Random sequential renormalization and agglomerative percolation in networks: Application to Erdös-Rényi and scale-free graphs

Golnoosh Bizhani, Peter Grassberger, and Maya Paczuski
Complexity Science Group, University of Calgary, Calgary, Canada T2N 1N4
(Received 23 September 2011; published 15 December 2011)

We study the statistical behavior under random sequential renormalization (RSR) of several network models including Erdös-Rényi (ER) graphs, scale-free networks, and an annealed model related to ER graphs. In RSR the network is locally coarse grained by choosing at each renormalization step a node at random and joining it to all its neighbors. Compared to previous (quasi-)parallel renormalization methods [Song *et al.*, Nature (London) **433**, 392 (2005)], RSR allows a more fine-grained analysis of the renormalization group (RG) flow and unravels new features that were not discussed in the previous analyses. In particular, we find that all networks exhibit a second-order transition in their RG flow. This phase transition is associated with the emergence of a giant hub and can be viewed as a new variant of percolation, called agglomerative percolation. We claim that this transition exists also in previous graph renormalization schemes and explains some of the scaling behavior seen there. For critical trees it happens as $N/N_0 \to 0$ in the limit of large systems (where N_0 is the initial size of the graph and N its size at a given RSR step). In contrast, it happens at finite N/N_0 in sparse ER graphs and in the annealed model, while it happens for $N/N_0 \to 1$ on scale-free networks. Critical exponents seem to depend on the type of the graph but not on the average degree and obey usual scaling relations for percolation phenomena. For the annealed model they agree with the exponents obtained from a mean-field theory. At late times, the networks exhibit a starlike structure in agreement with the results of Radicchi *et al.* [Phys. Rev. Lett. **101**, 148701 (2008)]. While degree distributions are of main interest when regarding the scheme as network renormalization, mass distributions (which are more relevant when considering "supernodes" as clusters) are much easier to study using the fast Newman-Ziff algorithm for percolation, allowing us to obtain very high statistics.

DOI: 10.1103/PhysRevE.84.066111 PACS number(s): 89.75.Da, 64.60.aq, 64.60.ah, 05.10.Cc

I. INTRODUCTION

Complex networks provide a useful representation for complex phenomena in a variety of settings including social, biological, and technological systems and have been studied extensively in the past decade [1–3]. A common property of many complex real world networks is the heterogeneity of nodes leading to wide (power-law, "scale-free") degree distributions [4].

For systems embedded in Euclidean space, scale-free statistics is often related to the notion of self-similarity. In statistical physics and critical phenomena this is usually studied by using the renormalization group (RG) technique, where degrees of freedom of the system are eliminated successively by coarse graining. The scaling behavior of the systems close to the fixed point of the RG flow is then examined and systems with similar scaling behavior are classified into universality classes [5,6].

While renormalization is well defined and extensively studied for spatially extended systems (including regular lattices and disordered systems), it is not clear whether it can be applied to complex networks that have no spatial structure, where the topology is given only by the network itself. Naively one would expect that the "small-world" property displayed by many real networks [7,8] means that they cannot be embedded in any finite dimensional space, and thus renormalization schemes should be less useful. Nevertheless, a real space renormalization transformation for such networks was introduced by Song *et al.* [9,10]. In this scheme, the entire network is covered in each RG step by a set of boxes, and each box is considered as a "supernode" in the next RG step. Several complex networks were claimed to have a finite self-similar or fractal dimension; that is, the number of boxes needed to cover the network seemed to show a power-law relation with the diameter of the box, in blatant contradiction to their small-world property. Although this issue was never solved, it was suggested that the fractality of real world networks depends on self-organization in the growth mechanism [11], assortativity of fractal networks [12], and fractality of their underlying structure [13–17].

This conflict between the "small-world" property and any fractality of complex networks was avoided by Radicchi *et al.* [18,19] by using an RG analysis based on the same box covering idea, but studying carefully the RG flow itself, without using any length scale dependence for making claims about fractality.

There are some technical concerns in these previous box covering methods for renormalizing networks. First, according to the original idea of Hausdorff [20], the sizes of boxes should be individually optimized, whereas in the suggested methods all boxes are of equal size. This is a particularly severe problem due to the heterogeneous connectivity in complex networks that leads also to very wide distributions of nodes per box, most of them being nearly empty. Second, even when boxes of the same size are used, the precise placement of boxes strongly affects the result, and optimizing their positions is not practically feasible. Although the suggested methods in Refs. [9–11,13] are claimed to overcome this problem, their results still depend on the order in which the boxes are laid down, making these schemes quasisequential. In particular, the number of nodes per box decreases strongly with the number of boxes already put down. Finally, during each RG step the size of the network decreases dramatically, which results in a

small number of data points in the RG flow. For networks with small-world property this is particularly serious, as the diameter of the networks scales only with $\log(N)$ (N being the size of the network). To compensate for this, only parts of the network have been coarse grained in Ref. [17] at each step of renormalization, which adds more complexity to the process and makes the results even more difficult to interpret.

In our previous work [21], we suggested a completely sequential renormalization scheme for undirected and unweighted graphs called random sequential renormalization (RSR). In RSR at each step of renormalization one node is chosen randomly, and all nodes within a given distance b are replaced by a single supernode. All links from the outside to the (removed) neighborhood are redirected to the supernode, and the supernode is then treated like any other node in the network. The parameter b is called the box radius.

RSR has the advantage that it does not involve any optimum tiling and is very easy to code and understand. It avoids the problem of mostly empty boxes. Furthermore, as the network is affected only locally and the decimation is considerably less at each step of RSR, the whole flow generates much more statistics, which allows a more detailed analysis.

Another advantage of RSR is that it can be interpreted as a cluster growth process, where initially all nodes are considered to be clusters of mass one. At each step of RSR a randomly chosen cluster grows by agglomerating with all its neighboring clusters. Using the fast Monte Carlo algorithm for percolation introduced by Newman and Ziff (NZ) [22,23], RSR can be easily implemented on networks with millions of nodes.

In our first paper on RSR [21], we applied this method to critical trees. Their simple structure makes it possible to study the renormalization flow analytically, giving perfect agreement with results from numerical simulations. We found three regimes in the evolution of critical trees under RSR. (i) First is an initial regime with small fluctuations in the region $N_0^{1/2} \lesssim N < N_0$ (with N_0 being the initial size of the network and N its size at a given renormalization step). (ii) The second is an intermediate regime for $N_0^{1/4} \lesssim N \lesssim N_0^{1/2}$ where the network is a fat, short tree whose structure is dominated by a giant hub. The transition between these two regimes is associated with emergence of a power-law degree distribution and is described by crossover functions exhibiting finite-size scaling. (iii) A third regime extends down to $N = 1$ where the network is a star with a central hub and many leaves.

The appearance of power-law distributions and scaling is associated with a continuous transition, called "agglomerative percolation" (AP) [24]. In one dimension (i.e., graphs consisting of a simple 1D chain), AP has been solved exactly [25,26]. There it shows nontrivial scaling with exponents that depend on the box size [25]. In two dimensions, AP is for triangular lattices in the same universality class as ordinary percolation (OP), whereas it shows different critical behavior for square lattices [24]. This is related to the bipartite structure of the square lattice as every site on the boundary of any cluster is on the same sublattice [27]. The fact that patently nonfractal structures like 1D and 2D lattices also exhibit scaling under RSR suggests that some of the scaling laws previously found in small-world networks are due to AP transition, rather than any underlying fractality of most networks.

In the present paper we study the behavior of sparse Erdös-Rényi (ER) graphs and of the scale-free model of Barabási and Albert (BA) [4] under RSR. For sparse ER graphs under RSR with $b = 1$, we find a continuous percolation transition at finite $x = N/N_0$. Using finite size scaling methods, we show that the corresponding critical exponents are consistent with a scaling theory based on two independent exponents. Within our error estimates, these exponents appear to be independent of the initial average degree of the ER graphs. For the BA model the transition seems to be pushed to $x = 1$, which makes it more difficult to obtain precise numerical results.

We also study RSR analytically using a mean-field theory based on generating functions. The behavior of graphs before the AP transition is consistent with this theory. After the transition the theory fails due to large fluctuations, as well as due to the effect of loops that are negligible before the transition. The predictions of the theory are in agreement with our simulations of an annealed model.

We introduce our model and simulation method in Sec. II, where we also define the graphs and the sizes of the ensembles under study. Section III presents our simulation results for ER graphs. We show evidence of a continuous percolation transition, find the scaling properties and the corresponding critical exponents numerically, and show that they obey the common scaling relations of OP. In Sec. IV we develop a mean-field theory for the evolution of ER graphs under RSR and compare its results with our simulations of an annealed model. We discuss the behavior of graphs beyond the percolation transition in Sec. V. Sections VI and VII are devoted to the results of simulations on ER graphs with different average degrees as well as RSR with larger box sizes. Finally, in Sec. VIII we examine BA networks, and we conclude our study in Sec. IX.

II. THE MODEL

A. Random sequential renormalization

Random sequential renormalization with radius $b (b = 1,2,\ldots)$ is the process of consecutively applying a local coarse graining (one step of RSR), on a given network G_0, which leads to a series of graphs $G_t (0 \leqslant t \leqslant T)$ with strictly decreasing sizes N_t. In the following, t is called *time*. The initial graph at $t = 0$ has N_0 nodes, and the final graph at $t = T$ is just a single node. We also assign masses to the nodes (initially $m_i = 1; \forall\, i \in G_0$). For any undirected and unweighted graph with N nodes of masses m_i, one step of RSR (as shown in Fig. 1) proceeds as follows.

(i) Choose randomly one of the nodes in the graph as the target.

(ii) Define the neighborhood $\mathcal{N}$ around the target to include all nodes within a distance $d \leqslant b$ from the target. Distance is measured by the number of links that make the shortest path between two nodes.

(iii) Delete all the nodes in $\mathcal{N}$, except for the target.

(iv) Delete all the internal links of $\mathcal{N}$.

(v) Redirect to the target all links that connect nodes in $\mathcal{N}$ to the rest of the network.

(vi) If a multiple link appears, replace it with a single link.

(vii) Update the mass of the target to $m = \sum_i m_i$, with $i \in \mathcal{N}$.

