On the evolution of scale-free graphs

D.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim School of Physics and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea (Dated: December 24, 2003)

We study the evolution of random graphs where edges are added one by one between pairs of weighted vertices so that resulting graphs are scale-free with the degree exponent γ . We use the branching process approach to obtain scaling forms for the cluster size distribution and the largest cluster size as functions of the number of edges L and vertices N. We find that the process of forming a spanning cluster is qualitatively different between the cases of $\gamma > 3$ and $2 < \gamma < 3$. While for the former, a spanning cluster forms abruptly at a critical number of edges L_c , generating a single peak in the mean cluster size $\langle s \rangle$ as a function of L, for the latter, however, the formation of a spanning cluster occurs in a broad range of L, generating double peaks in $\langle s \rangle$.

PACS numbers: 89.70.+c, 89.75.-k, 05.70.Jk

Recently, many studies have been performed on complex networks. Such studies are mostly influenced by the random graph theory proposed by Erdős and Rényi (ER) [1]. In the ER model, N number of vertices are present from the beginning and edges are added one by one in the system, connecting pairs of vertices selected randomly. A remarkable result ER obtained is that a giant cluster of size O(N), a spanning cluster, appears abruptly when L reaches its threshold value L_c , which is O(N). Note that the formation of such a spanning cluster can be viewed as a percolation transition occurring at the critical probability $p_c = 2L_c/N(N-1) = O(1/N)$. While the ER graph is pioneering, it is too random, and various properties of the ER graph are not in accordance with those of complex networks recently discovered in real world. For example, the distribution of the number of edges incident on each vertex, called the degree distribution, is Poissonian for the ER graph, while it follows a power law for many realworld networks, called scale-free (SF) networks [2, 3, 4].

It was shown that a SF network can be generated by following a similar way to used in the ER model [5, 6]. N number of vertices are present from the beginning and edges are added one by one. For SF networks, however, each vertex with the index of an integer i ($i = 1, \dots, N$) is not identical, but is assigned a normalized weight $w_i = i^{-\alpha} / \sum_{j=1}^{N} j^{-\alpha}$ with a control parameter $\alpha \in [0,1)$. Each edge connects a pair of vertices (i,j) selected with probability $w_i w_j$. Thus the ER graph is generated with $\alpha = 0$. The process of adding edges is repeated until the total number of edges in the system reaches L. This process of constructing networks is called the static model. When L is in the intermediate regime $L_{\ell} \ll L \ll L_{u} = L_{\ell}N$, with L_{ℓ} being specified below, the degree distribution follows a power law, $p_d(k) \sim k^{-\gamma}$ with $\gamma = 1 + 1/\alpha \in (2, \infty)$. However, when $L \lesssim L_{\ell}$ ($L \gtrsim L_u$), the network is too sparse (dense), so that the degree distribution does not follow the SF behavior. The static model was introduced to generate SF networks with various y, being used to study various problems. However, it has not been studied yet how clusters evolve as the number of edges L increases, which is the goal of this Letter.

The percolation problem of SF networks has been studied [7, 8, 9, 10], reversely, that is, by removing randomly-selected vertices as well as their attached edges. In this Letter,

we study how the cluster evolution of SF graphs proceeds as edges are added. Besides confirming the previous results in Ref. [8, 10], we show that the process of forming a spanning cluster for the case of $2 < \gamma < 3$ is fundamentally different from that of $\gamma > 3$. When $\gamma > 3$, as in the case of the ER graph, there exists a critical number of edges L_c at which a spanning cluster forms through many small clusters coalescing, and the mean cluster size diverges at finite L_c/N in the thermodynamic limit. In other words, a percolation transition occurs at L_c . When $2 < \gamma < 3$, however, large or small clusters grow in a similar manner as a whole without sudden coalescence occurring. As a result, the mean cluster size does not diverge anywhere, but instead exhibits two peaks at L_{p1} and L_{p2} . Near L_{p1} , some small clusters merge together forming a much larger one, but it does not span the entire system. After passing L_{p1} , the largest cluster as well as smaller ones continue to grow, and the largest one becomes as large as O(N)around L_{p2} . Throughout this Letter, we will denote the case of $\gamma > 4$ as (I), $3 < \gamma < 4$ as (II), and $2 < \gamma < 3$ as (III). The schematic diagram of the cluster formation is shown in Fig. 1. We obtain characteristic numbers of edges for each case as a function of N and summarize them in the phase diagram shown in Fig. 2. Moreover, we derive scaling forms for the cluster size distribution and the largest cluster size analytically and numerically.

