Distribution of diameters for Erdős-Rényi random graphs

A. K. Hartmann*
Institut für Physik, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany

M. Mézard†
Département de Physique de l'ENS, PSL University, 75005 Paris, France

(Received 16 October 2017; published 23 March 2018)

We study the distribution of diameters d of Erdős-Rényi random graphs with average connectivity c. The diameter d is the maximum among all the shortest distances between pairs of nodes in a graph and an important quantity for all dynamic processes taking place on graphs. Here we study the distribution $P(d)$ numerically for various values of c, in the nonpercolating and percolating regimes. Using large-deviation techniques, we are able to reach small probabilities like 10^{-100} which allow us to obtain the distribution over basically the full range of the support, for graphs up to $N = 1000$ nodes. For values $c < 1$, our results are in good agreement with analytical results, proving the reliability of our numerical approach. For $c > 1$ the distribution is more complex and no complete analytical results are available. For this parameter range, $P(d)$ exhibits an inflection point, which we found to be related to a structural change of the graphs. For all values of c, we determined the finite-size rate function $\Phi(d/N)$ and were able to extrapolate numerically to $N \to \infty$, indicating that the large-deviation principle holds.

DOI: 10.1103/PhysRevE.97.032128

I. INTRODUCTION

For each connected *component* κ of a network or a graph $G = (V, E)$ [1–3], the *diameter* $d(\kappa)$ is the maximum over all pairs of the component's vertices i, j of the shortest-path distance $i \leftrightarrow j$. The diameter $d(G)$ of the *graph* is the maximum of $d(\kappa)$ over all components κ. The diameter is an important measure of the network. It has a strong influence on, e.g., dynamical processes taking place on these networks, since it characterizes a typical long length scale for the transport of information. Examples where the diameter plays an important role are rumor spreading [4], energy transport in electric grids [5], or oscillations in neural circuits [6]. Furthermore, for networks changing over time, the temporal evolution of the diameter can give important insights into the structure of the dynamics [7].

Not much is known about the behavior of the diameter of random network ensembles. At least it is known that the average diameter for many ensembles grows logarithmically with the number of nodes [8]. Nevertheless, a full description, i.e., the probability distribution of network diameters over the instances of an random graph ensemble, has been obtained only in very limited cases, to the knowledge of the authors.

Here, we deal with the most fundamental and least-structured graph ensemble, Erdős-Rényi (ER) random graphs [9]. Let $N = |V|$ denote the number of vertices. Each realization of an ER random graph is generated by iterating over the $N(N-1)/2$ pairs i, j of nodes and adding the edge $\{i, j\} \in V^{(2)}$ with probability p. Here we concentrate on the sparse case $p = c/N$, c being the average connectivity.

In the nonpercolating phase $c < 1$, close to the percolation threshold $c \nearrow 1$, the distribution of diameters is described in theorem 11(iii) of Ref. [10]. The distribution is asymptotically ($N \to \infty, c \to 1$) given by a Gumbel (extreme-value) distribution

$$P_{\mathrm{G}}(d) = \lambda e^{-\lambda(d-d_0)}\, e^{-e^{-\lambda(d-d_0)}}. \tag{1}$$

Here, d_0 is the maximum of the distribution, which scales logarithmically with the number N of nodes. λ is the Gumbel parameter describing the exponential behavior $P_{\mathrm{G}} \sim e^{-\lambda d}$ for large values. It describes the variance, which is proportional to $1/\lambda^2$. In this $c \to 1$ limit, the Gumbel parameter λ as a function of the connectivity c is given by

$$\lambda(c) = -\ln(c). \tag{2}$$

The fact that $P(d)$ takes the form of an extreme-value distribution is intuitively clear: Below the percolation threshold, each graph consists of a large number of trees, hence the diameter is obtained by maximizing over these trees.

Note that Ref. [10] also contains results for general values of $c < 1$. Although they are given in a more complex and partially implicit form, they indicate that the asymptotic distribution is also the Gumbel distribution, Eq. (1), with a parameter λ also given by Eq. (2).

Due to finite-size corrections, the distribution of diameters in finite-size graphs does not follow the Gumbel distribution. In earlier numerical studies of another problem, sequence alignments [11–13], the data were well fitted by "modifying" the Gumbel distribution by a Gaussian factor:

$$P_{\mathrm{mG}}(d) = \lambda' P_{\mathrm{G}}(d) e^{-a(d-d_0)^2}, \tag{3}$$

*a.hartmann@uni-oldenburg.de
†marc.mezard@ens.fr

where λ' is given through the normalization $\int P_{\rm mG}(\delta)\,d\delta = 1$. This distribution will be used in our analysis.

The probabilities $P(d)$ for values of d which deviate from the typical size are often exponentially small in N. Hence, one uses the concept of the large-deviation *rate function* [14,15] by writing

$$P(d) = e^{-N\Phi(d/N)+o(N)} \quad (N \to \infty)\,. \tag{4}$$

Note that the normalization is part of the $e^{o(N)}$ factor. One says that the *large-deviation principle* holds if, loosely speaking, the empirical rate function

$$\Phi_N(d/N) \equiv -\frac{1}{N}\ln P(d) \tag{5}$$

converges to $\Phi(d/N)$ for $N \to \infty$. Due to the logarithm the normalization and the subleading term of $P(s)$ become an additive contribution to Φ, which go to zero for $N \to \infty$.

