

Large Deviation Properties of Minimum Spanning Trees for Random Graphs

Mahdi Sarikhani¹ and Alexander K. Hartmann²

¹*Department of Physics, College of Science, Shiraz University, Shiraz, Iran*

²*Institute of Physics, University of Oldenburg, Oldenburg, Germany*

(Dated: December 15, 2025)

We study the large-deviation properties of minimum spanning trees for two ensembles of random graphs with N nodes. First, we consider complete graphs. Second, we study Erdős-Rényi (ER) random graphs with edge probability $p = c/N$ conditioned to be connected. By using large-deviation Markov chain sampling, we are able to obtain the distribution $P(W)$ of the spanning-tree weight W down to probability densities as small as 10^{-300} . For the complete graph, we confirm analytical predictions with respect to the expectation value. For both ensembles, the large deviation principle is fulfilled. For the connected ER graphs, we observe a remarkable change of the distributions at the value of $c = 1$, which is the percolation threshold for the original ER ensemble.

I. INTRODUCTION

Graphs and networks [1–6], are ubiquitous in modeling systems like chemical molecules, disordered materials, energy grids, spread of diseases, or social networks. Graphs can be analyzed in many ways, like the calculation of the connected components, graph diameter, resilience to node or edge removal, communities, minimum matchings, spectral properties, and many more; all these have found applications in physical systems.

Here we consider *spanning trees*, which are subsets of edges of a graph that connect all vertices. Being a tree, for N nodes the spanning tree has $N - 1$ edges, and from each vertex to any other one, there exists a unique path. If the edges carry weights, like distances or costs, the *minimum spanning tree* is among all spanning trees the one where the sum W of the weights of the tree edges is a minimum.

The study of optimum trees dates back [7] to the works of Fermat and Torricelli in the 17th century. In practical applications and research, spanning trees have found many applications, e.g., to set up transportation networks [8], reconstruct evolutionary trees [9], cluster data [10], detect communities [11], determine patterns in images [12], and many more. In statistical physics, among several other applications, spanning trees have been analyzed by counting [13] or enumerating [14] them. Also, spanning trees have been used to analyse atomic deposition structures on surfaces [15], avalanches in sandpile models [16], transport [17] and other technical networks [18], stock markets [19, 20], and fractal properties of percolation clusters [21–23]. Finally, spanning trees have also attracted the attention of mathematicians, where we use some of the work [24–29] for comparison with our results below.

In this work, we study numerically the distribution $P(W)$ of the weights of minimum spanning trees for two random ensembles. The first one consists of complete graphs, where the edge weights are randomly, independently and identically distributed (iid) uniformly in the interval $[0, 1]$. For this case, the expectation value of W in the limit of a large number N of nodes is known

[25], also that the probability to deviate from this value decreases exponentially with N . But details about the distribution are not known. The second ensemble we consider consists of Erdős-Rényi (ER) random graphs conditioned to be connected, again the weights are iid uniform in $[0, 1]$. Here, to the best of our knowledge, nothing is known about the distribution $P(W)$. The approach we use utilizes a Markov-chain Monte Carlo large-deviation sampling [30, 31] of random graphs, which allows us to access the tails of the distributions down to exponentially small probabilities such as 10^{-300} . This means we can obtain the distribution over a very large range of support. This ability is useful to verify the so-called large-deviation principle [32, 33]. Also, we hope that our study motivates more analytical work, possibly involving approximation techniques, which can be compared to our results.

The paper is organized as follows. In Sec. II we introduce minimum spanning trees, the graph ensembles we use, the quantities we measure, and mention previous results, which are relevant for our work. Next, we explain the algorithms we have used to obtain the distribution of spanning-tree weights over hundreds of decades in probability density. In Sec. IV we present our results before we finish with a summary and outlook.

II. MODELS

A *graph* $G = (V, E)$ consists of N nodes $i \in V$ and M undirected edges $\{i, j\} \in E \subset V^{(2)}$. For each edge $\{i, j\} \in E$, the nodes i and j are called *adjacent*. The edge is said to be *incident* to these two adjacent nodes. The *degree* of a node $i \in V$ is the number of nodes adjacent to i . A sequence $P = (i_1, \dots, i_l)$ of nodes, where all pairs i_k, i_{k+1} of consecutive nodes are adjacent, is called a *path* of length $l - 1$; the two nodes i_1 and i_l are called *connected* by the path P . If a path connects a node with itself, i.e., $i_1 = i_l$, the path is called a *cycle*. A graph that does not contain a cycle is called *acyclic*. If in a graph all nodes are connected, the graph is also called *connected*. Here we consider *weighted* graphs, i.e., each

edge $\{i, j\} \in E$ is assigned a real-valued weight $w_{i,j}$.

A *spanning tree* $T = (V, E_T)$ for a given graph $G = (V, E)$ consists of the set V of nodes and an acyclic subset E_T of E that connects all nodes, i.e., a tree that spans the graph G . The total edge weight of a spanning tree is the sum

$$\tilde{W} \equiv \tilde{W}(T) = \sum_{\{i,j\} \in E_T} w_{i,j} \quad (1)$$

of the weights of the edges of the tree. The *minimum spanning tree* (MST) of graph G is a spanning tree where the weight $\tilde{W}$ is minimal; we denote the minimum weight as W . Note that in general the minimum spanning tree may not be unique, but for the ensembles we consider here this is the case, so we can speak of *the* MST, and we also write $W(G)$ for $W(T(G))$.