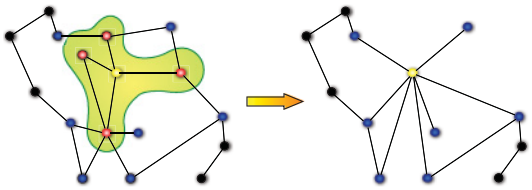

FIG. 1. (Color online) One step of RSR with radius $b = 1$. The randomly chosen target node [yellow (white)] absorbs all its nearest neighbors [red (light gray)]. All links to the absorbed nodes [blue (dark gray)] are then redirected to the target. Alternatively, one can view the supernode as a cluster that grows by eating all its neighboring clusters. RSR with any $b > 1$ can be performed by applying the above procedure on the same target b times.

Hence, the target node and all its neighbors up to distance b in the network are replaced by a supernode. This process preserves all the links to the outside but discards the internal details in the target's neighborhood, analog to course graining in real space renormalization. The supernode is then treated like any other node in the network. We consecutively repeat this procedure until the graph is reduced to a single node. Alternatively, one can also define RSR such that the target node is chosen with probability proportional to its mass [24,26] or degree, but we only discuss the unweighted form here.

For $b = 1$ the target absorbs only its nearest neighbors. The easiest way to implement RSR with any $b > 1$ is to apply RSR with $b = 1$ on the same target for b successive steps. Although this is slightly slower than an optimal coding, we use it in our simulations to reduce code complexity and potential sources of errors.

As indicated in Fig. 1 RSR can also be interpreted as a cluster growth process on the graph. The target cluster is chosen at random and grows by absorbing all clusters within distance b of it. Hence, the fast NZ algorithm for growth of percolation clusters can be easily adapted to this problem, and it makes sense to speak of a percolation transition beyond which one of the clusters occupies a finite fraction of the nodes.

B. The graph ensembles under discussion

We mainly focus on connected ER graphs with average degree $\langle k\rangle$ slightly larger than 2. The ensemble is produced in the following way: For each graph size N_0 we make several ER graphs with fixed size $N^* > N_0$ and a fixed number of links such that the average degree $\langle k\rangle^* = 2$ and determine their giant component (which contains about 80% of the nodes for this value of $\langle k\rangle^*$). If the size of the giant component is $N_0 \pm 1\%$—corresponding to $N_0 = (0.80 \pm 0.01)N^*$—we add the giant component to the ensemble; otherwise it is discarded. Notice that this leads to a slight scatter of N_0 and of the average degree of the graphs at the start of RSR. The latter is $\approx\langle k\rangle_0 = 2.4$. For each N^* the ensembles typically contain $\approx 10^4$ networks, and we apply several realizations of RSR on each of them.

We also examine RSR on ER graphs with $\langle k\rangle^* \neq 2$, as well as the scale-free model of Barabási and Albert [4]. In each case, the ensemble of connected graphs is generated in a similar manner.

C. Algorithms and quantities of interest

In network studies much attention has focused on the statistics of the number of links (degree) of nodes in a network. The degree distribution, the maximum degree, the average, and higher moments of the distribution are also often considered. However, keeping track of the degrees of all nodes under RSR is time consuming and seriously confines the system sizes and statistics of numerical studies. In this paper we have performed numerical simulations with degree measurements for networks up to $N^* = 2.4 \times 10^5$ nodes.

As mentioned previously, the NZ algorithm can be adapted to keep track of cluster masses rather than their degrees. With the NZ algorithm large network sizes with high statistics can be simulated in a reasonable time. We have performed RSR with mass analyses on networks up to $N^* = 10^7$ nodes. Unfortunately, all our efforts to track the degrees of the nodes using the NZ or other algorithms have led to extremely long running times; thus, we restrict our analysis to degrees of smaller networks and measure only masses for larger ones. As far as critical behavior is concerned, we show that mass and degree distributions lead to similar conclusions.

D. Averaging over the ensemble

When discussing ensemble averages, one can use different quantities as independent control parameters. In particular, one can average over RSR trajectories at fixed N or at fixed t. As shown in Fig. 2, these two ways of averaging give different results at late times (and hence small N), due to large fluctuations in the number of nodes eliminated per RSR step in the hub dominant phase. In the same figure we also show the result of a mean-field theory (MFT) discussed in Sec. IV.

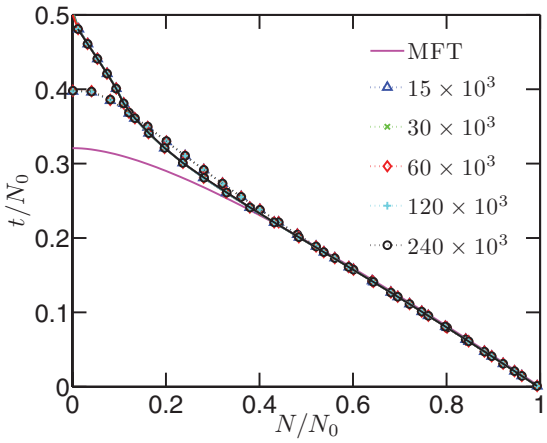

FIG. 2. (Color online) Time dependence of network size, N, in rescaled units. The size decreases monotonically under RSR. Data is obtained from ER graphs with $\langle k\rangle^* = 2$ under RSR with $b = 1$. The curves with dashed lines are obtained by averaging t values corresponding to fixed N, and the curves with solid lines are obtained by averaging N for fixed t. The magenta (gray) solid line shows the mean-field theory prediction (see Sec. IV). The two averages differ when mean-field theory breaks down due to fluctuations. In the rest of the paper we choose N as the independent variable and average all other quantities at fixed N. Numbers in the legend show the initial size of the ER graph, N^*, from which the initial giant components are obtained.

During the initial stages of the flow, MFT gives an accurate description of RSR, but breaks down when different ensembles lead to different results. Some RSR flows last a much longer time than others and since we want to keep the number of members in the ensemble more or less fixed to obtain each data point, we choose to average at fixed N (rather than t) in the rest of this paper.

III. SIMULATION RESULTS FOR ER GRAPHS WITH INITIAL $\langle k\rangle^*=2$

We focus in detail on the behavior of the giant component of ER graphs with $\langle k\rangle^* = 2$ under RSR with $b = 1$. For these graphs the average degree of the giant component is $\langle k\rangle_0 = 2.4$. We find evidence for a continuous AP transition in the evolution of these networks under RSR. The transition is associated with the emergence of a giant hub or the percolation of a giant cluster on the network. We study scaling properties at this transition and measure the corresponding critical exponents numerically. We show that these exponents obey scaling relations associated with percolation, although RSR represents a different universality class than OP, even in the mean-field limit.

A. Evidence for a phase transition

We begin by studying the behavior of the maximum degree $k_{\max}$, as a function of N/N_0, for various initial system sizes, as shown in Fig. 3. Note that the direction of the renormalization flow—or *time*—is from right to left. The initial ER graph has a narrow Poisson degree distribution with no hubs and $k_{\max}/N_0$ is $\mathcal{O}(1/N_0)$. As RSR aggregates nodes locally, although higher degree nodes appear in the system, $k_{\max}/N_0$ remains small. However, as shown in Fig. 3, $k_{\max}/N_0$, suddenly at $N/N_0 \sim 0.7$, starts to increase more rapidly. This implies the existence of at least two regimes in the evolution of ER graphs under RSR: first, a *no-hub* (or mean-field; see Sec. IV) regime, where the degree distribution is narrow, fluctuations are negligible, and a mean-field theory describes the evolution of the system; second, a *hub* regime where a growing hub exists and our mean-field theory breaks down. This is due to large fluctuations as well as the effect of loops in the network. Loops are present in the networks initially, but they typically are large and the graphs are locally treelike. As RSR proceeds, these loops become shorter and the graphs no longer remain locally treelike. As indicated in Fig. 3, the transition between these two regimes becomes sharper on increasing system size N_0.

The same behavior can be observed for the mass of the larger cluster, $M_{\max}$, as shown in the inset of Fig. 3. Initially $m = 1$ for all nodes. Although clusters grow under the renormalization flow in the mean-field regime, the maximum mass remains $\mathcal{O}(1)$. In the critical region a node with the largest mass percolates and separates itself from the rest of the distribution in terms of both size and degree.

This is also indicated in Fig. 4, where both $k_{\max}$ and the second largest degree $k_{\max,2}$ are plotted vs N/N_0. While the two largest degrees are about the same size in the mean-field regime, after the transition the largest hub grows and the second largest degree shrinks, which is another indication of a percolation transition. Similar behavior for $M_{\max}$ and the second largest mass, $M_{\max,2}$, is shown in the inset.

The detailed relation between mass and degree is discussed in the Appendix. No singular behavior in $k_{\max}$ vs $M_{\max}$ appears in the critical region, and this smoothness holds statistically for the mass and degree of other nodes as well. Thus, either variable can be used to extract the critical properties of the phase transition. Since RSR with mass measurement is much

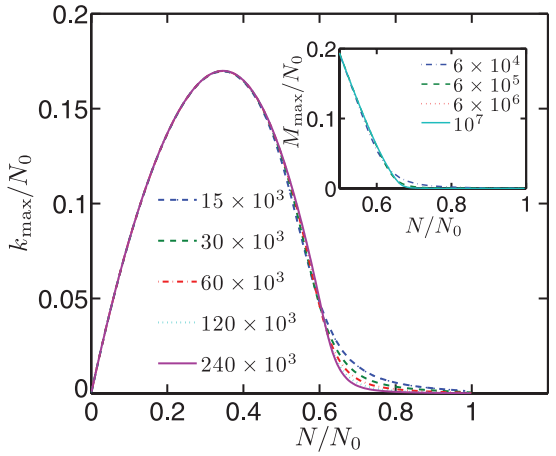

FIG. 3. (Color online) Plot of $k_{\max}/N_0$ vs N/N_0 for ER graphs with $\langle k\rangle^* = 2$ and several initial sizes. Note that the direction of the RSR flow is from right to left. While $k_{\max}/N_0$ is close to zero in the mean-field regime, the hub at late times absorbs a finite and increasing fraction of the nodes. The transition gets sharper with increased system size. (Inset) Similar behavior for the rescaled maximal cluster mass $M_{\max}/N_0$. Note that $M_{\max}$ always increases monotonically under RSR, whereas $k_{\max}$ has to finally decrease. Using the Newman-Ziff algorithm mass related properties can be measured on much larger systems than degree related properties.