In the present work, we study numerically the distribution $P(d)$ of diameters of ER random graphs in the sparse regime $p = c/N$ by using Markov-chain Monte Carlo simulations. Using a large-deviation technique which is based on studying a biased ensemble characterized by a finite temperature-like parameter, see Sec. II, we are able to obtain the distributions over almost the full range of the support, down to very small probabilities like 10^{-100}. For the nonpercolating regime $c < 1$, we compare our numerical results to the available analytical results and find a good agreement. In particular we find that the asymptotic $c \nearrow 1$ result of a suitably scaled Gumbel distribution, modified by a Gaussian finite-size correction, is compatible with our results for all values $c < 1$. Also, we find the dependence Eq. (2) of the Gumbel parameter λ as a function of c, for $N \to \infty$. This confirms the validity of our approach.

We are also able to obtain $P(d)$ numerically for $c > 1$. Here we find that the distributions exhibit an inflection point. This leads to a first-order transition in our finite-temperature ensemble and makes the numerical determination of the distribution much harder.

Nevertheless, for all values of c we determined the rate functions for various numbers N of nodes and obtained, where necessary, the limiting rate function via extrapolation. In all cases we observed a good convergence, indicating that the large-deviation principle holds.

II. SIMULATION AND REWEIGHTING METHOD

We are interested in determining the distribution $P(d)$ of the diameter in an ensemble of random graphs. The distribution can be obtained in principle for any graph ensemble, here we apply it to ER random graphs. *Simple sampling* is straightforward: One generates a certain number K of graph samples and obtains the diameter $d(G)$ for each sample G. This means each graph G will appear with its natural ensemble probability $Q(G)$. The probability that the graph has diameter d is given by

$$P(d) = \sum_G Q(G)\delta_{d(G),d}\,. \tag{6}$$

Therefore, by calculating a histogram of the values for d, an estimation for $P(d)$ is obtained. Nevertheless, with this simple sampling, $P(d)$ can only be measured in a regime where $P(d)$ is relatively large, about $P(d) > 1/K$. Unfortunately, the distribution usually decreases very quickly, e.g., exponentially in the system size N when moving away from its typical (peak) value, like in Eq. (4). This means that, even for moderate system sizes N, the distribution will be inaccessible through this method on almost its complete support.

A. Markov-chain Monte Carlo approach

To estimate $P(d)$ for a much larger range of diameters, we use a different *importance sampling* approach [11,16]. For self-containedness, the method is outlined subsequently. The basic idea is to generate random graphs with a probability that includes an additional Boltzmann factor $\exp(-d(G)/T)$, T being a "temperature" parameter, in the following manner: A standard Markov-chain Monte Carlo simulation [17,18] is performed, where the current state at "time" t is given by an instance of a graph $G(t)$. The procedure is similar to the standard use of Monte Carlo (MC) simulations for sampling configuration of spin systems by sequentially generating spin configurations. In the present case we generate a sequence of graphs. We use the Metropolis algorithm [19] as follows: At each step t a *candidate* graph G^* is created from the current graph $G(t)$. One then computes the diameter of the candidate graph, $d(G^*)$. To complete a step of the Metropolis algorithm, the candidate graph is *accepted* $[G(t+1) = G^*]$ with the Metropolis probability

$$p_{\rm Met} = \min\{1, e^{-[d(G^*)-d(G(t))]/T}\}\,. \tag{7}$$

Otherwise (with probability $1 - p_{\rm Met}$) the current graph is kept $[G(t+1) = G(t)]$. The temperature parameter T allows us to bias the sampling toward graphs of larger or smaller diameters. Using an infinite temperature leads to acceptance of all configurations; this gives access to the typical graph diameter. With positive T temperature, one samples graphs with diameters smaller than the typical diameter. The use of negative T allows us to sample graphs with diameters larger than the typical one.

Here, the generation of G^* is done using the following local update rule: A node $i \in V$ of the current graph is selected randomly, with uniform weight $1/N$, and all incident edges are deleted. Next, the node i is reconnected again; for all other nodes $j \in V$ the corresponding edge $\{i,j\}$ is added with a probability c/N (and not added with probability $1 - c/N$), which corresponds to its contribution to the natural weight $Q(G)$ of an ER graph.

By construction, the algorithm fulfills detailed balance. Clearly the algorithm is also ergodic, since within N steps, each possible graph may be constructed. Thus, in the limit of infinitely long Markov chains, the distribution of graphs will follow the probability

$$q_T(G) = \frac{1}{Z(T)}Q(G)e^{-d(G)/T}\,, \tag{8}$$

where $Z(T)$ is the *a priori* unknown normalization factor. Note that for $T \to \infty$ all candidate graphs will be accepted and the distribution of graphs will follow the original ER weights.

