gives the probability that the disease does not spiral into an epidemic. The underlying probability distribution assumes that all individuals in the network are equally likely to introduce the disease, and the disease results from a single first instance. Percolation theory teaches that the giant component of the network is connected. Therefore, $G_0(u)$ also gives the percentage of the population that does not belong to the giant component and hence remains uninfected in the event of an epidemic. This conclusion is only as robust as the two assumptions going into it.

Figure 2 shows an explicit decomposition of the data in Fig. 1(b) into components Q and R. The task is carried out by taking the final curve in Fig. 1(b) and noticing that it has converged to a static value up to around $l=32$ (the precise cut point does not matter much) and is continuing to evolve for larger l. For $l>32$, Q_l is estimated by a power-law fit. The area under Q_l found this way is 0.3791 which compares well with the value predicted by Eq. (12) of 0.3790.

IV. SIZE OF INFECTED CLUSTER

One can work out analytically the average size of the infected and recovered cluster as a function of time. Note that $F^{(n)}(1)=1$ and let

$$M_n = \frac{d}{dx} F^{(n)}(x)\big|_{x=1}.$$

Then

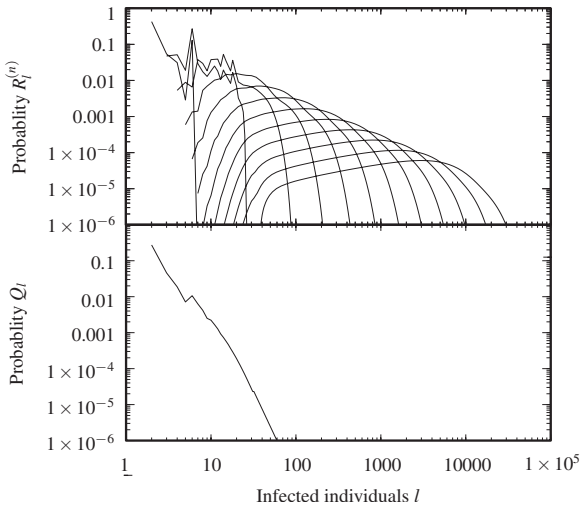

FIG. 2. Decomposition of the data in Fig. 1(b) into static and growing components Q_l and $R_l^{(n)}$. This is basically done by finding Q_l and subtracting it from successive P_l's. Specifically, compute $P_l^{(12)}$, set $Q_l = P_l^{(12)}$ for $l \leq 32$, fit Q_l to a power law for $l>32$, and subtract Q_l obtained in this way from distributions $P_l^{(1)}, \ldots, P_l^{(11)}$.

$$M_n = G_1'(1)[F^{(n-1)}(1) + M_{n-1}] = \frac{z_2}{z_1}(1 + M_{n-1}). \tag{13}$$

Using $M_0=0$, one can solve this iterated map exactly as a power series, which has the compact final expression

$$M_n = \frac{z_2}{z_1}\sum_{l=0}^{n-1}\left(\frac{z_2}{z_1}\right)^l = \frac{z_2}{z_1}\left(\frac{1-(z_2/z_1)^n}{1-z_2/z_1}\right). \tag{14}$$

Then the average number of individuals in the cluster is

$$\langle l\rangle_n = \frac{d}{dx} H^{(n)}(x)\big|_{x=1} = 1 + z_1\left[1 + \frac{z_2}{z_1}\left(\frac{1-(z_2/z_1)^{n-1}}{1-z_2/z_1}\right)\right]. \tag{15}$$

If $z_2<z_1$, one obtains the expected result [8,20,21] for large n that

$$\langle l\rangle = 1 + z_1\left(1 + \frac{z_2}{z_1 - z_2}\right) = 1 + \frac{z_1^2}{z_1 - z_2}. \tag{16}$$