Here we do not consider just single specific given graphs, but ensembles of a fixed or random connected graphs with, in any case, randomly chosen edge weights, which means that the minimum spanning-tree weight W is a random variable and will be described by a probability distribution with density $P(W)$. If we denote the probability density to obtain a certain graph G , which is meant to include the edge weights, by $Q(G)$, one has

$$P(W) = \sum_G \delta_{W(G), W} Q(G). \quad (2)$$

Note that since the weights are continuous variables, the $\sum_G$ actually involves also an integral, but we keep the summation for simplicity. Obtaining, i.e., numerically estimating $P(W)$, including the low-probability tails, is the purpose of this work.

The first ensemble we consider consists of weighted complete graphs $G = (V, V \times V)$, where only the weights $w_{i,j}$ are uniform iid random variables in $[0, 1]$, i.e.,

$$w_{i,j} \sim U(0, 1).$$

For this ensemble, the expectation value $\langle W \rangle_N$ was obtained analytically [25] in the limit of infinite graph size $N \rightarrow \infty$ as

$$\langle W \rangle_N \rightarrow \zeta(3)/D \text{ where } \zeta(3) = \sum_{k=1}^{\infty} 1/k^3 = 1.202 \dots \quad (3)$$

This holds actually for arbitrary distribution functions $F(w_{i,j})$ of *non-negative* edge weights, and $D = F'(0) > 0$ is the probability density at 0. For the present case with uniform distribution in $[0, 1]$, we have $D = 1$. This is compatible with a previous result [24] that $\langle W \rangle_N \leq 2(1 + \log n/n)$. Recently, also the finite-size corrections of $\langle W \rangle_N$ have been calculated [29]. With respect to the distribution, the typical part becomes Gaussian in the limit of large N with variance about 1.6857 [27]. Concerning the tails, which is our main concern here, it has been shown [26, 28] that the deviations from the typical value decrease exponentially in the graph size N , i.e.,

fulfill the large-deviation principle; see the end of this section.

Second, we study a special ER [34] random graph ensemble. Standard ER random graphs with parameter $p \in [0, 1]$ are constructed as follows. One starts with a set of N vertices. Then one iterates over the $N(N-1)/2$ possible pairs of nodes and adds each edge $\{i, j\} \in V^{(2)}$ with probability p . Of particular interest are sparse ER graphs where $p = c/N$ and c denotes the average degree of a node in the graph.

To study MSTs, we require the graphs to be connected. Thus, the second ensemble consists of ER random graphs conditioned to be connected. Technically, depth-first search [35] is used to verify connectiveness. If a graph is not connected, it is discarded in the analysis. Note that for small values of c , typical ER graphs will not be connected. Here the Markov chain Monte Carlo (MCMC) approach presented below is more efficient to generate connected ER random graphs.

In large-deviation theory [32, 33], one considers the probabilities $P(X)$ for random quantities X . The tails of $P(X)$ describe the large deviations from the typical values of X . Here we consider $X = W$, and the corresponding intensive quantity $w = W/(N-1)$. Often these deviations are for large values of N exponentially small in N as

$$P(W) \simeq e^{-N\Phi(w)}, \quad (4)$$

with the *rate function* $\Phi(w)$ which often depends on the intensive quantity. This form of the distributions has a very specific dependence on the system size N , separated from the intense quantity w . Using the small- o notation, this means that

$$P(W) = e^{-N\Phi(w) + o(N)}, \quad (N \rightarrow \infty). \quad (5)$$

The so-called *large deviation principle* holds if, loosely speaking, the distribution has shape Eq. (4) and the empirical rate function

$$\Phi_N(w) \equiv -\frac{1}{N} \ln P(wN), \quad (6)$$

converges to $\Phi(w)$ as $N \rightarrow \infty$. Due to the logarithm, both the normalization and the subleading term of $P(W)$ become additive contributions to Φ , which go to zero as $N \rightarrow \infty$.

III. ALGORITHMS

An MST can be conveniently calculated by using Prim's algorithm [8, 35]. It begins with a randomly selected node i which is the first node of the tree. Then the algorithm treats i and all subsequent nodes added to the tree in the same way: all edges adjacent to it are added to a priority queue. Iteratively the current minimum-weight edge $\{k, l\}$ is pulled out of the queue. At least one node of the edge will already belong to the tree. It is

checked whether even both nodes k and l belong to the tree. If yes, the edge is disregarded. If not, say k is in the tree but not l ; l is added to the set of tree nodes. Also the edge $\{k, l\}$ is added to the tree edges and the other edges incident to l , i.e., except $\{k, l\}$, are inserted into the priority queue. When, during the selection from the priority queue, m edges with the equal (smallest) weight appear, the degeneracy is broken by randomly selecting one with probability $1/m$. As mentioned above, due to the real-valued edge weights, there will be no degeneracy here. The algorithm stops when all nodes belong to the tree. The running time of Prim's algorithm as described above is of the order of $O(M \log N)$.

To estimate the distribution $P(W)$ in the high probability region, direct sampling is straightforward: one generates K graph samples and determines the weight of the minimum spanning tree $W(G)$ for each sample G . Each graph G occurs with its natural ensemble probability density $Q(G)$. Therefore, by calculating a histogram of the values for W , an estimation for $P(W)$ can be obtained. Also, the mean or the variance can be estimated rather accurately. Nevertheless, with this direct sampling, $P(W)$ can only be measured in a regime where $P(W)$ is relatively large, about $P(W) > 1/K$. Unfortunately, the distribution usually decays exponentially in the system size N when moving away from its typical (peak) value. Thus, even for moderate system sizes, the distribution remains unknown over most of its support.