B. Obtaining the distribution

The probability to measure d at any temperature T is given by

$$\begin{aligned} P_T(d) &= \sum_G q_T(G)\delta_{d(G),d} \\ &\overset{(8)}{=} \frac{1}{Z(T)} \sum_G Q(G) e^{-d(G)/T} \delta_{d(G),d} \\ &= \frac{e^{-d/T}}{Z(T)} \sum_G Q(G)\delta_{d(G),d} \\ &\overset{(6)}{=} \frac{e^{-d/T}}{Z(T)} P(d) \\ \Rightarrow \quad P(d) &= e^{d/T} Z(T) P_T(d)\,. \end{aligned} \tag{9}$$

Hence, the target distribution $P(d)$ can be estimated, up to a normalization constant $Z(T)$, from sampling at finite temperatures T. For each temperature, a specific range of the distribution $P(d)$ will be sampled, with a bias toward smaller d when positive temperatures are used, and a bias toward larger d when negative temperatures are used. In both cases, temperatures of large absolute value will cause a sampling of the distribution close to its typical value, while temperatures of small absolute value are used to access the tails of the distribution. Hence one chooses a suitable set of temperatures $\{T_{-N_n}, T_{-N_n+1}, \ldots, T_{N_p-1}, T_{N_p}\}$ with N_n and N_p being the number of negative and positive temperatures, respectively. A good choice of the temperatures is such that the resulting histograms of neighboring temperatures overlap sufficiently. This allows us to "glue" the histograms together, see next paragraph. By obtaining the distributions $P_{T_{-N_n}}(d), ..., P_{T_{N_p}}(d)$, such that $P(d)$ is "covered" as much as possible, one can measure $P(d)$ over a large range, possibly on its full support. The range where the distribution can be obtained may be limited, e.g., when the MC simulations at certain temperatures T_k do not equilibrate. This happens usually for small absolute values $|T_k|$, where the system might also have a glassy behavior. Another difficult case is when $P(d)$ is not concave: Since $P_T(d)$ is proportional to the product of $P(d)$ and the exponential $e^{-d/T}$, this product will always have one maximum if $\log P(d)$ is concave. If $\log P(d)$ has an inflection point, $P_T(d)$ exhibits two local maxima, and the global maximum depends on the value of T. In this case a first-order transition will appear [16] as a function of T, which might prevent one from obtaining $P(d)$ in some regions of the support for large systems.

The normalization constants $Z(T)$ can easily be computed, e.g., by including a histogram generated from simple sampling, which corresponds to the temperature $T = \pm\infty$. Using suitably chosen temperatures T_{+1}, T_{-1}, one measures histograms which overlap with the simple-sampling histogram on its left and right border, respectively. Then the corresponding *relative* normalization constants $Z_r(T_{\pm 1})$ can be obtained by the requirement that, after rescaling the histograms according to Eq. (9), they must agree in the overlapping regions with the simple-sampling histogram within error bars. This means, the histograms are "glued" together. Here, we achieve this by minimizing the mean-squared difference between the two rescaled histograms of, e.g., temperatures T_i and T_{i+1}:

$$\sum_d{}' [e^{d/T_i} Z(T_i) P_{T_i}(d) - e^{d/T_{i+1}} Z(T_{i+1}) P_{T_{i+1}}(d)]^2\,,$$

where the sum $\sum_d'$ runs over the histogram entries d in the overlapping region. The minimization constraint is equivalent to the requirement that, after dividing the above expression by $Z(T_i)^2$, its derivative with respect to $Z_r(T_{i+1}) \equiv Z(T_{i+1})/Z(T_i)$ is equal to zero. This leads to

$$Z_r(T_{i+1}) = \frac{\sum_d' e^{d/T_i} P_{T_i}(d)}{\sum_d' e^{d/T_{i+1}} P_{T_{i+1}}(d)}. \tag{10}$$

In practice, in this computation we exclude data points at boundaries of the histograms, where the statistics is bad. One could also include weights which are inverse to the statistical errors of the data points.

In this manner, the range of covered d values can be extended iteratively to the left and to the right by choosing additional suitable temperatures $T_{\pm 2}, T_{\pm 3}, \ldots$, and gluing the resulting histograms one to the other. The histogram obtained in the end can be normalized (with constant Z), such that the probabilities sum up to unity. This also yields the actual normalization constants $Z(T) = Z_r(T)/Z$ from Eq. (9). Note that one could also not only glue together neighboring histograms, but also use for each bin value d all available data, as is done, e.g., within the multihistogram approach by Ferrenberg and Swendsen [20]. For the present case, it was sufficient to use the histograms just pairwise. A pedagogical explanation of the gluing process and examples of this procedure can be found in Ref. [21].

C. Equilibration

To obtain the correct result, the MC simulations generating Markov chains of different graphs must be equilibrated. In our case, there is an easy test that can indicate when equilibration has not been reached. One can run two simulations starting with two different initial graphs $G(t_{\mathrm{MCS}} = 0)$:

(1) A graph which is arranged in a line, i.e., it has a linear structure with N nodes and $N-1$ edges. The diameter of this graph is $N-1$.