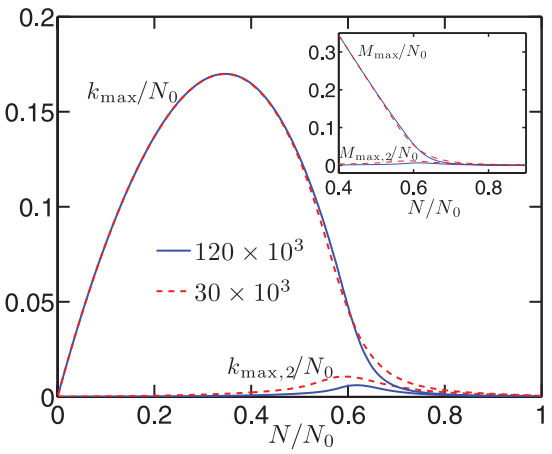

FIG. 4. (Color online) The largest degree $k_{\max}$ and the second largest degree, $k_{\max,2}$, are of comparable size in the mean-field regime, but in the hub regime a giant hub takes over and the second largest degree shrinks. This behavior is also consistent with a continuous percolation transition and shows that there is only one outstanding hub (cluster) in every network. The inset shows the same behavior for the largest and the second largest mass. The data are obtained from ER graphs with $\langle k\rangle^* = 2$.

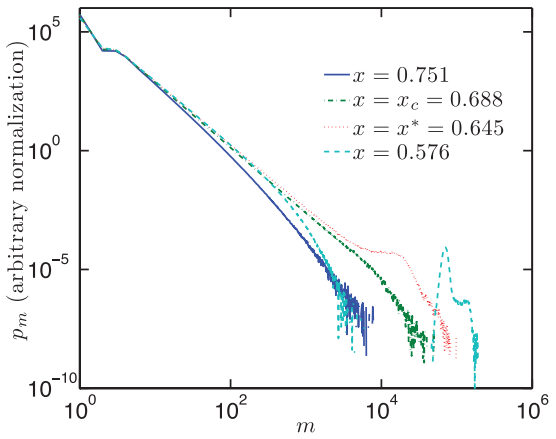

FIG. 5. (Color online) Cluster mass distribution at different stages of the RSR flow for ER graphs of $N^* = 10^6$ nodes, and $\langle k \rangle^* = 2$. This distribution broadens and approaches a power law $p_m \sim m^{-\tau}$ as $x = N/N_0$ decreases. The power law is broadest at $x^*(N_0) = 0.645$, for this system size. For $N_0 \to \infty$, the critical point converges to $x^* \to x_c = 0.688$ (the green dashed-dotted curve). For $x < x^*$ a giant cluster emerges and a gap expands between this cluster and the rest of the distribution. Note that the size distribution of the giant cluster has a shoulder on the right (unlike OP). This is due to the possibility of selecting the hub as a target node and is discussed in more detail in Sec. V.

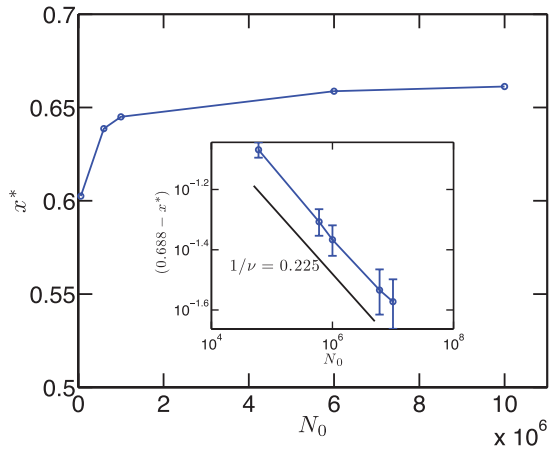

FIG. 6. (Color online) Convergence of the effective critical point, $x^*(N_0)$, to $x_c = 0.688$ as the system size increases. (Inset) The critical point x_c and the exponent $1/\nu$ are consistent with the values $x_c = 0.688$ and $-1/\nu = -0.225$, as indicated by the slope of the straight line.

faster using the NZ algorithm, we mostly base our discussions on the masses of nodes.

B. Finite size scaling analysis

In order to analyze the RG flow in the critical region, we perform a finite size scaling (FSS) analysis on a number of quantities and their distributions. Initially, all the nodes have mass $m = 1$. As shown in Fig. 5, the mass distribution broadens with the number of RSR steps, until a power-law distribution $p_m \sim m^{-\tau}$ emerges in the critical region. As the RSR flow continues, an expanding gap appears between the giant cluster and the rest of the clusters. As shown by the curve for $x \equiv N/N_0 = 0.576$ (below the transition), the peak corresponding to the giant cluster has a pronounced shoulder on the right. This is different from OP, where the peak is featureless. As discussed in more detail in Sec. V, this shoulder results from the giant cluster being chosen repeatedly as the target of RSR. These are rare events, but they have dramatic effects on the flow.

Setting $x = N/N_0$, the effective critical point for a finite system, $x^*(N_0)$, is defined as the value at which the system has the broadest power law in its mass distribution. In Fig. 6, we illustrate the convergence of x^* as the system size increases. The limiting value for infinite system size, x_c, is consistent with $x_c = 0.688$, as shown in the inset.

To proceed further, we make a conventional scaling ansatz for the mass distribution of a finite system in terms of a homogeneous scaling function [6],

$$p_m = m^{-\tau} g\left(n N_0^{1/\nu}, m/N_0^D\right), \tag{1}$$

where

$$n = (x - x_c)/x_c. \tag{2}$$

Note that such an ansatz is never perfect, and all critical parameters discussed in the following are obtained by compromises to get the best overall data collapses for all quantities of interest and by assuming the scaling relations between critical exponents implied by the FSS ansatz. A summary of all critical exponents, the equations defining them and the figures demonstrating numerical evidence, is given in Table I.

Our best estimate for the critical point—mainly from Fig. 6, but also taking into account the consistency checks in Sec. III C—is

$$x_c = 0.688 \pm 0.002. \tag{3}$$

The exponent $1/\nu$ in Eq. (1), describing the convergence of x^* to x_c with increase of system size, is determined to be

$$1/\nu = 0.225 \pm 0.005. \tag{4}$$

The exponent D, giving the scaling of the maximum mass with system size (see Fig. 7), is

$$D = 0.60 \pm 0.01. \tag{5}$$

It is related to the Fisher exponent τ by demanding that there is $O(1)$ cluster of size $\geqslant N_0^D$ and using Eq. (1) [6]:

$$\tau = \frac{1 + D}{D} = 2.67 \pm 0.03. \tag{6}$$

Degree distributions behave similar to the mass distributions. The initial ER graph has a Poisson degree distribution. As RSR proceeds, higher degree nodes appear and the degree distribution broadens. At the phase transition local hubs join together to make a single hub much larger than all others. Just before the giant hub emerges, the degree distribution is approximately a power law with a power τ_k that is consistent with the power τ of the mass distribution. Afterward, the hub continues to grow, but not forever. Figure 8 shows the degree distribution at different values of $x = N/N_0$. At $x \approx 0.7$ the distribution resembles a power law; at $x \approx 0.6$ a bump appears at the rightmost end of the distribution. Later, when $x < 0.6$, a growing gap forms between the giant hub and the rest of the

TABLE I. Summary of critical exponents for ER graphs with $\langle k\rangle^* = 2$ under RSR with $b = 1$. All exponents are obtained by best compromise for the data collapses shown in the figures listed in column 5 and by requiring the scaling relations in column 3 to hold, except for the exponent α (last line). The critical exponents are clearly different from those of mean-field OP (column 6) and for mean-field AP(last column).

Exponent	Value (RSR)	Scaling relation	Eq(s).	Fig(s).	Mean-field OP	Mean-field AP
D	0.60 ± 0.01		(1)	7	2/3	
ν	4.44 ± 0.10		(1)	6, 7	3	4.4 ± 0.3
τ	2.67 ± 0.03	$(1+D)/D$	(1), (6)	7	5/2	3
β	1.78 ± 0.08	$(1-D)\nu$	(7), (8), (9)	9	1	
σ	0.375 ± 0.015	$1/(D\nu)$	(11), (12), (13)	10	1/2	
γ	0.88 ± 0.10	$2D\nu - \nu$	(15), (16), (17)	11	1	1/2
α	6.8 ± 0.3	See text	(18)	12	4	

nodes. This continues until the shrinking system size forces the degree of the giant hub to decrease.

C. Consistency checks

In this section we check for consistency of our simulations with the scaling theory based on the FSS ansatz in Eq. (1) by showing data collapses for different quantities of interest. Notice that the well known scaling relations between critical exponents [6] follow from Eq. (1) by considering appropriate limits.

1. The order parameter

An order parameter is any property of a system that can unravel the singularity at the critical point, which is nonzero only on one side of the transition. Typically, P_∞, the probability that a given site belongs to the percolating cluster, is considered as an order parameter for percolation. For RSR on graphs both $k_{\max}/N_0$ and $M_{\max}/N_0$ can be used as order parameters. Notice that the latter is equal to P_∞.

An FSS ansatz for $M_{\max}$ follows by multiplying Eq. (1) by m^y, integrating over m, and taking the limit $y \to \infty$. Using also Eq. (6) gives

$$\frac{M_{\max}}{N_0^D} = h\left(nN_0^{1/\nu}\right). \tag{7}$$

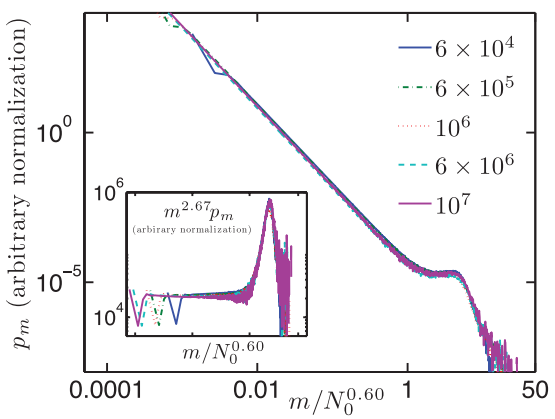

FIG. 7. (Color online) Data collapse using the FSS ansatz in Eq. (1) for the mass distribution at $x^*(N_0)$. The exponent $D = 0.6$ gives the best data collapse, and $\tau = 2.67$ fits the power law (see the inset). These values are consistent with Eq. (6).