To estimate $P(W)$ for a wider range of weights of the minimum spanning trees, including very small probability densities of the order of 10^{-100} , we use a different approach [30, 31, 36]. Here we display only the main ingredients of the algorithm, for a pedagogical introduction, see Ref. [37]. The basic idea is to generate graphs with a modified density that includes an additional exponential (Boltzmann) factor $\exp(-W(G)/\theta)$, where θ is a temperature-like parameter, which controls the weight.

To sample according to the original weights modified by the additional Boltzmann factor, we perform standard MCMC simulations, with the current state at step t is given by an instance $G(t)$ of a graph. The Metropolis-Hasting algorithm [38–41] is applied as follows. At each step t , a candidate graph G^* is created from the current graph $G(t)$. Different techniques to create candidate graphs are used for different graph ensembles:

- In the case of complete graphs, an edge $\{i, j\}$ is selected randomly, with uniform probability $1/M$, where $M = N(N - 1)/2$. The weight of the selected edge is changed to a new random value with uniform probability in the range $[0, 1]$.
- In the case of ER random graphs, a node i is selected randomly with uniform probability $1/N$. All edges adjacent to i are removed. Then, for all pairs i, j , new edges $\{i, j\}$ are added with probability $p = c/N$, and their weights are drawn uniformly from the range $[0, 1]$, respectively.

If the candidate graph G^* is not connected, it will be

immediately rejected. Note that the initial graphs also need to be connected. Therefore, most of our simulation begins with a complete graph, and the described procedure runs until the equilibration is reached; see below.

If a connected candidate graph G^* is created, we calculate the weight of its minimum spanning tree $W(G^*)$. The candidate graph is accepted ($G(t + 1) = G^*$) with the Metropolis probability

$$p = \min \left\{ 1, e^{-[W(G^*) - W(G(t))]/\theta} \right\}. \quad (7)$$

In case the candidate graph is rejected, the current graph is kept, i.e., $G(t + 1) = G(t)$.

The temperature parameter θ allows us to bias the sampling toward graphs with larger or smaller MST weights. An infinite temperature leads to acceptance of all configurations, which gives access to the typical values. Positive (negative) θ allows us to sample graphs with smaller (larger) weights of the MST than the typical one.

By construction, the algorithm fulfills detailed balance. It is also ergodic, since within M or N steps, respectively, each possible graph may be constructed. To check equilibration, we monitor the time series $W(t)$. The most simple idea is to consider the time series equilibrated if no systematic trend is visible anymore, only fluctuations. This is in particular well visible if one starts two independent runs with initial configurations that lead to very different initial weights $W(0)$, respectively, such that the time series will converge to the equilibrium range from different sides. As an example, we consider the ER ensemble with $N = 512$ and $c = 5$. For one initial graph, we take the complete graph with all weights set to $w_{ij} = 1$. Thus $W(0) = 511$. For the other initial graph, we start with a line graph of 511 edges where all weights are randomly close to zero. Here we have $W(0)$ about 50. The resulting time series $W(t)$ are shown [42] in Fig. 1 for two different temperatures, $\theta = -0.1$ and $\theta = -1$. For the first case, equilibration takes some time, more than 10^5 steps, while for $\theta = -1$, equilibration is reached faster. Note that for $\theta = -0.1$, the observed weights are close to the maximum possible value $W = 512$, and with small fluctuations. So for the simulations to obtain the distributions, we did attempt to reach such extreme values of the weight, i.e., the far tails, by suitably chosen temperature values.

Thus, in the limit of infinitely long Markov chains, the distribution of graphs will follow the probability

$$Q_\theta(G) = \frac{1}{Z(\theta)} e^{-W(G)/\theta} Q(G), \quad (8)$$

where $Q(G)$ is the above introduced original probability of the graph within its ensemble and $Z(\theta)$ is an a priori unknown normalisation constant.

The probability distribution for W at temperature θ is given by, where the $\sum_G$ is meant to involve again an

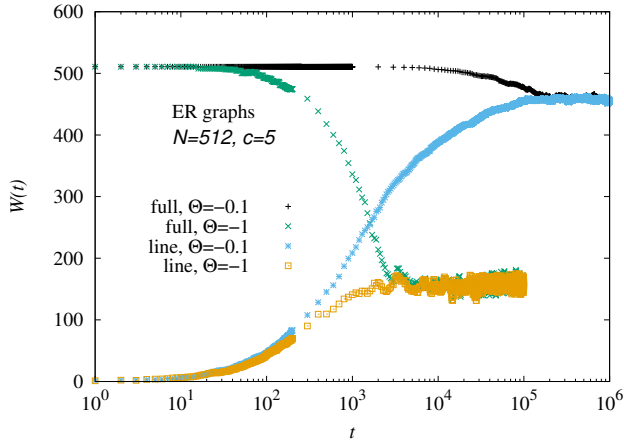


FIG. 1. Equilibration of Markov chain: time series $W(t)$ for ER graphs with $N = 512$ and $c = 5$ for two different initial configurations with full graphs (all $w_{ij} = 1$) and a line graph (all w_{ij} near 0.)

integral over the possible weight values for all edges:

$$\begin{aligned}
 P_\theta(W) &= \sum_G \delta_{W(G), W} Q_\theta(G) \\
 &= \frac{1}{Z(\theta)} \sum_G \delta_{W(G), W} e^{-W(G)/\theta} Q(G) \\
 &= \frac{e^{-W/\theta}}{Z(\theta)} \sum_G \delta_{W(G), W} Q(G) \\
 &= \frac{e^{-W/\theta}}{Z(\theta)} P(W) \\
 \Rightarrow P(W) &= e^{W/\theta} Z(\theta) P_\theta(W).
 \end{aligned} \tag{9}$$