(2) A complete graph, which contains all $N(N-1)/2$ possible edges. For this graph the diameter is unity.

For each of these two different initial conditions, the evolution of $d(t_{\mathrm{MCS}})$ will approach the equilibrium range of values from two different extremes, which allows a simple equilibration test: If the measured values of d disagree within the range of fluctuation, equilibration is not achieved. We shall assume that, conversely, if the measured values of d agree, then equilibration has been obtained.

Note that one can also use generalized ensemble methods like the Multicanonical method [22] or the Wang-Landau approach [23], in particular when a first-order transition as function of T appears, to obtain the distribution $P(d)$. While these methods in principle do not require to perform independent simulations at different values for the temperatures, it turns out that for larger system sizes, one still has to perform multiple simulations. The reason is that the interval of interest must be split into smaller overlapping subintervals in

order to make the simulation feasible. Here, the Wang-Landau algorithm was used only for the case of $c = 3$, where the temperature-based approach did not work well, see below. For other values of c, it turned out to be much easier to guide the simulations toward the regions of interest, e.g., where data are missing using the so-far-obtained data, and to monitor the equilibration process.

III. RESULTS

We have numerically determined the distribution of diameters for ER random graphs for different connectivities below, above, and at the percolation threshold $c_c = 1$.

To set up the simulations, one first has to choose suitable temperatures to address different ranges of the support. For the present study, for connectivities smaller than 1, only a few different values were sufficient. For example, for $c = 0.6$ and $N = 1000$, the temperature set $\{-1.5, -1.2, -1\}$ was sufficient. For $c = 0.9$ and $N = 1000$ the set $\{-3, -2, -1.5, -1\}$ was used. Above the percolation threshold, more temperatures were needed. For $c = 2.0$, $N = 1000$ a larger set of 13 different temperatures in the range $[-10, 1]$ was necessary. Equilibration, in the cases where it was achieved, was rather fast. The longest equilibration time scales were observed for $c = 0.9, N = 1000$ and $c = 2, N = 500$, which were about 1000 MC sweeps. One sweep means N trials for isolating and randomly reconnecting a randomly chosen node. In these two cases, another 7000 sweeps where used for sampling. On a standard PC such a simulation (per temperature one Intel Core i3, 3GHz) takes about one day, i.e., about five sweeps per minute. Note that most computational effort goes into the calculation of the pairwise shortest paths. Here we used the *all-pairs-shortest-path* algorithm provided by the LEDA library [24]. This implementation has a running time which scales like $O(N^2 \ln N)$ for diluted graphs. For the case $c = 3$, see below, we also performed Wang-Landau simulations, because the Boltzmann-biased MC simulations did not equilibrate. For the Wang-Landau case nevertheless the numerical effort was larger, but convergence was achieved in a large range of the support. For the largest size $N = 1000$, 119 independent MC simulations had to be performed.

A. Connectivity $c < 1$

We start with the nonpercolating regime, where we can compare with exact asymptotic results [10]. In Fig. 1 $P(d)$ is shown at $c = 0.6$ for three different graph sizes. Using the approach explained in the previous section, probabilities as small as 10^{-100} are easily accessible. In the linear-log plot, clearly a curvature in the data is visible for large values of the diameter, which could be partially due to strong finite-size corrections. We have fitted the data to a modified Gumbel distribution, Eq. (3), and obtained good fit qualities. We studied the strength a of the Gaussian correction, see inset of Fig. 1. One observes a clear power-law behavior. Hence, the numerical data support that asymptotically the full distribution becomes Gumbel or Gumbel-like below the percolation threshold.

We have studied the behavior in the nonpercolating phase for various values of the connectivity c and different system

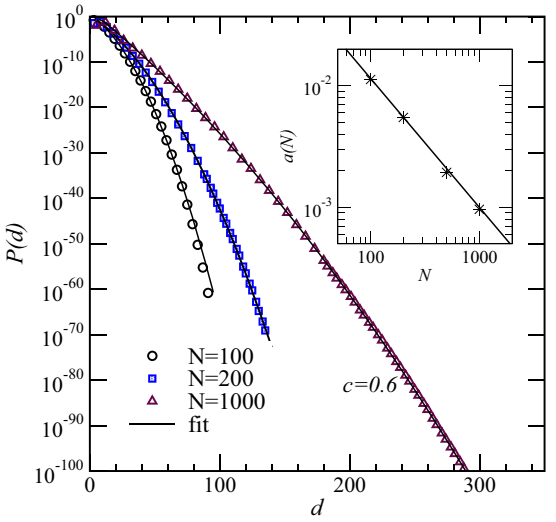

FIG. 1. Distribution of diameters for Erdős-Rényi random graphs with average connectivity $c = 0.6$ for three different graph sizes $N = 100, 200$, and 1000. The lines show fits to the modified Gumbel distribution according to Eq. (3). The inset shows the dependence of the parameter a of Eq. (3) on the system size N. The line displays the power law $a \sim 1.51 N^{-1.06}$.