Assume now that $h(z)$ satisfies a power law, $h(z) \sim z^\beta$ for $z \to 0$. Equation (7) gives then in the supercritical case $n < 0$ (where we expect $M_{\max} \propto N_0$),

$$D = 1 - \frac{\beta}{\nu} \tag{8}$$

and

$$\frac{M_{\max}}{N_0} \sim |n|^\beta. \tag{9}$$

Figure 9 shows a data collapse according to Eq. (7), with $x_c = 0.688$ and critical exponents as given in Table I. The analogous FSS ansatz for $k_{\max}$,

$$\frac{k_{\max}}{N_0^{1-\beta_k/\nu_k}} = h_k\left(nN_0^{1/\nu_k}\right), \tag{10}$$

with $\nu_k = \nu$ and $\beta_k = \beta$ is shown in the inset of Fig. 9. The exponents for maximal mass and degree are equal within our errors.

2. The cutoff scale for the cluster size distribution

The size of the second large cluster, $M_{\max,2}$ (resp. $k_{\max,2}$), determines the cutoff for the finite clusters (excluding the hub).

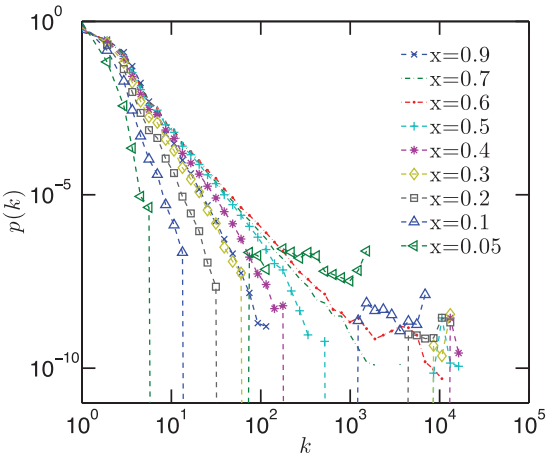

FIG. 8. (Color online) Degree distributions for $N^* = 1.2 \times 10^5$ at different stages of the RSR flow. The initial, narrow distribution gets broader and approaches a power law $p_k = k^{-\tau_k}$ close to the transition. Then a giant hub stands out and a gap opens between the hub and the rest of the nodes.

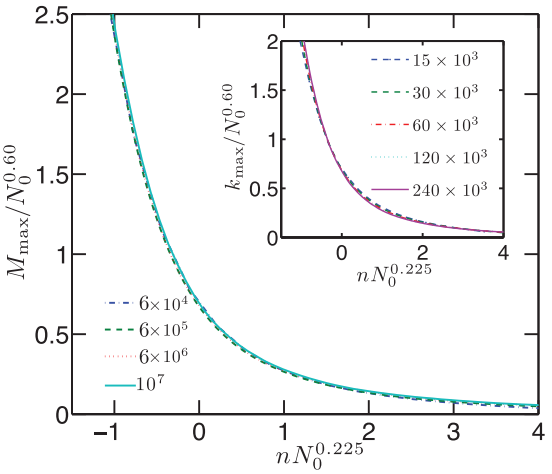

FIG. 9. (Color online) Scaling plot of $M_{\max}$ in the critical region. The value of $x_c = 0.688$ in here and the following figures is the same as in Fig. 6, and the critical exponents are those given in Table I. The inset shows that the exponents ν_k and β_k for the maximum degree are the same as those for the maximum mass within error.

An FSS ansatz based on Eq. (1) gives

$$\frac{M_{\max,2}}{N_0^{1/\sigma\nu}} = h_2\left(nN_0^{1/\nu}\right), \tag{11}$$

and for the infinite system limit

$$M_{\max,2} \sim n^{-1/\sigma}. \tag{12}$$

The exponent σ is related to other exponents by

$$\sigma = \frac{1}{D\nu} = 0.375 \pm 0.015. \tag{13}$$

One can write similar equations for $k_{\max,2}$. Figure 10 shows data collapse plots with $1/\sigma\nu = 1/\sigma_k\nu = 0.57$ for the second largest mass and degree. This leads to

$$\sigma = \sigma_k = 0.395. \tag{14}$$

This estimate was chosen as it gives the best data collapse. It is consistent with the value obtained in Eq. (13), within error.

3. Average cluster size

The average size of the cluster to which a randomly chosen node of the original network belongs is equal to the second moment of the mass distribution. An FSS ansatz for the average cluster size can be written as

$$\frac{\langle m^2\rangle}{N_0^{\gamma/\nu}} = J\left(nN_0^{1/\nu}\right), \tag{15}$$

and in the limit of $N_0 \to \infty$

$$\langle m^2\rangle = n^{-\gamma}. \tag{16}$$

The exponent γ obeys the scaling relation

$$\gamma = (2D-1)\nu = 0.88 \pm 0.10. \tag{17}$$

Figure 11 shows the corresponding FSS analysis, with $\gamma/\nu = 0.17$ chosen for an optimal data collapse. Within errors, this is consistent with the value $\gamma/\nu = 0.20 \pm 0.20$ obtained in Eq. (17). In the inset we show the second moment of the mass

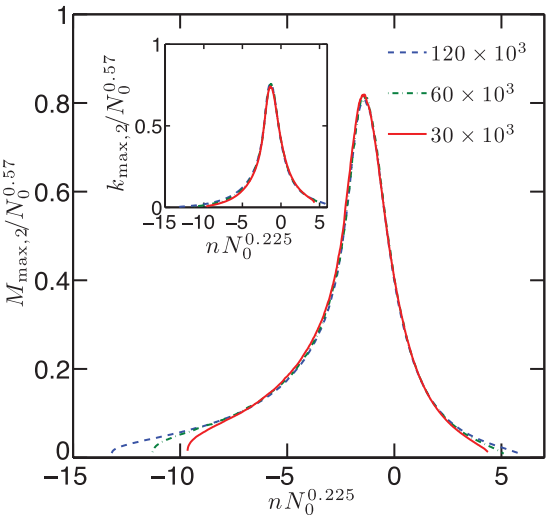

FIG. 10. (Color online) Data collapse for the second largest mass and the second largest degree. Again ν is taken from Table I, while σ is fitted for optimal collapse. The value of σ_k used in the inset is equal to σ, which is here $\sigma = 0.395$.

distribution excluding the largest cluster, $\langle m^2\rangle_{ex}$, with the same scaling exponents. The exponent γ_k for the degree moment is found to be the same as that for the mass moment within error (data not shown).

4. Variance of the maximal cluster size

The variance of $k_{\max}$ and $M_{\max}$ also diverge at the critical point. Because of technical problems we do not have precise values of the latter, and we concentrate on the variance of $k_{\max}$. It should scale as

$$\frac{Var[k_{\max}]}{N_0^{\alpha_k/\nu_k}} = J'\left(nN_0^{1/\nu_k}\right). \tag{18}$$

Figure 12 shows the corresponding scaling plot. In OP the standard deviation of the order parameter has the same

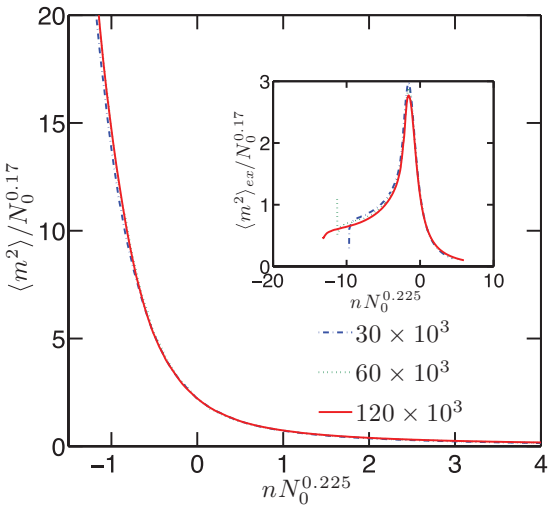

FIG. 11. (Color online) Scaling plot of the second moment of the mass distribution, $\langle m^2\rangle$, for ER graphs, with $\gamma/\nu = 0.17 \pm 0.03$. The inset shows the same plot for $\langle m^2\rangle_{ex}$. The same exponents are obtained.

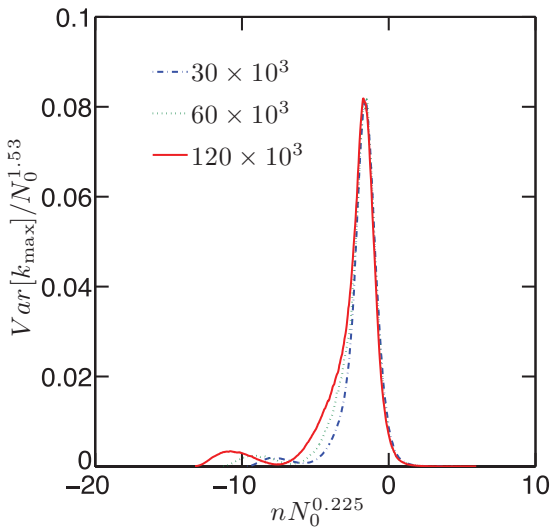

FIG. 12. (Color online) Scaling plot for the variance of the maximum degree, with $x_c = 0.688$, $1/\nu = 0.225$, and $\alpha_k/\nu = 1.53 \pm 0.02$.

critical exponent as the order parameter itself [28], implying $\nu - \beta = \alpha/2$. This is not what we find for RSR on ER graphs, if we assume $\alpha = \alpha_k$. To clarify this we directly looked at the distribution of $k_{\max}$ at x^* for each of the system sizes shown in Fig. 12 (data not shown). The distribution is flat on the left side, but has an approximate power-law tail on the right. The fluctuations in this case grow faster than the average, unlike in OP. We thus believe that the observed violation of the scaling relation is not due to $\alpha \neq \alpha_k$, but shows that the relation $\nu - \beta = \alpha/2$ is violated in AP.