Thus, the target distribution $P(W)$ can be estimated from $P_\theta(W)$ at finite temperatures, up to a normalization constant $Z(\theta)$. Note that for a finite sample of graphs, each selected temperature θ will lead to an observation of values of W within a specific range and will therefore allow us to estimate $P(W)$ there. A bias toward smaller (larger) values of W is obtained when positive (negative) temperatures are used. In both cases, temperatures of large absolute $|\theta|$ 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. Thus, performing the simulations for a suitably chosen set of temperatures allows us to estimate $P(W)$ over a large range, possibly on its full support.

The normalization constants $Z(\theta)$ can be computed by comparing histograms from neighboring temperatures. In their overlapping region, relative normalization constants can be obtained by requiring that the rescaled histograms must agree within error bars. This means the histograms are “glued” together. The range of covered W values can be extended iteratively by choosing additional suitable temperatures and gluing the resulting histograms to each other. For details, see Refs [31, 36].

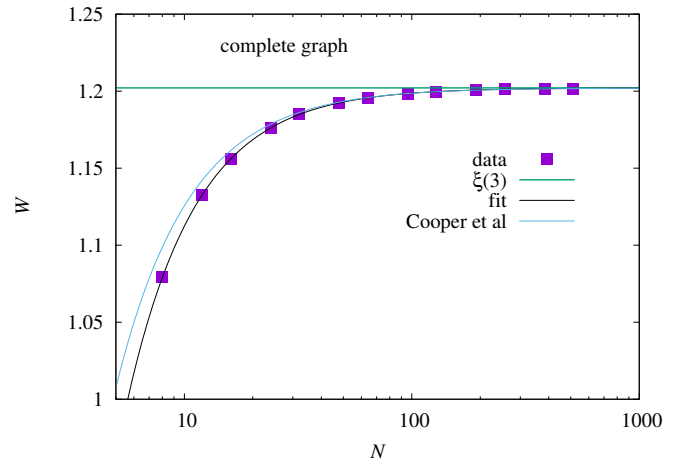


FIG. 2. Average spanning tree weight W as a function of the number N of nodes for the complete graph. The horizontal line indicates the limiting value obtained analytically, see Eq. (3), while the upper solid curve prerepresents Eq. (11). The lower solid curve indicates a fit to Eq. (10).

IV. RESULTS

We have performed simulations for complete graphs and connected ER graphs with varying connectivity c for different number N of nodes in the range $N \leq 512$. For both ensembles, the weights are uniformly distributed in $[0, 1]$. First, we consider the results for the complete graphs; in the second subsection, the ER case.

A. Complete graphs

The average spanning-tree weight for the complete graph is shown as a function of the graph size in Fig. 2. A convergence to the analytical result Eq. (3) is well visible, confirming our simulations. Note that there exist exact results [43] for small graphs $N \leq 9$. The result, $\frac{199462271}{184848378} \approx 1.07906$, agrees to the five leading digits with our simulation data. We have fitted a power law function

$$W_{\text{avg}}(N) = W_{\text{avg}}^\infty + bN^{-c}, \tag{10}$$

resulting in $W_{\text{avg}}^\infty = 1.20211(2)$, $b = -2.349(5)$, and $c = 1.418(1)$. As visible from the figure, the finite-size dependence of the weight follows the power-law very well. On the other hand, it has been predicted by Cooper et al. [29] that the finite corrections have the form

$$W(N) = \zeta(3) + c_1 N^{-1} + c_2 N^{-4/3} + \dots, \tag{11}$$

where $c_1 = 0.0384956\dots$, $c_2 = -1.7295$, and the $\dots$ in the equation refers to even faster decaying correction terms. We have included this function without a free parameter in Fig. 2. Here the agreement is worse, showing that the additional corrections are significant.

Next, we consider the full distribution $P(W)$. In Fig. 3 it is shown for the complete graph with $N = 256$ nodes.

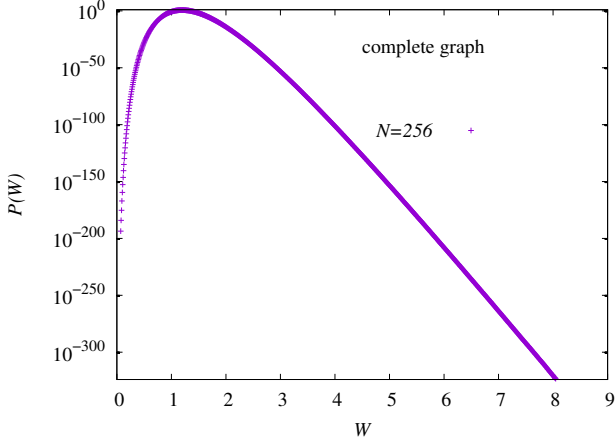


FIG. 3. Distribution of spanning tree weight W for the complete graph with $N = 256$ nodes.

By using the large-deviation approach, we are able to reach the tails of the distributions to values of the probability density as small as 10^{-300} . Since $W > 0$, we tried to fit the data to a generalized extreme-value distribution with shape parameter $\xi > 0$, but the fit did not converge. Still, the right tail resembles an exponential, but is slightly bent. A fit of the right tail to a stretched exponential $e^{-\lambda(W-W_0)^\beta}$ (not shown) yields an exponent $\beta = 1.22(2)$.