sizes, see Fig. 2 for $c = 0.9$. Each time we observe qualitatively the same convergence to a Gumbel distribution. From the fits, for each value of the connectivity c and each system size N, a value of $\lambda(c, N)$ is obtained. To extrapolate the dependence of the Gumbel parameter $\lambda(c)$ to large graph sizes, the following heuristic dependence was applied:

$$\lambda(c, N) = \lambda(c) + bN^{-\alpha}. \tag{11}$$

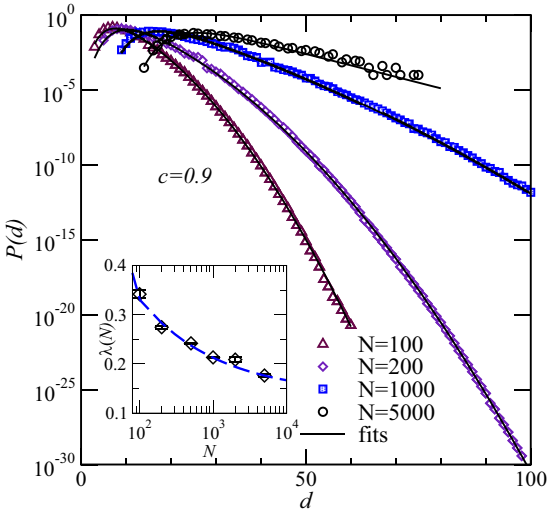

FIG. 2. Distribution of diameters for Erdős-Rényi random graphs with average connectivity $c = 0.9$ for four different graphs sizes $N = 100, 200, 1000$, and 5000. The lines show fits to the modified Gumbel distribution according to Eq. (3). The inset displays the dependence of the Gumbel parameter λ as a function of graph size N. The inset shows the result of a fit to the function Eq. (11).

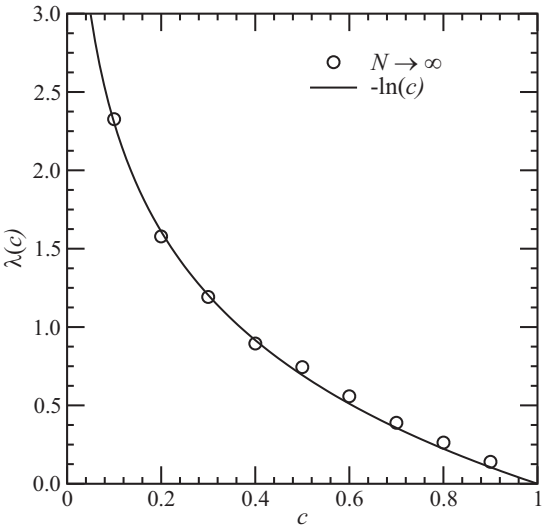

FIG. 3. Dependence of the Gumbel parameter λ as a function of the connectivity c. The symbols show the extrapolation of the numerical results using Eq. (11). The error bars are of order of symbol size. The solid line represents the mathematical result Eq. (2) of Ref. [10].

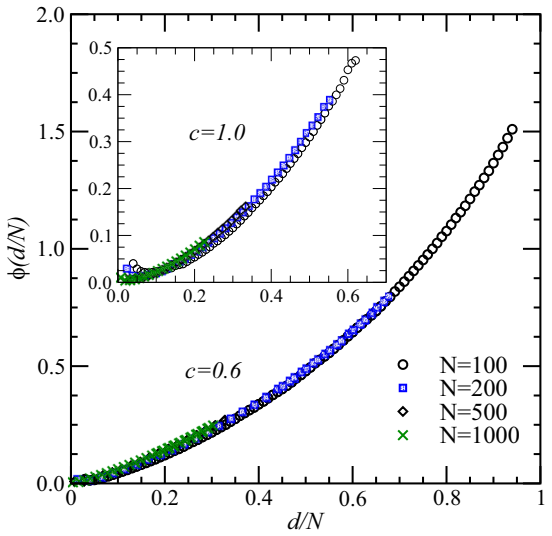

FIG. 4. Large-deviation rate function Φ as a function of the rescaled diameter d/N for $c = 0.6$ and different system sizes. The inset shows the large-deviation rate function for the percolation threshold $c = 1$.

Note that we have no justification for such a relation from analytical arguments concerning the Gumbel distribution. The relation is purely heuristic and based on the fact that such a behavior is often observed in standard finite-size scaling theory [25,26] used for the analysis of phase transitions.

The inset of Fig. 2 shows the behavior of $\lambda(c = 0.9, N)$ together with the fit as a function of graph size N. Here, we found the values $\lambda = 0.16(2)$ and $\alpha = 0.4(1)$. For smaller values of c the convergence is even faster, for $c \leqslant 0.5$ basically the same value of λ is obtained for all sizes within fluctuations. The resulting values for λ as a function of the connectivity c are shown in Fig. 3 together with the asymptotic result of Eq. (2), yielding a nice agreement. This shows that indeed the numerical approach allows us to reliably study the distribution of diameters for finite sizes and to extrapolate to large graphs.