This stems from the difference in the growth process in agglomerative and OP. Adding a bond (or site) in OP might merge only a few clusters into the giant cluster, leading to an additive growth of its size (and degree). In contrast, if the hub is chosen as the RSR target, it absorbs all its neighboring clusters. This leads to multiplicative growth. Thus, in RSR we expect to see larger relative fluctuations in the hub size near the transition comparing to OP (see also Sec. V).

IV. MEAN-FIELD THEORY AND AN ANNEALED MODEL

We now approach the problem analytically using a mean-field theory (MFT) based on generating functions [29]. We show that the critical exponents for mean-field RSR do not agree with the ones for ordinary mean-field percolation.

A. General formalism

Let n_k be the number of nodes with degree k. The total number of nodes in the network is $N = \sum_k n_k$ and the probability of picking a node with degree k is $p_k = n_k/N$. The change of n_k in one step of RSR can be written as the sum of a loss term r_k associated with eliminating a k-degree node and a gain term q_k associated with creating one,

$$\frac{dn_k}{dt} = r_k + q_k. \tag{19}$$

The loss term r_k is

$$r_k = -p_k - \sum_{k'} k' \frac{kp_k}{\sum_l l p_l} p_{k'} = -(k+1)p_k. \tag{20}$$

The first term in the central expression is the probability of targeting a k-degree node, and the second term is the probability that any of the neighbors of the target have degree k. Note that the mean-field assumption is to ignore any potential correlations between the degrees of neighboring nodes.

In order to obtain an equation for dN/dt one does not need to know q_k in detail; one just has to know that exactly one new node is created, whence $\sum_k q_k = 1$. Summing Eq. (19) over k leads then, indeed, to

$$\frac{dN}{dt} = -\langle k \rangle, \tag{21}$$

as expected from the fact that all neighbors of a randomly chosen node are eliminated in one RSR step.

To get q_k, assume that the target has m neighbors with degrees $k_1, k_2, \ldots, k_m$. The new degree of the target will be the number of its second nearest neighbors. If all degrees are uncorrelated and the target's neighbors are not connected among themselves,

$$q_k = \sum_m p_m \sum_{k_1,k_2,\ldots,k_m} \prod_{i=1}^{m} \frac{k_i p_{k_i}}{\langle k \rangle} \delta_{k_1+\cdots+k_m,k+m}. \tag{22}$$

We use generating functions to proceed. The degree distribution is generated by

$$G(x) = \sum_k p_k x^k, \tag{23}$$

and q_k by

$$Q(x) = \sum_k q_k x^k. \tag{24}$$

The degree distribution of the neighbors of the target is proportional to kp_k; thus, their remaining degree is generated by

$$\frac{\sum_k k p_k x^{k-1}}{\sum_k k p_k} = \frac{G'(x)}{\langle k \rangle}. \tag{25}$$

Equation (22) gives then

$$Q(x) = \sum_m p_m \left(\frac{G'(x)}{\langle k \rangle} \right)^m = G\left(\frac{G'(x)}{\langle k \rangle} \right). \tag{26}$$

Using Eqs. (19) through (26) one can write the master equation for the generating function of the degree distribution as

$$\frac{d}{dt} G(x) = \frac{1}{N} \left[G\left(\frac{G'(x)}{\langle k \rangle} \right) + \langle k - 1 \rangle G(x) - x G'(x) \right]. \tag{27}$$

B. The average degree

The moments of the distribution can be obtained from

$$\langle k^m \rangle = \left[\left(x \frac{d}{dx} \right)^m G(x) \right]_{x=1}. \tag{28}$$

One can check that the time derivative of the zeroth moment is zero, that is, normalization is correct. The time derivative of the first moment is given by

$$\frac{d\langle k\rangle}{dt} = \frac{1}{N}[\langle k\rangle^2 - 2\langle k\rangle]. \tag{29}$$

Using Eq. (21) to convert to the derivative with respect to N (for subtleties in this, see [21]) and integrating gives

$$\langle k\rangle = \frac{aN_0}{N} + 2, \tag{30}$$

where $a = \langle k\rangle_0 - 2$ and the subscript zero refers to the initial value.

To test the results of MFT we have simulated RSR for an *annealed random graph model* (AM) in the following way: We start with the degree sequence of the giant component of the ER graphs studied in the previous section, remembering for each of the N_0 nodes its degree, but remove all links. During each RSR step we first pick a random target node and read its degree k. Then we pick k other random nodes $i = 1, \ldots, k$, this time with probabilities proportional to their degrees k_i. Finally, we update the degree of the target to $k' = (\sum_1^k k_i) - k$ and discard the other k nodes.

Figure 13 compares Eq. (30) to the simulation results of the AM and of the model discussed in the last section starting with ER graphs. In all three cases we used $\langle k\rangle^* = 2$. Due to loops in the ER graphs, the average degree of the ER graphs is always less than or equal to that of the AM or MFT. Note that ER graphs are locally treelike and the effect of loops can be ignored initially. Thus, before the transition—in the mean-field regime—there is complete agreement between the results of MFT, the AM, and the ER graphs. However, after the transition, the effect of loops as well as fluctuations (which we discuss later) results in a breakdown of the mean-field assumptions and the average degree of the ER graphs no longer agrees with the other two cases.

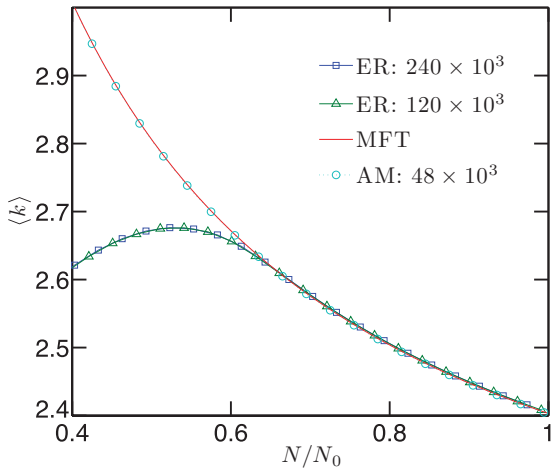

FIG. 13. (Color online) Comparison between the annealed model (AM), mean-field theory (MFT), and ER graphs with $\langle k\rangle^* = 2$. There is good agreement between theory and data in the mean-field regime $x \geqslant x_c$. After the transition the effect of loops in ER graphs can no longer be ignored and results in smaller $\langle k\rangle$ for ER graphs.

In the mean-field regime the system size, N, can be found as an implicit function of t by using Eqs. (21) and (30) to get

$$t = -\frac{1}{2}\left\{N - N_0 - \frac{aN_0}{2}\ln\left[\frac{a + 2N/N_0}{a+2}\right]\right\}. \tag{31}$$

This result is shown in Fig. 2 and is in good agreement with simulation results in the mean-field regime.

C. Divergence of degree fluctuations

Also within MFT, the variance of the degree distribution diverges at the transition point. For ease of calculations we switch to factorial moments of the degree distribution obtained by consecutive derivatives of the generating function,

$$\begin{aligned}\langle k_-^m\rangle &= \langle k(k-1)\cdots(k-m+1)\rangle \\ &= \left[\left(\frac{d}{dx}\right)^m G(x)\right]_{x=1}.\end{aligned} \tag{32}$$

Using Eq. (27), the time derivative of the second factorial moment is

$$\frac{d}{dt}\langle k_-^2\rangle = \frac{\langle k_-^2\rangle}{N}\left[\frac{\langle k_-^2\rangle^2}{\langle k\rangle^2} + \langle k\rangle - 3\right]. \tag{33}$$

We next define a variable $u = \langle k_-^2\rangle/\langle k\rangle$ and use Eqs. (21), (29), and (30) to get

$$\frac{du}{u^3 - u} = -\frac{dN}{aN_0 + 2N} = -\frac{dx}{a + 2x}, \tag{34}$$

with $x = N/N_0$. Integrating this equation leads to

$$u^2 = \frac{a + 2x}{a - c + 2x}, \tag{35}$$

where

$$c = \frac{u_0^2 - 1}{u_0^2}(a+2), \tag{36}$$

and u_0 is the initial value of u. Since the average degree $\langle k\rangle$ does not diverge at the transition, the divergence of u is the same as the divergence of the variance of the degree distribution. The quantity u diverges when the denominator of Eq. (35) vanishes, so the critical point is at

$$x_c = \frac{N_c}{N_0} = \frac{1}{2}\left[c - (\langle k\rangle_0 - 2)\right]. \tag{37}$$

Equations (36) and (37) result in $x_c = 0.718\ldots$ for the AM model with $\langle k\rangle_0 = 2.4$ we study here (notice that the initial degree distribution is not strictly Poissonian due to the restriction to the giant component of the original ER graph). Substituting $n = (x - x_c)/x_c$ into Eq. (35) we get

$$\langle k^2\rangle \sim u \sim n^{-\gamma} \quad \text{with} \quad \gamma = 1/2. \tag{38}$$

For a finite system, we make the FSS ansatz

$$\frac{\langle k^2\rangle}{N_0^{\gamma/\nu}} = f\left(nN_0^{1/\nu}\right). \tag{39}$$

Figure 14 shows an FSS analysis of $\langle k^2\rangle$ for the annealed model, close to criticality, and for several system sizes. The values of $x_c = 0.718$ and $\gamma = 0.112/0.225 = 0.5$ used in

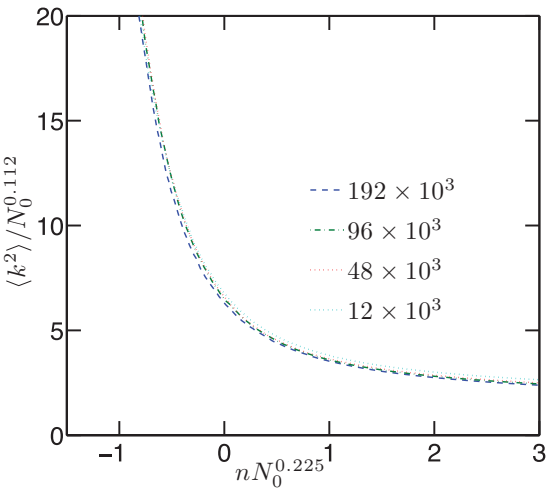

FIG. 14. (Color online) FSS analysis of the second moment of the degree distribution, $\langle k^2\rangle$, close to criticality for the AM. The values of $x_c = 0.718$ and $\gamma = 0.5$ used in the plot are those obtained from the MFT. While the exponent $1/\nu = 0.225 \pm 0.015$, obtained from the FSS data collapse, is similar to ER graphs, the exponent γ is different from $\gamma = 0.88$ in ER graphs.

the FSS analysis are taken from MFT and give excellent agreement.