Next, we analyze the rate function Eq. (6), which is shown for different number N of nodes in Fig. 4. Note that for $N = 512$ for the far tail of the distribution, the MCMC simulations do not fully equilibrate. Thus, we cannot use this part of the data, and we are restricted here to somewhat smaller values of W . While for small values of N some finite size effects are visible, the curves, at least for values of W up to about 6, collapse onto each other for $N \geq 128$. This means that the convergence of the empirical rate function to a limiting rate function is compatible with the data. Still, we cannot say much about the shape, but the compatibility of the tail with a stretched exponential means that here the right part is compatible with a power law

$$\tilde{\phi}(W) = l(W - W_0)^\beta \quad (12)$$

with the same exponent $\beta = 1.22(2)$ as mentioned above.

Now, we want to understand what makes rare graph realizations exhibit particular large or small MST weights W . For the complete graph, the graph structure is fixed, so it must be due to the specific edge weights, i.e., the $N(N-1)/2$ weights of each graph G . For this purpose, we have stored, here for the largest considered size $N = 512$, for each statistically independent graph $G = (V, E)$ encountered in the MCMC simulation, the mean $W_G^{\text{mean}} = \frac{2}{N(N-1)} \sum_{\{i,j\} \in V} w_{i,j}$ weight. Also, we have for each node determined the minimum weight among all edges that are incident to the node. The empirical average over all nodes we denote as W_G^{min} . Since

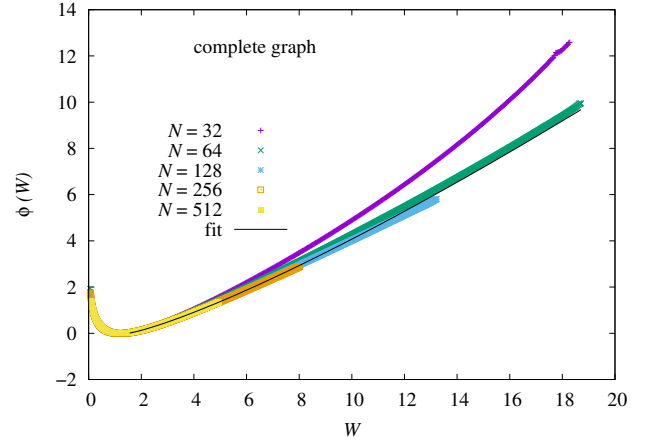


FIG. 4. Rate functions $\phi(W)$ for the complete graph for different values of the number N of nodes. The line shows the result of a fit of the data for $N = 256$ to a power law Eq. (12).

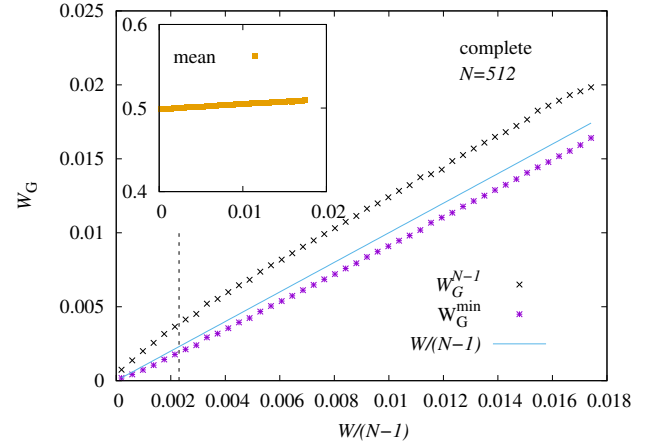


FIG. 5. Average of W_G^{min} , W_G^{N-1} , and (inset) W_G^{mean} , conditioned to the value W of the MST of G , respectively. The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.002$ indicates the typical edge weight of an MST, corresponding to $P(W)$ exhibiting a peak near $W = 1$ in Fig. 3 and $N = 512$.

each node is connected to all other nodes by at least one edge, $(N-1)W_G^{\text{min}}$ forms a lower bound and also a good approximation of W . In a similar way, since each spanning tree consists of $N-1$ edges, we have stored the weight W_G^{N-1} of the $(N-1)$ 'th largest weight in each graph G . This might serve as a good upper approximation, but not a strict upper bound. Finally, we also store the empirical variance W_G^{var} of the weights of each graph to see whether all edge weights get concentrated around some values.

We analyze now these quantities as a function of the MST W of each graph G . For this purpose, we show the averages of W_G^{min} , W_G^{N-1} , W_G^{mean} conditioned to $W = W(G)$ in Fig. 5, as a function, for better com-

parison, of $W/(N-1)$, i.e., the average weight of an edge in the MST. One sees a strong correlation of W with $W_G^{\min}$ and W_G^{N-1} , and indeed $W/(N-1)$ lies well between them. The observed linear correlation coefficients are 0.9997 and 0.9971, respectively. A scatter plot (not shown) reveals that the values of $W_G^{\min}$ and W_G^{N-1} are concentrated near the typical values, conditioned to W . For just a typical random graph, the MST weight is about 0.002, indicated by the vertical line in the figure. The $N-1$ smallest edge weights are below a weight of 0.004. This means that basically the smallest weights contained in the graph actually contribute to the MST, although the edges in the MST cannot be chosen independently. On the other hand, each complete graph contains many more edges, where most edges do not contribute to the MST. As the inset of Fig. 5 shows, a typical edge weight is always near the average weight of 0.5, but one can indeed observe a clear but weaker correlation, with a linear correlation coefficient of 0.9946. The mean conditioned to large MST weights is 0.509, only slightly larger than the unconditional mean of 0.5. A similar weak effect can be seen for the empirical variance (not shown as a figure), where we observed a clear linear correlation coefficient of -0.9637, which results from a small decay of the variance from 0.0839 for $W/(N-1) = 0.0005$ to a variance of 0.0805 for $W/(N-1) = 0.018$.