In the opposite case, $z_2>z_1$, Eq. (14) becomes $M_n \approx (z_2/z_1)^{n+1}/(1-z_2/z_1)$ and for large n the average size of the infected population is

$$\langle l\rangle_n \sim \frac{z_1(z_2/z_1)^n}{z_2/z_1 - 1}. \tag{17}$$

The width of the distribution is proportional to the mean. The dominant contribution to $\langle l^2\rangle_n$ at large n is

$$\sqrt{\langle l^2\rangle_n - \langle l\rangle_n^2} \sim \frac{(z_2/z_1)^n}{z_2/z_1 - 1}\sqrt{z_2 - z_1^2 + \frac{z_2 G_1''(1)}{(z_2/z_1 - 1)}}. \tag{18}$$

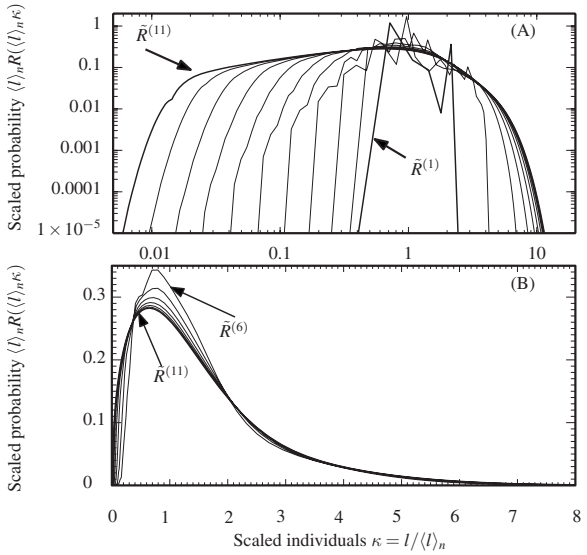

FIG. 3. Plot of $\langle l\rangle P_{\kappa\langle l\rangle}$, using the generating functions from Fig. 1(b) iterated ten times, appears to be converging on a scaling form. Convergence for small values of $\kappa=l/\langle l\rangle_n$ is pointwise on a logarithmic scale (a) and uniform on a linear scale (b).

V. WHEN INFECTION IS NOT CERTAIN ACROSS AN EDGE

Newman [8] describes the case where infection is not certain across an edge connecting two nodes, but occurs with probability T. In this case, the probability of infecting neighbors starting with a randomly chosen node is generated by

$$G_0(1+T(x-1)), \tag{19}$$

the probability of infecting neighbors starting with a randomly chosen edge, excluding the incoming edge is generated by

$$G_1(1+T(x-1)), \tag{20}$$

and by employing these two generating functions, the evolution equation (7) is unchanged, while Eq. (15) for the average size of the outbreak generalizes to

$$\langle l\rangle_{n+1} = \frac{d}{dx}H^{(n+1)}(x)\big|_{x=1} = 1 + \frac{z_1^2T - z_2T(z_2T/z_1)^n}{z_1 - z_2T}. \tag{21}$$

Essentially z_2 is replaced by T_{z_2}.

VI. INDIVIDUALS INFECTED AT EACH TIME STEP

Another interesting quantity to track is the probability of infecting l new individuals at each time step. This can be done by adding a subscript to the variable x that tracks the time step at which an individual has entered the cluster. Doing so, one has

$$H^{(0)}(x_0) = x_0, \quad H^{(1)}(x_0,x_1) = x_0G_0(x_1), \tag{22}$$

$$H^{(2)}(x_0,x_1,x_2) = x_0G_0(x_1G_1(x_2)). \tag{23}$$

Continuing in this fashion, one has

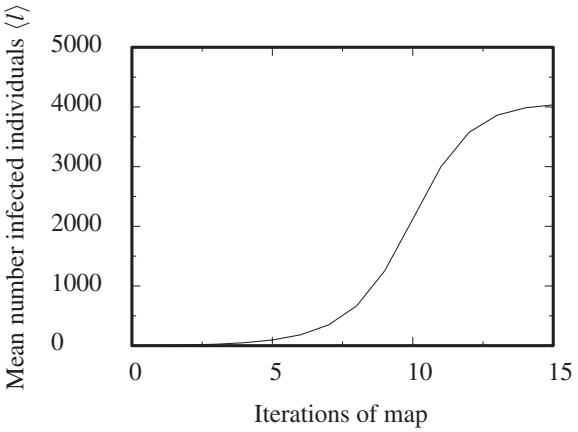

FIG. 4. Time evolution of mean number of infected individuals in the presence of saturation. The simulation is identical to Fig. 1(b), but the size of the giant component is taken to be 6554, and all weight in the distribution function greater than 6554 is attributed to that number of infected individuals, with the result that the mean number of infected individuals reaches a plateau at around 4000.