Nevertheless, the scaling of the Gaussian correction parameter is close to $a \sim 1/N$, hence when looking at the data for the rescaled diameter $\tilde{d} = d/N$, the size dependence exactly drops out. Hence, the rate function of Eq. (5) as a function of $\hat{d}$ is studied next, as displayed in Fig. 4. The data collapse is good, which means that even for small system sizes the rate function is well converged. This indicates that the distribution of diameters can indeed be described well by a rate function, hence the large-deviation principle [14,27] holds. It is thus likely that this model is accessible to the mathematical tool of large-deviation theory. Note that due to the curvature of the rate function, the Gumbel distribution is not visible when inspecting the result on the d/N scale, which makes the most important finite-size contribution drop out. Only when one looks at the data at fixed values d is the convergence to a Gumbel a meaningful statement. A similar result has been observed previously for the distribution of scores of sequence alignments [13].

B. Connectivity $c = 1$

The results for the nonpercolating phase gave us confidence in the numerical method. We now use it to study cases where no exact results for the full distribution are available. In Fig. 5, the distribution of diameters is shown right at the percolation transition $c = 1$. Here, the random graph consists of a large extensive tree plus small components and no Gumbel distribution is expected, since $\lambda = 0$.

Nevertheless, it is still possible to fit the finite-size data to the modified Gumbel distribution, see Fig. 5, since any finite-size

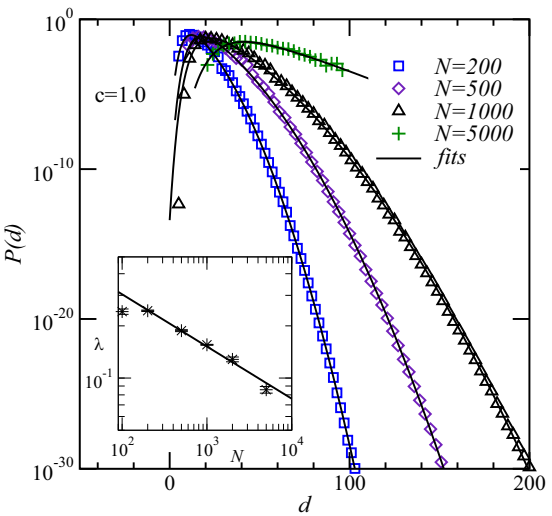

FIG. 5. Distribution of diameters for Erdős-Rényi random graphs with average connectivity $c = 1.0$ for different graph sizes. The lines show fits according to Eq. (3). The inset shows the dependence of the fitting parameter λ as a function of graph size using a double-logarithmic axis. The lines shows the power-law $\sim N^{-\alpha}$ with $\alpha = -0.3$ obtained from fitting the $\lambda(N)$ data for $N \geqslant 200$.

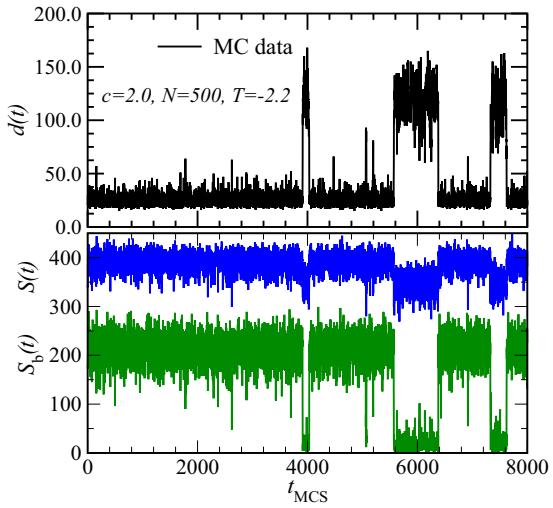

FIG. 6. Top: Time series for the diameter d as a function of the number t_{MCS} of MC sweeps, for $c = 2$, graph size $N = 500$ at artificial temperature $T = -2.2$. Bottom: for the same run, the size S of the largest component and the size S_{b} of the largest biconnected component as a function of t_{MCS}.

graph for $c = 1$ cannot be distinguished from the case c close to 1. For example, fitting the case $N = 1000$ to Eq. (3) resulted in

$$\lambda = 0.151(1), \quad d_0 = 22.9(1), \quad a = 1.304(6) \times 10^{-3}.$$

The fit matches the data well, also for other systems sizes. Nevertheless, when studying the dependence of λ on the system size, a convergence toward zero seems most likely, see inset of Fig. 5. In the double-logarithmic plot the data appear to be compatible with a straight line, meaning a power-law decrease and maybe even a faster decrease. We verified this by fitting $\lambda(N)$ according to Eq. (11), where we obtained a negative value for λ for $N \to \infty$ with an error bar of almost the same size, showing that indeed the distribution $P(d)$ differs from the Gumbel distribution for $c = 1$.

This can also be seen from studying the large-deviation rate function $\Phi(d/N)$, see inset of Fig. 4 where also an upward-bending function is seen, as for the case $c = 0.6$.