Note that such a large range for the values of $W/(N-1) \in [0, 0.018]$ can only be accessed with a large-deviation approach. With a direct sampling of, say, 10^6 samples, one would be able to observe values and therefore measure correlations for about $W/(N-1) \in [0.002, 0.003]$.

B. Erdős-Rényi graphs

For the connected ER ensemble, the behavior can be richer, since not only the edge weights but also the graph structure varies. We first analyse the mean minimum spanning tree weight, which we denote by W as well for simplicity, as a function of the number N of nodes. Since the connected ER ensemble is sparse, the number of available edges to construct an MST is much smaller, of the order $O(N)$. Thus, doubling the size of the graph will basically double the weight of the MST; hence, W , being the sum of $N-1$ weights, should scale linearly with $N-1$. This is well visible in Fig. 6. We have fitted the data for each value of c to a linear function $W = a(N-1)$. For small values of c , the original ER graphs contain $cN/2$ edges, which is smaller than $N-1$, the number of edges required to form a spanning tree. Hence, the connected ER ensemble will usually contain $N-1$ edges or a few more, despite the small value of c . Thus, almost all edges are contained in the spanning tree, which means that, because the average edge weight is $1/2$, W will be about $(N-1)/2$, i.e., $a = 1/2$. As shown in the inset of Fig. 6, the obtained values of a are indeed close to $1/2$ for small values of c . Note that the value $c = 1$, which is the percolation transition point of the standard ER ensemble,

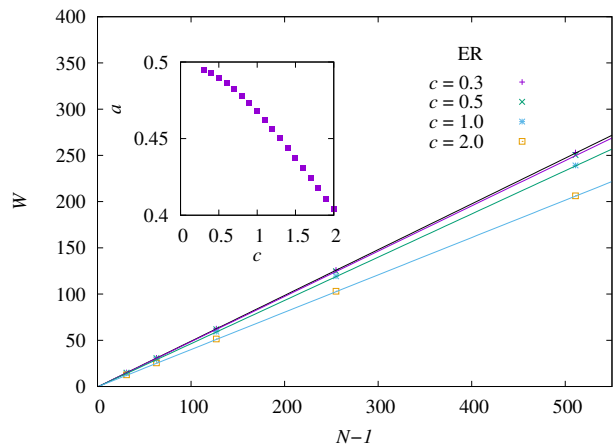


FIG. 6. Average spanning tree weight W as a function of the number N of nodes for the connected ER ensemble. The lines indicate fits to a linear function with slope a . The inset shows the slope a as a function of the connectivity c .

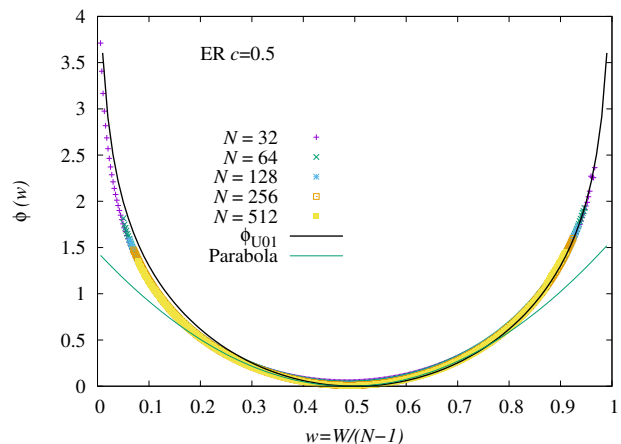


FIG. 7. Rate functions $\phi(W)$ for the connected ER graph with $c = 0.5$ for different values of the number N of nodes. The black line shows the rate function ϕ_{U01} for the average of $N-1$ iid random numbers distributed uniformly in $[0, 1]$. The green line shows the result of a fit of the data for $N = 512$ in the high-probability region to a parabola Eq. (13).

does not show any special behavior here.

Next, to investigate the actual distributions, we consider the rate functions $\phi(w)$ with $w = W/(N-1)$ being the spanning-tree weight per edge. The result for $c = 0.5$ is shown in Fig. 7. The empirical rate functions for different sizes collapse almost onto each other, so a convergence to a limiting rate function in the limit of large graph sizes can be expected, and the large-deviation property is fulfilled.

As mentioned above, the restriction to connected graphs means we expect that here there are basically graphs with $N-1$ edges for $c = 0.5$ such that almost all edges contribute to the spanning tree. Thus, the spanning tree weight is very simply almost given by the

sum of $N - 1$ random numbers that are iid in $[0, 1]$, i.e., $W/(N-1)$ is almost the average weight of these numbers.

For a first simple comparison, we consider the region of larger probabilities. From the central limit theorem, we therefore expect that the typical part of the distribution is Gaussian $\exp(-(w - w_p)^2/(2\sigma^2))$, centered about the mean value 0.5 and with variance $\sigma^2 = 1/(12(N-1))$, with $1/12$ being the variance of the uniform distribution in $[0, 1]$. This is reflected visibly in the shape of the rate function. Indeed, a fit to a parabola

$$\tilde{\phi}(w) = l(w - w_0)^2 \quad (13)$$

in the range $w \in [0.3, 0.7]$ yields a good fit; see Fig. 7. Here we find $w_0 = 0.489(1)$ and $l = 6.08$. The differences to the expected values $w_0 = 0.5$ and $l = 12/2$, respectively, are likely due to the finite size of the system. E.g., the variance behaves like $1/12(N-1)$, but for the rate function we divide by N instead of $N-1$. The result of the fit confirms the expectations about the shape of minimum spanning trees in this low-connectivity region.