$$H^{(n)}(\vec{x}) = H^{(n-1)}(x_0,x_1,\ldots,x_{n-2},x_{n-1}G_1(x_n)). \tag{24}$$

One recovers the results in Eq. (7c) by removing all the indices from the variables x. To focus upon the individuals infected at step n, just set all variables x_l to 1 except the last. Denote by $s_l^{(n)}$ the probability that l individuals have been infected at time step n, and let $J(x)$ be the generating function for this probability. Then

$$J^{(1)}(x) = G_0(x), \quad J^{(n)}(x) = J^{(n-1)}(G_1(x)). \tag{25}$$

One can now calculate the mean number of individuals infected at each time step, δl_n,

$$\delta l_n \equiv \sum s_l^{(n)} l, \quad \delta l_1 = z_1, \tag{26}$$

$$\delta l_2 = \frac{z_2}{z_1}z_1 \cdots \delta l_n = \left(\frac{z_2}{z_1}\right)^{n-1} z_1. \tag{27}$$

VII. SCALING FORM FOR EPIDEMIC

It would seem natural for the growing part of the probability distribution R_l to adopt a scaling form at long times. To capture the growing part of the distribution, one computes

$$R_l^{(n)} \approx \frac{1}{\langle l\rangle_n}\tilde{R}(\kappa) \quad \text{where } \kappa = l/\langle l\rangle_n. \tag{28}$$

As shown in Fig. 3, this scaling form does appear to describe R after sufficiently many iterations, On a logarithmic scale the tail of $\tilde{R}$ for small $\kappa=l/\langle l\rangle_n$ converges pointwise, but on a linear scale convergence is uniform.

VIII. SATURATION

The approach taken in this paper does not naturally lend itself to studying the saturation effects that occur when the total number of infected individuals becomes comparable to

the total population size. A simple estimate can be obtained by allowing the probability distribution to evolve, but choosing some number N for the population of the giant component and treating all weight in the probability distribution above N as describing an infection of the N individuals in the giant component. The results of such a computation appear in Fig. 4. It would be interesting to compare this admittedly crude technique with mean field methods such as that of Volz [17].

ACKNOWLEDGMENTS

I would like to thank Lauren Ancel Meyers for introducing me to this problem and making many helpful comments on the manuscript, and Shweta Bansal for patiently explaining to me some of the perplexing features. Thanks also to the National Science Foundation for support through Grant No. DMR-0401766.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 035108(R) (2002).
[3] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[4] L. A. Meyers, M. E. J. Newman, M. Martin, and S. Schrag, Emerg. Infect. Dis. **9**, 204 (2003).
[5] L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham, J. Theor. Biol. **232**, 71 (2005).
[6] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, J. Theor. Biol. **240**, 400 (2006).
[7] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2003).
[8] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[9] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[10] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[11] M. E. J. Newman and D. J. Watts, Phys. Rev. E **60**, 7332 (1999).
[12] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).
[13] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[14] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[15] Y. Moreno and A. Vazquez, Eur. Phys. J. B **31**, 265 (2003).
[16] A. Vazquez, Phys. Rev. E **74**, 056101 (2006).
[17] E. Volz, e-print arXiv:physics/0508160.
[18] D. Stauffer, Phys. Rep. **54**, 1 (1979).
[19] J. W. Essam, Rep. Prog. Phys. **43**, 833 (1980).
[20] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[21] M. Molloy and B. Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).