Power-law degree distribution — The probability $p_{d,i}(k)$ that a vertex i has degree k follows approximately a Poissonian form as $p_{d,i}(k) \simeq \langle k_i \rangle^k \exp(-\langle k_i \rangle)/k!$ for large k and large k. Here the average degree of the vertex k is given by k is k in k. Where k is k is k in k is k in k

$$p_d(k) = \frac{1}{N} \sum_{i=1}^{N} p_{d,i}(k) \simeq c k^{-\gamma},$$
 (1)

where c is given as $c \simeq (\gamma - 1)[K(L)]^{\gamma - 1}/N$.

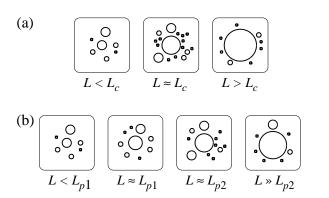


FIG. 1: Schematic picture for the comparison of cluster evolution between (I,II) (a) and (III) (b).

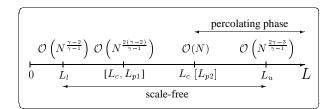


FIG. 2: Schematic phase diagram of the static model. The SF behavior of the degree distribution appears between L_{ℓ} and L_{u} . A spanning cluster emerges at L_{c} for (I,II), and around L_{p2} for (III). The quantities in [...] are only for (III).

Branching process approach — As L increases beyond L_{ℓ} , small clusters form. However, clusters are still sparse and of tree structure [1]. The formation of such sparse clusters can be understood through the multiplicative branching process approach [11]. Here we introduce the probability distribution P(s) as the number of vertices belonging to s-clusters, clusters with s vertices, divided by N [12]. Also we define another probability distribution R(s) as the number of edges followed by s-clusters divided by 2L. The generating functions of those quantities are defined as $\mathcal{P}(z) = \sum_s P(s) z^s$ and $\mathcal{R}(z) = \sum_s R(s) z^s$, respectively. Both summations run over finite clusters only [11], and then when clusters are sparse the following relations hold:

$$\mathcal{R}_{\cdot}(z) = zf(\mathcal{R}_{\cdot}(z))$$
 and $\mathcal{P}(z) = zg(\mathcal{R}_{\cdot}(z)),$ (2)

where $g(\omega) = \sum_{k=0}^{\infty} p_d(k) \omega^k$ and $f(\omega) = g'(\omega)/\langle k \rangle$ with $\langle k^n \rangle = \sum_{k=0}^{\infty} k^n p_d(k)$.

To apply Eq. (2) to the static model, we use the following form valid in the limit $1 - \omega \ll 1$; $g(\omega) \simeq (1/N)\sum_{i=1}^{N} \exp[\langle k_i \rangle (\omega - 1)]$. Then it is obtained that $z = \omega + \sum_{n=1}^{\infty} a_n (1 - \omega)^n$ for $K(L)(1 - \omega) \leq 1$, where $a_1 = f'(1)$, a_2 a negative constant, and so on, while

$$z = \omega + \sum_{n=1}^{\lfloor \gamma - 2 \rfloor} a_n (1 - \omega)^n + A (1 - \omega)^{\gamma - 2} + \dots$$
 (3)

for $K(L)(1-\omega)\gg 1$, where $\lfloor x\rfloor$ is the floor function of x and $A=\Gamma(3-\gamma)[K(L)/N^{1/(\gamma-1)}]^{\gamma-2}$. The generating function $\omega=$

TABLE I: Cluster size distribution P(s). Here $\Delta \equiv (L-L_c)/L_c$, and the scaling exponents (τ,σ) are given as (3/2,1/2) for (I) and $((\gamma-1)/(\gamma-2),(\gamma-3)/(\gamma-2))$ for (II).