To compare also the tails, the rate function of the average of $N-1$ iid $U(0, 1)$ random numbers can be calculated. This is the correct model for a tree, where all edges contribute to the spanning tree; see above. To obtain the rate function of a random variable (RV) S_N using the Gärtner-Ellis Theorem [33], we need the scaled cumulant generating function $\lambda(k) = \lim_{N \rightarrow \infty} \ln \frac{1}{N} E[e^{NkS_N}]$, where the expectation $E[\cdot]$ is with respect to the N random variables $\{X_i\}$ contributing to S_N . For the average $S_N = \frac{1}{N} \sum_i X_i$, where all RVs X_i are the same X , we have $\lambda(k) = \lim_{N \rightarrow \infty} \ln \frac{1}{N} E[e^{k \sum_i X_i}] = \ln E[e^{kX}]_X$, where $E[\cdot]_X$ is the average of the random RV X . For the uniform distribution $U(0, 1)$, this yields $\lambda(k) = \ln \int_0^1 dx e^{kx} = \ln((e^k - 1)/k)$. By using the Gärtner-Ellis Theorem, one has $\phi(s) = \sup_k (ks - \lambda(k))$. We have performed this maximization numerically for many values of $s \in [0, 1]$; the result, which we denote as ϕ_{U01} , is also shown in Fig. 7. A very good agreement with the numerical data is visible; the small differences are likely due to the fact that also graphs with more than $N-1$ edges contribute, in particular for small values of w .

We have also considered $c = 0.7$ and $c = 1$ in the (about) non-percolating region; here the behavior is very similar.

For larger values of c , the shape of the distribution changes. In Fig. 8, we show the rate functions for the largest value of $c = 5$ we have considered. The function is clearly not symmetric. For larger values of N , no significant finite-size dependence is visible, indicating that a convergence $N \rightarrow \infty$ exists and therefore the large-deviation principle holds also here. In the intermediate regime, $w \in [0.3, 0.7]$, we have fitted a power law Eq. (12) yielding an offset $w_0 = 0.19$ and an exponent $\beta = 1.92(2)$, which is close to but significantly different from 2. This matches the fact that for small values of c , a parabolic shape $\beta = 2$ was found, while for the complete graphs, corresponding to $c \rightarrow \infty$, in Sec. IV A, a much smaller exponent was obtained.

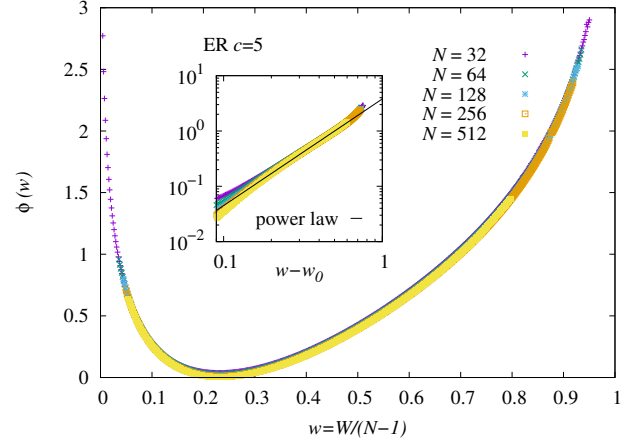


FIG. 8. Rate functions $\phi(W)$ for the connected ER graph with $c = 5$ for different values of the number N of nodes. In the inset, the same data is shown with a double log scale. The line shows the result of a fit of the data for $N = 512$ to a power law Eq. (12).

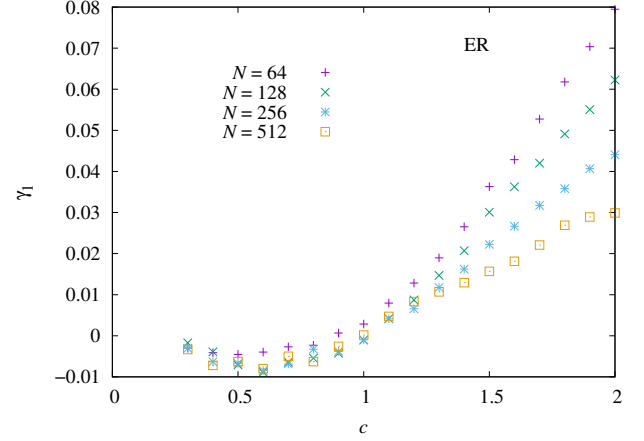


FIG. 9. Skewness γ_1 as a function of the connectivity parameter c for the connected ER ensemble, for different system sizes N .

Note that for larger values of w , the data bends up; thus, a growth with a larger exponent is possible. Such change of the behavior of rate functions, sometimes termed “phase transitions”, has been observed also for the rate functions of other random systems like those described by the Kardar-Parisi-Zhang equation [44].