C. Connectivity $c > 1$

In the percolating phase $c > 1$, the numerical results show that in the artificial finite-temperature ensemble there appears to be a first-order phase transition as a function of the temperature, similar to the distribution of the size of the largest component [16]. To visualize this, we here show an example of a MC time series for the diameter, see Fig. 6. Clearly the diameter oscillates between two distinct regimes, showing the coexistence of two "phases" with small and large diameters. This corresponds at this temperature to a distribution of diameters $P_T(d)$ with two peaks.

The data in Fig. 6 of the size S of the largest component and the size S_{b} of the largest bi-connected component suggest that the system oscillates between two states. When the diameter is small, around 30 here, the largest biconnected component is large, it contains about 200 nodes. On the other hand,

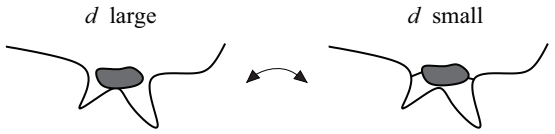

FIG. 7. Bimodal characteristics of the graphs at phase coexistence ($c > 1$). There is a large linear structure plus a tightly connected structure (plus a lot of small components which are not so important). At criticality, the system oscillates between states where the strongly connected structure is attached several times to the linear structure, allowing for shortcuts, or only weakly connected (sometimes even not at all).

when the diameter is large, about 130, the largest biconnected component has a size of only about 30. Nevertheless, the size of the largest components changes only a little bit. This we interpret in the following way.

There is always one large line-like object present and a tightly (bi)connected cluster, see Fig. 7. As we have verified explicitly in our numerical data, the tightly connected cluster typically has a small diameter while the line-like object has a large diameter. In one state, the line-like object is connected to the tightly connected cluster only at a few nodes or not at all. Hence, the diameter path is basically along the line-like object and the diameter is large. In the other state, the line-like object is connected to the tightly connected cluster at several distant points, such that the diameter path makes a shortcut. Thus, the diameter is small and the biconnected component relatively large. Our measurements showed that although the diameters of the two states differ strongly, the number of edges differ only slightly (not shown here). This allows a quick transition between the two states. Note that in the mathematical literature [28] for a logarithmically growing (but not too large) connectivity for diluted ER random graphs, the diameter is concentrated around a finite number of values. This might be related to the observed oscillations of the diameter.

The existence of an inflection point in $\log P$ translates into the existence of a first-order transition in the finite temperature ensemble with some critical temperature T_{c}. This leads to a bimodal structure of $P_{T_c}(d)$ exhibiting a very small probability in the region between the two peaks. Hence, concerning the numerical effort, obtaining the full distribution $P(d)$ becomes difficult, in particular for large graphs, because the intermediate region for values of d between the two peaks of $P_T(d)$ is rarely or even not at all sampled.

Therefore, for the case $c = 2.0$, only system sizes up to $N = 500$ could be equilibrated deep into the large-diameter regime (corresponding to negative temperatures with small absolute values). Larger sizes, in particular for even larger connectivities c, cannot be addressed by the MC sampling with a Boltzmann weight at temperature T. For this reason, we used the Wang-Landau algorithm for larger connectivities, see below. The first-order nature of the transition, i.e., the two-peak structure of the distributions of the diameter at finite temperature close to the transition temperature, becomes visible by a "dip" in the distribution of diameters for $N = 500$, see Fig. 8.

The dip is more pronounced when going to larger connectivities. This can be seen in Fig. 9, where the large-deviation rate function is displayed for $c = 3.0$ for different system sizes.

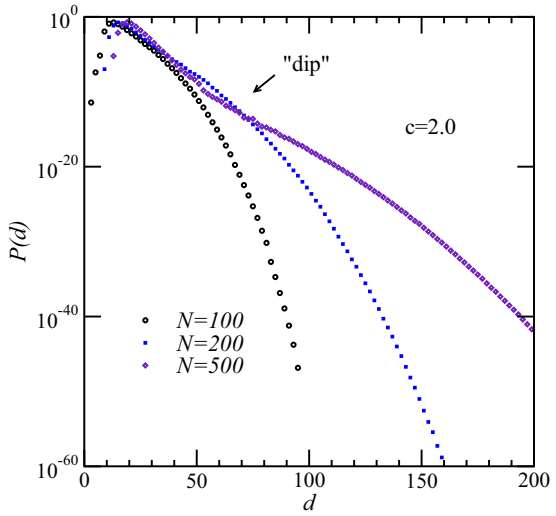

FIG. 8. Distribution of diameters for Erdős-Rényi random graphs with average connectivity $c = 2.0$ for graph size $N = 100$, 200, and 500.

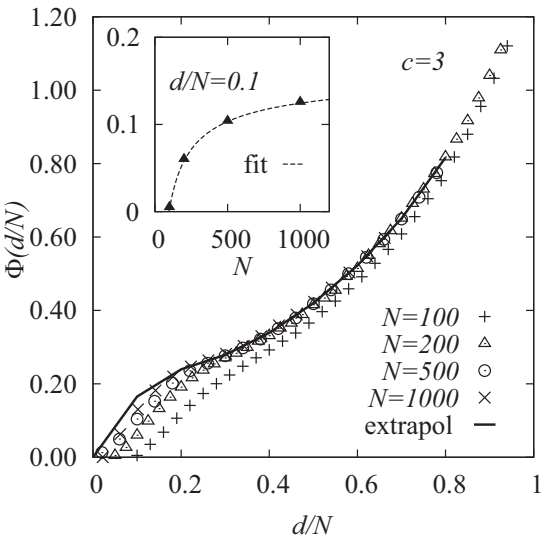

FIG. 9. Large-deviation rate function Φ as a function of the rescaled diameter d/N for $c = 3.0$ and different system sizes $N = 100$, 200, 500, and 1000 (symbols). The line shows the rate function obtained from extrapolation $N \to \infty$. Inset: extrapolation of rate function as function of system size, for $d/N = 0.1$.