	(i)	(ii)	(iii)	
	$(L < L_c)$	$(L=L_c)$	$(L > L_c)$	
(I,II)	$s^{-\tau}$ $(s \ll s_c)$		$s^{- au}$	$(s \ll s_c)$
	$s^{-\tau}$ $(s \ll s_c)$ $(\Delta s)^{-(\gamma-1)}$ $(s \gg s_c)$		$\exp(-s/s_c)$	$(s\gg s_c)$
	$s_c \sim \Delta ^{-1/\sigma}$			
(III)	$(Ns/L)^{-(\gamma-1)}$ $(s \ll K(L))$		$(Ns/L)^{-(\gamma-1)}$	$(s \ll s_c)$
	$N^{-\frac{4-\gamma}{2(\gamma-1)}}e^{-s/s_c}$ $(s\gg K(L))$		$(Ns/L)^{-(\gamma-1)} \qquad (s \ll s_c)$ $(L/N)^{\frac{4-\gamma}{2(3-\gamma)}} e^{-s/s_c} (s \gg s_c)$	
	$s_c \sim L^2/(N^{3(\gamma-2)/(\gamma-1)} \Delta ^2)$		$s_c \sim (L/N)^{-(\gamma-2)/(3-\gamma)}$	

 $\mathcal{R}(z)$ can be obtained by inverting $z = \omega/f(\omega)$, and $\mathcal{P}(z)$ is then obtained by using $\mathcal{P}(z) = zg(\mathcal{R}(z))$.

Critical point — The values of $\mathcal{P}(1)$ and $\mathcal{R}(1)$ are 1 only when $f'(1) \leq 1$, while they are smaller than 1 when f'(1) > 1. Thus the condition f'(1) = 1, which is the same as the condition $\langle k^2 \rangle / \langle k \rangle = 2$ [13], leads to a characteristic number of edges L_c ,

$$L_c = \frac{1}{2} \frac{\zeta_N[1/(\gamma - 1)]^2}{\zeta_N[2/(\gamma - 1)]}.$$
 (4)

For large N, L_c is $N(\gamma-1)(\gamma-3)/(2(\gamma-2)^2)$ for (I,II), and $N^{2(\gamma-2)/(\gamma-1)}(\gamma-1)^2/(2(\gamma-2)^2\zeta_N[2/(\gamma-1)])$ for (III). We will show later that for (I,II), a spanning cluster appears at L_c , but for (III), the size of the largest cluster does not reach O(N) at L_c . Thus more edges are needed to generate a spanning cluster.

Cluster size distribution — The asymptotic behaviors of P(s) and R(s) for large s can be obtained from the singular parts of $\mathcal{P}(z)$ and $\mathcal{R}(z)$ as $z \to 1$, respectively. In the static model, the characteristic behaviors of $\mathcal{P}(z)$ and $\mathcal{R}(z)$ depend on the degree exponent γ , classifying them into the three cases, (I), (II), and (III). Each case is again classified into (i) subcritical $(L < L_c)$, (ii) critical $(L = L_c)$, and (iii) supercritical cases $(L > L_c)$. Our results for P(s) are listed in Table I for each

Emergence of spanning cluster — In cases of (i) and (ii), the size of the largest cluster S is obtained self-consistently through the relation, $\sum_{s\neq S} P(s) = 1 - S/N$ using P(s) in Table I. For example, when $P(s) \sim s^{-\tau}$, S is obtained to be $S \sim N^{1/\tau}$. Thus when $L = L_c$ (ii), $S \sim N^{2/3}$ for (I) [1] and $N^{(\gamma-2)/(\gamma-1)}$ for (II) [14]. For (III), using $P(s) \sim (Ns/L_c)^{1-\gamma}$, we obtain $S \sim N^{(\gamma-2)/(\gamma-1)}$, but we show below that this is not the incipient spanning cluster. The size of the largest cluster for the subcritical case (i) is $S \sim \max\{K(L)/|\Delta|, |\Delta|^{-1/\sigma}\}$ for (I), $\sim K(L)/|\Delta|$ for (II), and $\sim K(L)$ for (III), with $\Delta = (L-L_c)/L_c$.