D. Other critical exponents

To get all other exponents in MFT we use Eqs. (27) and (32) to find the time derivative of the third factorial moment, $h = \langle k_-^3\rangle$,

$$\frac{dh}{dt} = \frac{hu^3}{N} + \cdots, \tag{40}$$

where the dots stand for terms that are less divergent for $x \to x_c$. Together with $du/dt \sim u^3/N + \cdots$ this gives near the critical point

$$\ln h \sim u \sim n^{-1/2} \Rightarrow h \sim e^{1/n^{1/2}}, \tag{41}$$

suggesting that the third moment has an essential singularity. The latter seems to be contradictory to scaling theory, but it really is not, and there exists a consistent solution showing these features. Assume the scaling ansatz

$$p_k = k^{-\tau} f(k/k_{\text{cutoff}}) \tag{42}$$

for the degree distribution near the critical point, with k_{cutoff} diverging at $x \to x_c$. For u to diverge, τ must be $\leqslant 3$. If τ were strictly <3, we would have $u \sim k_{\text{cutoff}}^{3-\tau}$ and $h \sim k_{\text{cutoff}}^{4-\tau}$; that is, there would be a power relation between them: $h \sim u^{(4-\tau)/(3-\tau)}$. The only way to obtain $u \sim \ln h$ is by having a logarithmic divergence of the sum $\sum_k k^2 p_k$, that is,

$$\tau = 3. \tag{43}$$

In order to have $\gamma = 1/2$, one needs furthermore $k_{\text{cutoff}} \sim e^{1/n^{1/2}} \sim N_0$, giving then also Eq. (41). The fact that k_{cutoff} diverges faster than a power for $x \to x_c$ means that there is no simple scaling theory near the transition due to the singularity.

E. Limiting behavior for $\langle k\rangle^* \to 1$

In the limit $\langle k\rangle^* = 1$, the giant components of ER networks become trees with $\langle k\rangle_0 = 2$. Since trees remain trees during RSR, $\langle k\rangle = 2$ during the entire RSR flow, consistent with Eq. (29). On the other hand, $\langle k^2\rangle$ does increase with t. Equation (35) leads to

$$u^2 = \frac{2x}{2x - c}, \tag{44}$$

and Eq. (37) gives

$$x_c = \frac{u_0^2 - 1}{u_0^2} > 0. \tag{45}$$

This is in contrast to the result of [21], where we found $x_c = 0$ for critical trees. Indeed, the limit $\langle k\rangle^* \to 1$ of the present model is *not* the model of critical trees that was treated in [21].

This follows from how the critical trees of [21] and critical ER graphs are generated. In ER graphs links are distributed among nodes completely at random. If a node is picked at random, the degrees of all its neighbors are distributed according to

$$kp_k \Big/ \sum_l l p_l, \tag{46}$$

and there is no further structure. In contrast, the critical trees of [21] are generated by a critical random branching process that starts from one particular node and imprints on them a *rooted* structure. Therefore, if a node is picked randomly, there are relations that hold seperately for its mother and its daughters. While the degree distribution for the mother satisfies Eq. (46) with k replaced by $k-1$, the degree distribution of the daughters is simply p_k. One might think that this subtle difference can be neglected in a mean-field approximation, but this is not true: Since each RSR step affects three generations of nodes, a consistent grandmother-mother-daughter relationship has an effect on the RSR flow. However, it is not intuitively clear why this small difference has such a strong influence on the threshold for AP. Notice that an even more surprising dependence on minor details, leading indeed to a violation of universality, is seen also in AP on 2D lattices [24].

We did not study the case $\langle k\rangle^* = 1$ numerically, because the size of the largest component in critical ER graphs of size N_0 is $\sim N_0^{2/3}$, making it very difficult to create large initial connected graphs.

V. FLUCTUATIONS IN THE HUB PHASE

If the giant cluster (or hub) is itself a target of RSR, the size of the network decreases significantly in that time step. This gives rise to large fluctuations in the size of the network.

Figure 15 shows a scatter plot of t vs N in rescaled units for an ensemble of ER networks with $N^* = 1.2 \times 10^5$ and $\langle k\rangle^* = 2$. The x and y axes in this plot are coarse grained into 500 and 200 bins, respectively, giving 100 000 pixels. The color of each pixel represents the frequency of this (N,t) pair relative to the frequency of the most populated pixel. As one can see, an envelope exists corresponding to the largest N at a given time (and biggest time for a given N). There is also a second band (of high probability) which corresponds to

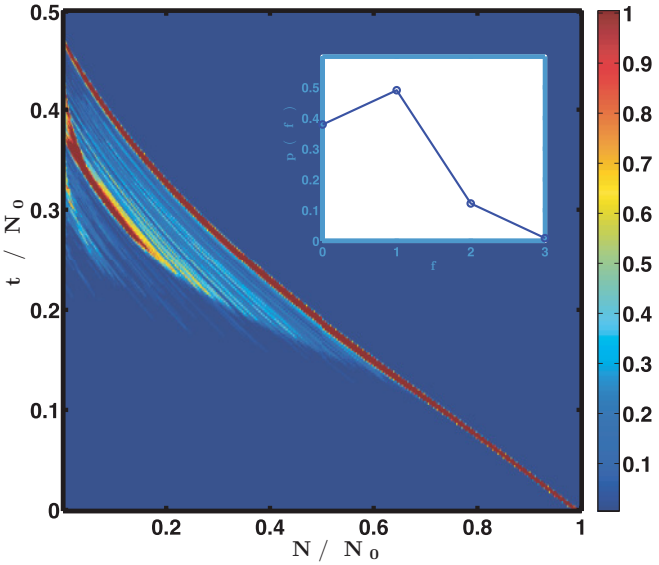

FIG. 15. (Color online) Scatter plot of t vs N in rescaled units for ER graphs with $\langle k\rangle^* = 2$ and $N^* = 120 \times 10^3$. The color map shows the relative frequency of each (N,t) pair in the ensemble. The main, intermediate, and weak bands correspond to realizations where the giant hub has been hit zero, one, or two times, respectively. The inset shows the probability that the giant hub is targeted f times by RSR.

intermediate N. A third but weak band also appears at smaller system sizes, which is more difficult to distinguish due to the considerable fluctuations.

The inset of Fig. 15 shows the fraction of realizations in which the giant hub is hit f times, conditioned on $k_{\max}/N_0 > 0.1$. Most often the hub is hit only once and never more than three times. Evidently, the envelope (the uppermost band) comes from the realizations in which the giant hub was not hit at all, the intermediate band results from cases where the giant hub is hit once, and the third weak band is due to rare cases where the giant hub is hit twice.

The slopes of the main bands are also informative. The uppermost band starts with slope $\frac{-1}{\langle k\rangle_0}$, in agreement with Eq. (21). At the final stages, where the structure is starlike (as discussed in the following section), the bands have slope -1, which means that in most cases a leaf is targeted and thus one node is removed in one time step. The wide range of values for realizations as shown in Fig. 15 explains why in Fig. 2 averaging over t at fixed N gave a different result than averaging over N at fixed t.

The distribution of times T for the networks to reach $N = 1$ corresponds to the leftmost column in Fig. 15. This distribution has a shoulder where the uppermost band hits the y axis, (at $t/N_0 \approx 0.47$) and a peak where the second one hits it ($t/N_0 \approx 0.38$). These distributions show perfect data collapses for different system sizes (data not shown). For networks with $\langle k\rangle^* > 2$ the shoulder turns into a second peak which grows and becomes the dominant peak on increasing $\langle k\rangle^*$. It should disappear for $\langle k\rangle^* \to 1$.

A. Scaling behavior at late times

Eventually, as the networks shrinks, $k_{\max}$ starts to decrease and at the same time the network topology moves toward a starlike structure.

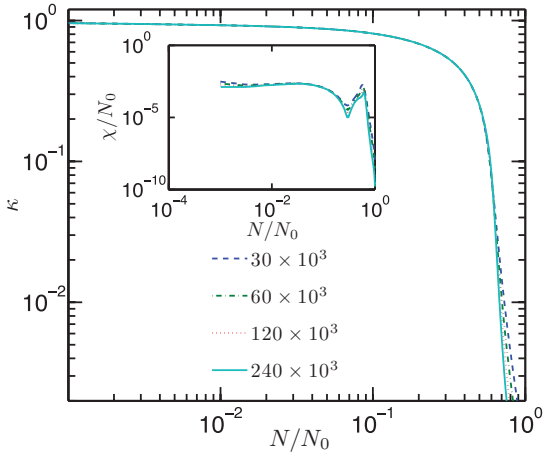

FIG. 16. (Color online) The relative maximum degree, $\kappa = k_{\max}/(N-1)$, vs N/N_0 for different system sizes. The plot shows that the network is starlike at late times since κ approaches 1. (Inset) Variance of κ vs N/N_0. Although the qualitative behavior of graphs under RSR at late times is the same as for the (quasi-)parallel renormalization method [18], the quoted exponents are different.

The relative size of the largest hub $\kappa = k_{\max}/(N-1)$ is a good measure for the similarity of a graph to a star which is a graph whose nodes are at most a distance two apart. Figure 16 shows κ and its variance χ as a function of the relative system size. As one can see, at late stages of RSR κ is close to one, and thus the network has a starlike structure.

The starlike regime was also observed in previous (quasi-) parallel methods used for renormalizing networks [18,19]. Comparing our analysis with those studies, RSR shows scaling and criticality in the flow at early times that was not picked up previously, because the renormalization steps in the quasiparallel method were too large and jumped over the AP transition. Thus, only the scaling at late times was observed in [18,19]. Although the graphs under RSR look qualitatively similar to those obtained with the quasiparallel method at late times, the quoted exponents are different (our Fig. 16 should be compared with Fig. 1 in Ref. [18]).