Here, the Wang-Landau approach was used, which allowed us to sample the region of the dip much better.

In Fig. 9 the corresponding rate function is shown. Here, in particular for small values of d/N, stronger finite size effects are visible. Thus, we have considered the functions $\Phi_N(d/N)$ at various fixed ratios $\tilde{d} = d/N$ and performed an extrapolation via fitting

$$\Phi_N(\tilde{d}) = \Phi(\tilde{d}) + bN^{-\beta}, \tag{12}$$

where b and β are fitting parameters which are determined for each considered value of $\tilde{d}$ separately, i.e., point-wise. An example of the extrapolation is shown in Fig. 9, together with the extrapolated values $\Phi(d/N)$. For large values of d/N, above 0.5, the finite size effects are small, while for small values of d/N the extrapolated function differs slightly from the results for finite values of N. The change from a concave to a convex function near $d/N = 0.3$ is well visible. A similar qualitative behavior has been found previously for the rate function Φ for the distribution of the size of the largest component for ER random graphs [16].

IV. SUMMARY

We have studied the distribution of the diameter for dilute ER random graphs with connectivities c. Using large-deviation simulation techniques, we were able to obtain the distributions over a large range of the support. In the nonpercolating region of small connectivities $c < 1$, the distributions are concave and can be well fitted to the Gumbel distributions with a Gaussian correction. The extrapolated parameter of the Gumbel distribution agrees well with mathematical results. In the percolating regime $c > 1$ the distribution of the diameters is not available analytically. Within the numerical result, we observed a change of the rate function from concave to convex behavior, thus a more complex distribution. Nevertheless, for all values of c studied, we were able to obtain and extrapolate the rate function. This means that the distribution of diameters follows the large-deviation principle.

Since the diameter is of importance for many physical processes taking place on networks, it would be interesting to obtain the distribution over a large range of the support for other graph ensembles, like scale free graphs. The results obtained in the present work show that this should in principle be possible.

ACKNOWLEDGMENTS

We thank Hendrik Schawe for many valuable discussions and for critically reading the manuscript. This project was supported by the German Humboldt Foundation. M.M. thanks the University of Oldenburg for its hospitality. A.K.H. is grateful for the hospitality and financial support of the LPTMPS, Université Paris-Sud, where he spent part of his sabbatical. The simulations were performed on the HERO cluster of the University of Oldenburg jointly funded by the DFG (INST 184/108-1 FUGG) and the ministry of Science and Culture (MWK) of the Lower Saxony State.

[1] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University, New York, 2010).

[2] M. Newman, *Networks: An Introduction* (Oxford University, Oxford, 2010).

[3] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University, Oxford, 2011).

[4] B. Doerr, M. Fouz, and T. Friedrich, Commun. ACM **55**, 70 (2012).

[5] A. K. Hartmann, Eur. Phys. J. B **87**, 114 (2014).
[6] A. Goldental, R. Vardi, S. Sardi, P. Sabo, and I. Kanter, Front. Neural Circ. **9**, 65 (2015).
[7] B. Blonder, T. W. Wey, A. Dornhaus, R. James, and A. Sih, Meth. Ecol. Evol. **3**, 958 (2012).
[8] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[9] P. Erdős and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).
[10] T. Łuczak, Rand. Struct. Alg. **13**, 485 (1998).
[11] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).
[12] S. Wolfsheimer, B. Burghardt, and A. K. Hartmann, Alg. Mol. Biol. **2**, 9 (2007).
[13] L. Newberg, J. Comp. Biol. **15**, 1187 (2008).
[14] F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2000).
[15] H. Touchette, Phys. Rep. **478**, 1 (2009).
[16] A. K. Hartmann, Eur. Phys. J. B **84**, 627 (2011).
[17] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon, Oxford, 1999).
[18] D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University, New York, 2000).
[19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[20] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[21] A. K. Hartmann, in *New Optimization Algorithms in Physics*, edited by A. K. Hartmann and H. Rieger (Whiley-VCH, Weinheim, 2004), p. 253.
[22] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[23] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[24] K. Mehlhorn and S. Näher, *The LEDA Platform of Combinatorial and Geometric Computing* (Cambridge University, New York, 1999).
[25] J. Cardy, *Finite-Size Scaling* (Elsevier, Amsterdam, 1988).
[26] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, MA, 1992).
[27] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer, Berlin, 2010).
[28] F. Chung and L. Lu, Adv. Appl. Math. **26**, 257 (2001).