In case of (iii), the theory of the multiplicative branching process yields the size of an *infinite* cluster through $N(1-\mathcal{P}(1))$. Thus it can be identified with the largest cluster if it is larger than S at L_c . From Eq. (2), $1-\mathcal{P}(1) \sim$

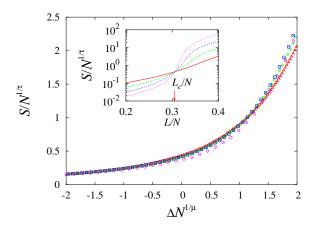


FIG. 3: Data collapse of $S/N^{1/\tau}$ versus $\Delta N^{1/\mu}$ for Eq. (5) with $\gamma=3.6$, and $N=10^4(\triangle)$, $10^5(\lozenge)$, $10^6(\square)$, and $10^7(\bigcirc)$. Here $\tau=(\gamma-1)/(\gamma-2)=13/8$, $\mu=(\gamma-1)/(\gamma-3)=13/3$, and $L_c/N\simeq0.305$ from Eq. (4) are used. Inset: the same data plotted versus L/N cross at $L/N\simeq L_c/N$.

 $(2L/N)(1-\Re(1))$, and the value of $1-\Re(1)=1-\omega$ is obtained by solving Eq. (3) with z=1. Thus we obtain, $S\sim\Delta^\beta N$ with $\beta=1$ for (I) and $\beta=1/(\gamma-3)$ for (II) in the regime of $\Delta N^{1/\mu}\gg 1$, while $S\sim (L/N)^{1/(3-\gamma)}N$ for (III) when $\Delta\gg 1$, where a new scaling exponent $\mu=3$ for (I) and $(\gamma-1)/(\gamma-3)$ for (II) was used. The behaviors of (i), (ii), and (iii) lead to a scaling ansatz for (I,II),

$$S \sim N^{1/\tau} \Psi_{(I,II)}(\Delta N^{1/\mu}),$$
 (5)

where τ is given in Table I, and $\Psi_{(I,II)}(x)$ is constant for $|x| \ll 1$ and behaves as x^{β} for $x \gg 1$ and $|x|^{-\delta}$ for $x \ll -1$ with $\delta = 2$ for (I) and 1 for (II). In the thermodynamic limit, a spanning cluster emerges if only $L \geq L_c$. On the other hand, for (III),

$$S \sim N^{(\gamma-2)/(\gamma-1)} \Psi_{(\text{III})}(\Delta),$$
 (6)

where $\Psi_{(\mathrm{III})}(x)$ is a constant for $|x| \lesssim 1$, behaving as (1+x) when $x \simeq -1$ and $x^{1/(3-\gamma)}$ for $x \gg 1$. The size of the largest cluster is $\mathcal{O}(N^{(\gamma-2)/(\gamma-1)})$, which is not as large as $\mathcal{O}(N)$, even when $L > L_c$, so that $L_c = \mathcal{O}(N^{2(\gamma-2)/(\gamma-1)})$ is not a percolation threshold. The largest cluster size S becomes $\mathcal{O}(N)$ only when L becomes as large as $\mathcal{O}(N)$.

To check such scaling behaviors of S, we perform numerical simulations for $\gamma=3.6$ and $\gamma=2.4$. As shown in Figs. 3 and 4, the data of $S/N^{(\gamma-2)/(\gamma-1)}$ with different N collapse with the scaling variables, $\Delta N^{1/\mu}$ for $\gamma=3.6$ (II), and Δ for $\gamma=2.4$ (III), respectively. For $\gamma=3.6$, $L_c/N\simeq0.305$ theoretically obtained in Eq. (4) is confirmed by the data crossing at $L_c/N\simeq0.306(2)$.

Mean cluster size — The difference in the cluster evolution for (I), (II), and (III) appears more apparently in the mean cluster size $\langle s \rangle$ defined as $\langle s \rangle \equiv \sum_{s \neq S} sP(s)$. The quantity $\langle s \rangle$ is similar to the susceptibility defined in the percolation theory but here we exclude the largest cluster even for $L < L_c$. For (I,II), as L increases, many small clusters grow by attaching

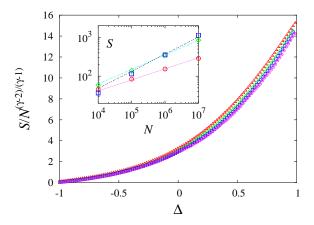


FIG. 4: Data collapse of $S/N^{(\gamma-2)/(\gamma-1)}$ versus Δ for Eq. (6) with $\gamma=2.4$ and $N=10^4(\triangle)$, $10^5(\lozenge)$, $10^6(\square)$, and $10^7(\bigcirc)$. Here $L_c/N^{4/7}\simeq 2.08$ from Eq. (4) is used. Inset: plot of S at L_c versus N for $\gamma=2.4(\circ)$, $2.6(\lozenge)$, and $2.8(\square)$, being in accordance with $S\sim N^{(\gamma-2)/(\gamma-1)}$ represented by the dotted, dashed-dotted, and dashed line, respectively.