B. The star regime

We define N_ℓ to be the last size of the network one step before it collapses into a single node. By definition the network has to be a pure star at this point. Figure 17 shows a data collapse for the distribution of N_ℓ for ER graphs of different sizes. It is a broad distribution following the scaling ansatz

$$p(N_\ell) \sim \frac{1}{N_\ell^{\tau_s}} f\left(\frac{N}{N_0^{D_s}}\right), \tag{47}$$

with $\tau_s = 1.40 \pm 0.15$ and $D_s = 0.25 \pm 0.05$.

The exponents τ_s and D_s are similar to the ones obtained for critical trees [21]. This suggests universality in the final structure of the graphs, regardless of the starting structure, as the graph collapses into a single node and all original structure is lost.

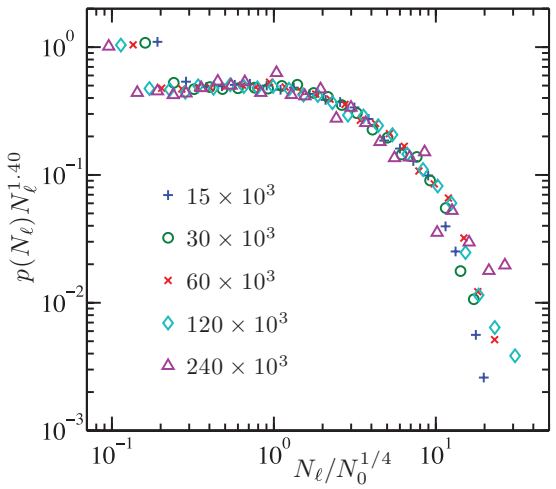

FIG. 17. (Color online) Data collapse for the distribution of the last sizes. The distribution follows the FSS ansatz in Eq. (47) with $\tau_s = 1.40 \pm 0.15$ and $D_s = 0.25 \pm 0.05$, except for the leftmost points. The reason for their special behavior is given in [21].

VI. STARTING WITH OTHER AVERAGE DEGREES

Up to now we studied the behavior of ER graphs with $\langle k \rangle^* = 2$. Here we discuss the effect of the initial average degree on RSR flow, still considering ER graphs.

Figure 18 represents $k_{\max}$ for ER graphs with different values of $\langle k \rangle^*$. The figure demonstrates similar critical behavior for these networks. For higher initial average degree the transition gets sharper and moves to the right, that is, toward earlier times. This is also predicted from Eq. (37). For larger $\langle k \rangle^*$, x_c approaches 1. Note that both $\langle k \rangle_0$ and $\langle k^2 \rangle_0$ affect the position of x_c.

Figure 19 shows an FSS analysis of $M_{\max}$ for ER graphs with $\langle k \rangle^* = 4$. The critical point $x_c = 0.865 \pm 0.010$ and the exponents $1/\nu = 0.215 \pm 0.030$ and $1 - \beta/\nu = 0.62 \pm 0.05$ are obtained by finding the best data collapse. The value of x_c is in agreement with Eq. (37) and the exponent ν and β agree with those for $\langle k \rangle^* = 2$, within our error estimates. For even higher average degrees (not shown) the exponents still agree with the ones obtained for $\langle k \rangle^* = 2$, although the error bars are rather large.

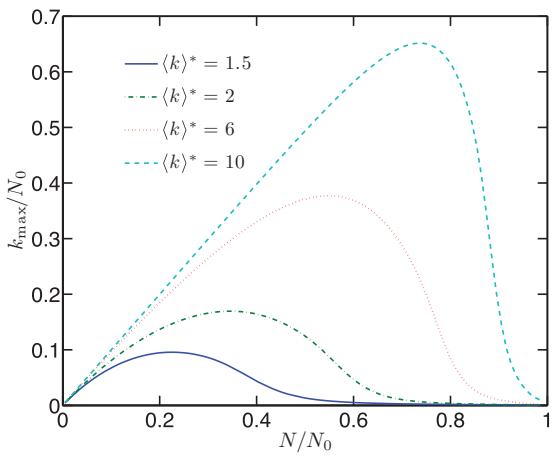

FIG. 18. (Color online) Rescaled maximum degree, $k_{\max}/N_0$ vs system size for ER graphs with $N^* = 30\,000$. The transition shifts to the right with increase of $\langle k \rangle^*$.

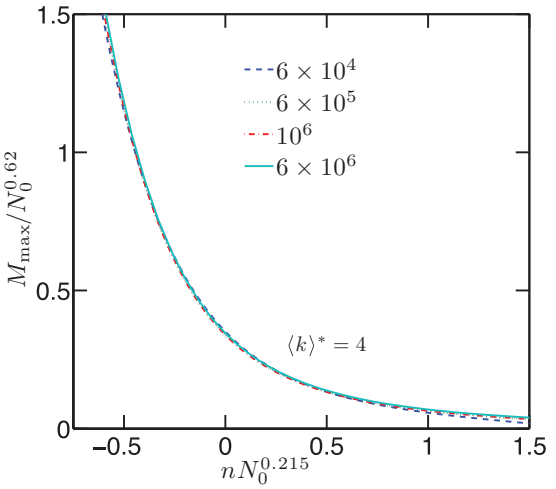

FIG. 19. (Color online) Scaling of $M_{\max}$ for ER graphs with $\langle k \rangle^* = 4$. The values $x_c = 0.865 \pm 0.010$, $1/\nu = 0.215 \pm 0.030$, and $1 - \beta/\nu = 0.62 \pm 0.05$, obtained by finding the best data collapse, agree with Eq. (37) and the exponents for ER graphs with $\langle k \rangle^* = 2$ within error bars.

VII. RSR WITH LARGER BOX SIZES

In this section we study RSR with box radius $b > 1$ on ER graphs with $\langle k \rangle^* = 2$. We see evidence for a transition at early times. Figure 20 shows the order parameter as a function of N/N_0 for networks of different system sizes under RSR with $b = 2$. Although one can clearly see evidence for a phase transition at early times, extrapolating the critical point for the infinite system with precision is not possible.

One of the main differences between the $b = 1$ case and $b > 1$ is that for larger box radii there is no star regime. Once the network has diameter two, it will die in the next step with probability one.

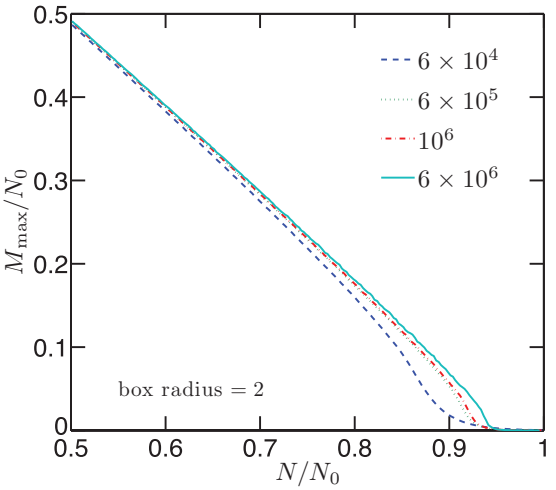

FIG. 20. (Color online) Flow of the order parameter under RSR with box radius $b = 2$ for ER graphs of $\langle k \rangle^* = 2$ and several sizes. The data show a sharp transition at early times, but a clean FSS analysis including precisely locating the critical point is not numerically tractable.

Another point to mention is that with any box size larger than one, the possibility to incorporate the hub at any step is large. The reason is that RSR with $b > 1$ is performed by targeting the same node b successive times. Although the target itself is not likely to be the hub, it is likely that it is the neighbor of the hub and thus merges with it. Hitting the same node again means then hitting the hub with high probability. With this argument any box size higher than one is similar to a *weighted* RSR, where nodes are being targeted with probability proportional to their mass or degree.

VIII. SCALE-FREE NETWORKS

Models with broad or "scale-free" degree distributions are often more interesting in view of their application to real-world networks. We have studied RSR on the Barabasi-Albert model [4]. Figure 21 shows the behavior of the maximum degree under RSR. Since these networks are scale free at the beginning, the transition is pushed all the way to $N/N_0 = 1$. There is perfect data collapse after the hubs are well established.

The critical point can also be obtained from Eq. (37). The value of $\langle k^2\rangle_0$—and thus also the value of $u_0 = \langle k^2_-\rangle_0/\langle k\rangle_0$—diverge for scale-free networks, giving $x_c \approx 1 - 1/u_0^2 \approx 1$.

When analyzing mass distributions for renormalized scale-free networks, it can be argued that one should not give masses $m = 1$ to all nodes of the initial graph. Instead, one might assign to every node a mass equal to its degree, as this allows one to consider mass as a proxy for the degree of nodes in the simulation of the RG flow.

This convention is used in Fig. 22, which illustrates an FSS analysis for the maximum cluster mass, $M_{\max}$, and the second moment of the mass distribution excluding the largest cluster, $\langle m^2\rangle_{ex}$, in BA networks of several sizes. Setting the critical point at $x_c = 1$, we obtained $1/\nu = 0.18 \pm 0.02$, $D = 0.5 \pm 0.1$, and $\gamma/\nu = 0.25 \pm 0.03$.

IX. CONCLUSION

In this paper we have extended random sequential renormalization (RSR) to several networks, namely Erdös-Rényi

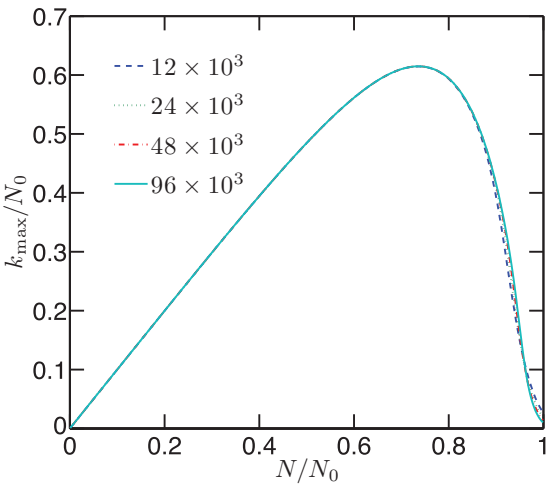

FIG. 21. (Color online) FSS analysis of $k_{\max}/N_0$ for the BA model. The critical point is pushed toward one, and there is a perfect data collapse after the hubs are well established.