edges, which continues up to $L=L_c$, and then a spanning cluster forms by the abrupt coalescence of those small clusters as shown in Fig. 1. Since we do not count the spanning cluster in calculating $\langle s \rangle$, $\langle s \rangle$ decreases rapidly as L passes L_c . Thus the mean cluster size exhibits a peak at $L=L_c$, which diverges in the thermodynamic limit $N\to\infty$. The scaling behaviors of S and P(s) in Table I lead to another scaling ansatz

$$\langle s \rangle = N^{1/\mu} \, \Phi(\Delta N^{1/\mu}), \tag{7}$$

where $\Phi(x)$ is a constant when $|x| \ll 1$, and behaves as x^{-1} when $x \gg 1$, and $|x|^{-1}$ for (I) and $|x|^{-(\gamma-3)/(\gamma-2)}$ for (II) when $x \ll -1$. Such behaviors can be confirmed with numerical data for $\gamma = 3.6$ in Fig. 5.

For (III), however, the mean cluster size does not diverge at any value of L but has two blunt peaks (Fig. 6). First, it has a small peak at $L = L_{p1}$, but it increases again as L increases beyond L_{p1} . Edges newly introduced either create new clusters of size larger than 1 or merge small clusters to the larger one with size not as large as O(N). When L reaches $L_{n2} = O(N)$ where the second peak arises, the largest cluster becomes as large as O(N). When L is near L_c , $\langle s \rangle$ can be evaluated through $\langle s \rangle = \sum_{s \neq S} s P(s)$, and it follows that $\langle s \rangle - 1 \sim \min\{s_c^{3-\gamma}, S^{3-\gamma}\}\$, where s_c is a characteristic cluster size defined in Table I, and the constant term 1 originates from the isolated vertices whose fraction is nearly 1. Since S (s_c) increases (decreases) with increasing L for $L > L_c$, $\langle s \rangle$ is maximal at $S = s_c$, occurring at $L_{p1} = bL_c = O(N^{2(\gamma-2)/(\gamma-1)})$ with b being a constant depending on γ . That is verified numerically as shown in the inset of Fig. 6. For $L_{p1} \ll L \ll N$, using $\langle s \rangle = \mathcal{P}'(1)$, we obtain $\langle s \rangle$ to be

$$\langle s \rangle \simeq 1 + \frac{2}{3 - \gamma} (\frac{L}{N}) - B \left(\frac{L}{N}\right)^{\frac{1}{3 - \gamma}} + C \left(\frac{L}{N}\right)^{\frac{\gamma - 1}{3 - \gamma}},$$
 (8)

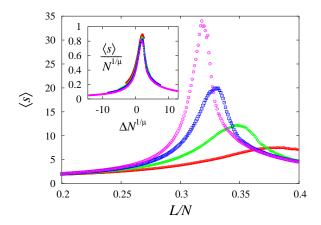


FIG. 5: Mean cluster size $\langle s \rangle$ as a function of L/N with $\gamma = 3.6$ for $N = 10^4(\triangle)$, $10^5(\lozenge)$, $10^6(\square)$, and $10^7(\bigcirc)$. The peak heights increase with N. The inset shows the data collapse of the rescaled mean cluster size $\langle s \rangle / N^{1/\mu}$ versus $\Delta N^{1/\mu}$.

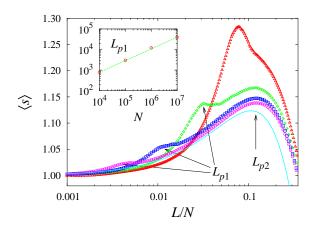


FIG. 6: Mean cluster size $\langle s \rangle$ as a function of L/N in semi-logarithmic scales with $\gamma=2.4$ for $N=10^4(\triangle)$, $10^5(\diamondsuit)$, $10^6(\square)$, and $10^7(\bigcirc)$. In addition to the peak at L_{p1} , another peak is shown at $L_{p2}\simeq 0.1N$ for $N=10^5$, 10^6 , and 10^7 , respectively. The measured values of $L_{p1}(\bigcirc)$ as a function of N are plotted in the inset together with the guide line whose slope is $2(\gamma-2)/(\gamma-1)$ for comparison. The solid line represents Eq. (8).

where $B=(5-\gamma)/(3-\gamma)[2((\gamma-2)/(\gamma-1))^{\gamma-2}\Gamma(3-\gamma)]^{1/(3-\gamma)}$ and $C=(2/(3-\gamma)[(2(\gamma-2)/(\gamma-1))^{\gamma-2}\Gamma(3-\gamma)]^{2/(3-\gamma)}-\Gamma(2-\gamma)[2(\gamma-2)/(\gamma-1)\Gamma(3-\gamma)]^{(\gamma-1)/(3-\gamma)}$. It exhibits a peak at $L_{p2}/N\simeq 0.1$ for $\gamma=2.4$, close to the location obtained by numerical simulations (Fig. 6).

Dense graph — The power-law degree distribution in Eq. (1) lasts up to $L \sim L_u \equiv \zeta_N[1/(\gamma-1)]N = O(N^{(2\gamma-3)/(\gamma-1)})$, around which the vertex i=1 is connected to nearly all vertices $(\langle k_1 \rangle \sim N)$ and the degree distribution $p_d(k)$ begins to develop a peak at k=N-1. As L increases beyond L_u , more vertices have such maximal number of edges, and the graph becomes denser.

Summary — We have studied how clusters of SF graphs are created and evolve as the number of edges increases. We obtained the cluster size distribution, the largest cluster size, and the mean cluster size as functions of the numbers of edges L and vertices N. Those quantities behave differently when $\gamma > 3$ and $2 < \gamma < 3$. For the former, a giant spanning cluster forms through a sudden coalescence of small clusters, exhibiting a percolation transition, while for the latter, it does gradually, and the mean cluster size shows double peaks at distinct numbers of edges, L_{p1} and L_{p2} . This result implies that the fragmentation process of SF graphs under random failures on edges is qualitatively similar to (different from) the one under intentional attack when $\gamma > 3$ (2 < γ < 3) [7, 8]. Finally, it is noteworthy that recently Aiello el al. [15] studied the possibility of forming a spanning cluster for given N and $L = (N/2)\zeta(\gamma - 1)/\zeta(\gamma)$ as a function of γ , and found that a spanning cluster can exist only for $\gamma < \gamma_c \simeq 3.48$. However, the way of constructing a SF graph in their model is different from ours.

The authors would like to thank S. Havlin for helpful comments on the manuscript. This work is supported by the KOSEF Grant No. R14-2002-059-010000-0 in the ABRL program.

- P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. 5, 17 (1960); Bull. Inst. Int. Stat. 38, 343 (1961).
- [2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
- [3] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. 51, 1079 (2002).
- [4] M.E.J. Newman, SIAM Review 45, 167 (2003).
- [5] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. 87, 278701 (2001).
- [6] G. Caldarelli, A. Capocci, P. De Los Rios, and M.A. Muñoz, Phys. Rev. Lett. 89, 258702 (2002); B. Söderberg, Phys. Rev. E 66, 066121 (2002).
- [7] R. Albert, H. Jeong, and A.-L. Barabási, Nature 406, 378 (2000)
- [8] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. 85, 4626 (2000); Phys. Rev. Lett. 86, 3682 (2001).
- [9] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. Lett. 85, 5468 (2000).
- [10] R. Cohen, D. ben-Avraham, and S. Havlin, Phys. Rev. E 66, 036113 (2002).
- [11] R. Otter, Ann. Math. Statist. **20**, 206 (1949); T.E. Harris, *The Theory of Branching Processes* (Dover, New York, 1989).
- [12] D. Stauffer and A. Aharony, *Introduction to percolation theory* (Taylor & Francis, London, 1992).
- [13] M. Molloy and B. Reed, Random Struct. Algorithms 6, 161 (1995).
- [14] R. Cohen, S. Havlin, and D. ben-Avraham, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H.G. Shuster (Willey-VCH, New York, 2002), Chap. 4.
- [15] W. Aiello, F. Chung, and L. Lu, Exp. Math. 10, 53 (2001).