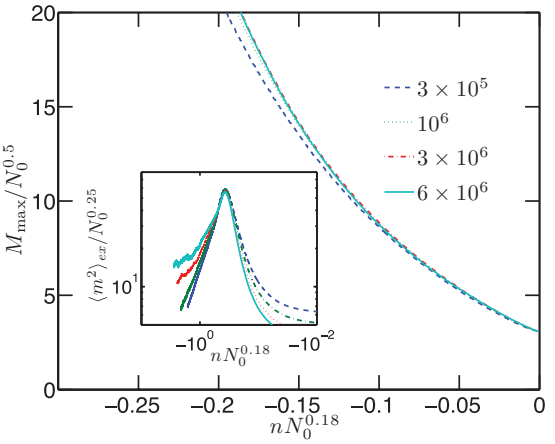

FIG. 22. (Color online) FSS analysis of $M_{\max}$ and $\langle m^2\rangle_{ex}$ for the BA model. The critical point is set at $x_c = 1$. The exponents $1/\nu = 0.18 \pm 0.02$, $D = 0.5 \pm 0.1$, and $\gamma/\nu = 0.25 \pm 0.03$ are obtained from the data collapse.

(ER) graphs and scale-free networks. In each step of RSR only a local part of the network within a fixed distance from a randomly chosen node is coarse-grained into one node. This is in contrast to (quasi-)parallel RG schemes that tile and coarse grain the whole network in one step, which, however, has to be broken up into sequential local substeps for technical reasons. Apart from simplicity of the algorithm, RSR generates considerably larger amount of statistics and allows for a more detailed analysis of the renormalization flow. RSR can be interpreted as a cluster growth process where at each step a randomly chosen cluster grows at its boundary by agglomerating to all its neighboring clusters. Hence, the fast Monte Carlo algorithm of Newman and Ziff [22,23] for percolation can be used to simulate RSR on networks of up to millions—or even billions—of nodes.

For all the graphs we studied, RSR leads to a continuous agglomerative percolation transition (AP) where the largest cluster (node) outgrows all others both in terms of its mass and degree. We found three universality classes (critical trees, sparse ER graphs, and mean-field AP) for evolution of networks under RSR. For sparse ER graphs we derived the corresponding critical exponents numerically and found that the exponents obtained by analysis of the masses of the clusters are not different from the ones obtained by analyzing the degrees of the nodes. Since mass analysis can be performed much faster with the help of the NZ algorithm, we suggest that mass analysis may be better suited to extracting scaling properties of large networks. Regardless of the initial average degree of the ER graph, we found the same critical exponents for the percolation transition, within error. At late stages of RSR, graphs experience a regime in which they switch to a star structure for $b = 1$. For both ER graphs and critical trees this regime extends in the range $1 < N < N_0^{1/4}$.

For scale-free networks the transition is forced to $x_c = 1$. Hence, our data collapse methods for finding the critical exponents of scale-free networks are not as neat as for ER graphs, and this makes it hard to decide whether BA and ER networks are in the same universality class.

While the scaling behavior of critical trees under RSR is similar to graph behavior under the (quasi-)parallel

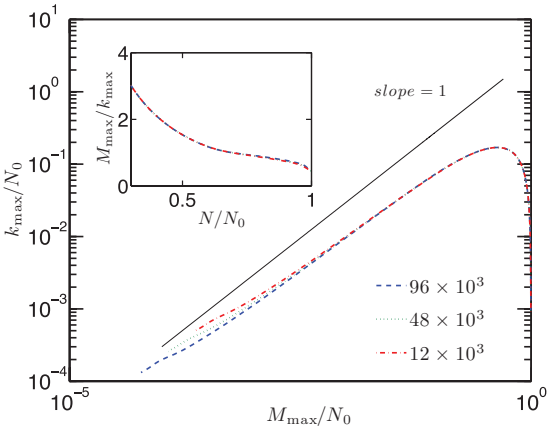

FIG. 23. (Color online) Log-log plot of normalized maximum degree $k_{\max}/N_0$ vs $M_{\max}/N_0$ for ER graphs with $\langle k\rangle^* = 2$ and size $N^* = 1.2 \times 10^5$. These are proportional to each other in a region close to criticality which extends over larger domains with increase of system size. While $M_{\max}$ increases with N_0 monotonically, $k_{\max}$ is confined to the current system size N and starts to decrease deep in the hub phase. The inset shows the ratio of maximal mass and degree on a linear scale. The curve is linear in the critical region.

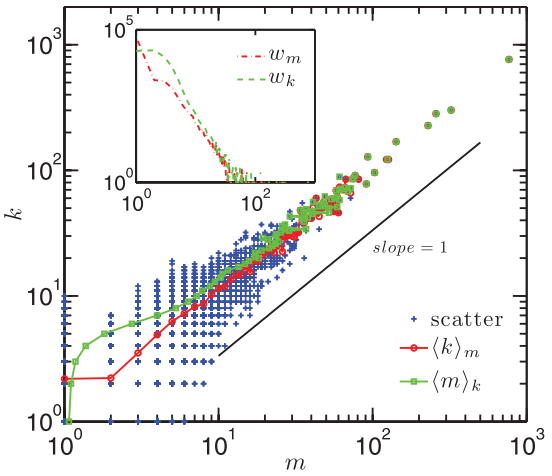

FIG. 24. (Color online) Scatter plot of masses and degrees of all clusters close to criticality. The data are obtained from one RSR trajectory of an ER graph of $N^* = 1.2 \times 10^5$ nodes, with $\langle k\rangle^* = 2$ at $N/N_0 = 0.688$. The red line with circles shows the average degree of clusters of a given mass, and the green line with squares shows the average masses of nodes with a given degree. The inset shows the number of clusters of a given mass (w_m) and the number of nodes of a given degree (w_k).

renormalization scheme studied by Radicchi *et al.* [18,19], the percolation transition revealed by our method in the early stages of the RG flow is not *seen* in their analysis. We conjecture that it exists also there in principle, but it would be very hard to study due to the coarseness of their RG flow observation. At final stages RSR and parallel schemes lead to the same qualitative picture, namely, a starlike structure for $b = 1$, but the scaling behavior and the corresponding exponents are different.

The simplicity of RSR as well as the fact that it is a percolation process both for networks and lattices makes it a useful tool for studying complex networks. For real-world networks finite-size scaling analysis is not generally possible since every network has a fixed (finite) size. However, even in that case high statistics of RSR flow and the efficiency of the algorithm make it possible to study the scaling properties of individual large networks.

ACKNOWLEDGMENTS

We thank Claire Christensen and Seung-Woo Son for numerous discussions and helpful comments.

APPENDIX: RELATIONSHIP BETWEEN MASS AND DEGREE OF NODES

Figure 23 depicts the linear relation between $k_{\max}$ and $M_{\max}$ in the critical region for ER graphs with $\langle k\rangle^* = 2$ and several system sizes. This shows that either of them can be used to extract renormalization flow properties near the transition. Since RSR can be simulated much faster if we only measure the mass-related quantities (instead of degree), we suggest that the RG analysis in the critical region can be confined to mass-related quantities. In the final stages of the flow ($N/N_0 \lesssim 0.3$), $k_{\max}$ decreases as it cannot exceed the number of nodes present in the system. $M_{\max}$, on the other hand, increases monotonically till the end of the process where $M_{\max} = N_0$.

The correlation between mass and degree of clusters close to criticality is shown in Fig. 24 for ER graphs of $N^* = 1.2 \times 10^5$ at $N/N_0 = 0.688$. Each point in the scatter plot shows one (m,k) pair in the whole network. Also shown are the average degree $\langle k\rangle_m$ of nodes of a given mass and the average mass $\langle m\rangle_k$ of clusters with a given degree. In the inset we show the number of clusters with a given mass, w_m, and the number of nodes with a given degree, w_k.

The average mass of nodes with degree one is close to 1, which means that most of them have not been hit by RSR, and the average degree of clusters with mass one is more than two, which means that the nodes that have not been hit by RSR keep their starting average degree. Since the two averages differ only for masses (degrees) less than 10, we suggest that one can use either of them to extract the properties of the percolation transition.

[1] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[2] M. J. E. Newman, SIAM Rev. **45**, 167 (2003).
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep. **424**, 175 (2006).
[4] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).
[5] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971).
[6] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1994).

[7] S. Milgram, Psychol. Today **1**, 61 (1967).
[8] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[9] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[10] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, J. Stat. Mech.: Theory Exp. (2007) P03006.
[11] C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).
[12] S.-H. Yook, F. Radicchi, and H. Meyer-Ortmanns, Phys. Rev. E **72**, 045105(R) (2005).
[13] J. S. Kim, K. I. Goh, G. Salvi, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **75**, 016110 (2007).
[14] J. S. Kim, K. I. Goh, B. Kahng, and D. Kim, New J. Phys. **9**, 177 (2007).
[15] J. S. Kim, K. I. Goh, B. Kahng, and D. Kim, Chaos **17**, 026116 (2007).
[16] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006).
[17] H. D. Rozenfeld, C. Song, and H. A. Makse, Phys. Rev. Lett. **104**, 025701 (2010).
[18] F. Radicchi, J. J. Ramasco, A. Barrat, and S. Fortunato, Phys. Rev. Lett. **101**, 148701 (2008).
[19] F. Radicchi, A. Barrat, S. Fortunato, and J. J. Ramasco, Phys. Rev. E **79**, 026104 (2009).
[20] K. J. Falconer, *The Geometry of Fractal Sets* (Cambridge University Press, Cambridge, 1985).
[21] G. Bizhani, V. Sood, M. Paczuski, and P. Grassberger, Phys. Rev. E **83**, 036110 (2011).
[22] M. E. J. Newman and R. M. Ziff, Phys. Rev. Lett. **85**, 4104 (2000).
[23] M. E. J. Newman and R. M. Ziff, Phys. Rev. E **64**, 016706 (2001).
[24] C. Christensen, G. Bizhani, S.-W. Son, M. Paczuski, and P. Grassberger, to appear in EPL December 2011.
[25] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, Europhys. Lett. **95**, 58007 (2011).
[26] S.-W. Son, C. Christensen, G. Bizhani, P. Grassberger, and M. Paczuski, Phys. Rev. E **84**, 040102(R) (2011).
[27] H. W. Lau, *et al.* (unpublished).
[28] M. Z. Bazant, Phys. Rev. E **62**, 1660 (2000).
[29] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).