To investigate more how the shape of the distribution changes when varying the parameter c , we analyse the skewness

$$\gamma_1 = \frac{\langle (W - \langle W \rangle)^3 \rangle}{\sigma^3}, \quad (14)$$

where σ^2 is the variance $\langle (W - \langle W \rangle)^2 \rangle$. The result is shown for different number N of nodes in Fig. 9. For values $c \leq 1$, the skewness is very close to zero, indicating a very symmetric distribution. Only from $c = 1$ onwards

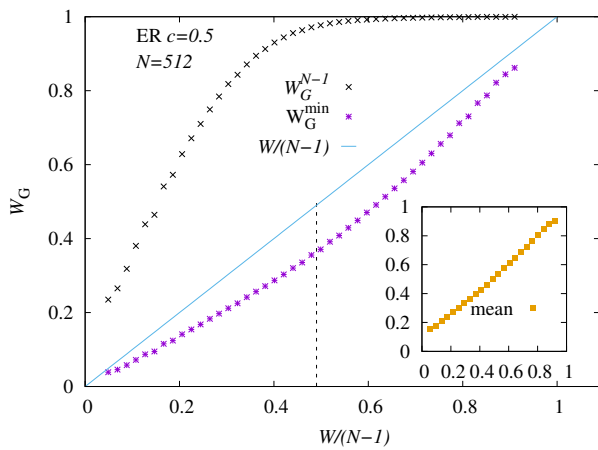


FIG. 10. Average of W_G^{N-1} , W_G^{min} , and (inset) W_G^{mean} , conditioned to the value W of the MST of G , respectively, for ER random graphs with $c = 0.5$. The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.49$ indicates the typical edge weight in an MST for this value of c .

a gradual increase is visible. Thus, for values of c that correspond to the percolating phase in the standard ER ensemble, the MST weight of the connected ER ensemble exhibits a significant asymmetric distribution.

Finally, we investigate how the structure of the graphs influences the MST weight. Similar to the results for the complete graph, we show in Fig. 10 for the case $c = 0.5$ the average graph weights W_G^{min} , W_G^{N-1} , and W_G^{mean} conditioned to a value of the weight $W/(N-1)$ per MST edge. Again we see that the average edge weight is between W_G^{min} and W_G^{N-1} . Compared to the complete graph, see Fig. 5, the upper value W_G^{N-1} is farther away from $W/(N-1)$. The reason is that for the connected ER graphs of such a small value of c , as discussed above, almost all edges contribute to the MST. Thus, W_G^{N-1} will be close to the upper limit of $w = 1$ for typical graphs and for atypical graphs with larger MST weights, as visible in Fig. 10. Only for atypical small values of W , W_G^{N-1} is likely smaller. Also, since almost all edges contribute to the MST, the average weight $W/(N-1)$ of the MST is very similar to the average edge weight W_G^{mean} in the graph, as visible in the inset.

The corresponding results for $c = 5$ are displayed in Fig. 11. Since more edges are present in the graph, but still the graph is sparse, the results look somehow in between the results for $c = 0.5$ and for the complete graph. Again, here the sparseness leads to a high influence of the mean edge weight W_G^{mean} of G on the MST weight W . We do not display results for the other values of c that we have studied in the same way, because they are intermediate between the two cases presented so far.

The fluctuations of the edge weights of the graph are not the only reason for particular small or large MST weights. In the ER ensemble, also the number of edges fluctuates. For any value of c , although the typical num-

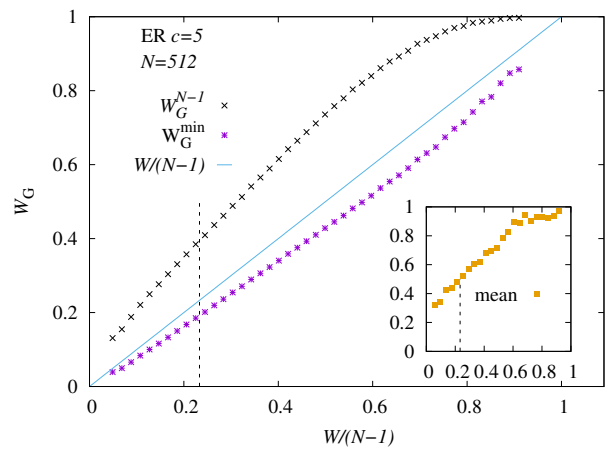


FIG. 11. Average of W_G^{N-1} , W_G^{min} , and (inset) W_G^{mean} , conditioned to the value W of the MST of G , respectively, for ER random graphs with $c = 5$. The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.23$ indicates the typical edge weight in an MST for this value of c .

ber of edges is $cN/2$, there will be much denser and much sparser graphs. In general, the more edges are available, the smaller the MST weight will be, because there are more weights to minimize over. This is visible in Fig. 12, where the average number M of edges conditioned to the MST edge weight $W/(N-1)$ per edge is shown for $N = 512$ and several values of c . For large values of W , the number of edges is very small and converges to the lower limit of $N-1$, since the graphs have to be connected. The dependency becomes more pronounced with increasing value of c . This is reasonable because for smaller values of c also typical connected graphs contain only few edges. Still, since the value $W/(N-1)$ is larger than the mean weight 0.5, these graphs are not only rare with respect to their structure, but also rare with respect to the edge weights, as discussed above. For small values of $W/(N-1)$, the number of edges is conversely atypically larger, so the minimization can lead to lower MST weights.

In summary, both atypical graph structure and atypical edge weights come together for weighted connected ER graphs which exhibit atypical MST weights.

V. DISCUSSION

We have studied the distribution of the weight of minimum-spanning trees for complete graphs as well as for connected ER random graphs with connectivity c . With respect to simple properties as the expectation value, our results are compatible with previous analytical studies. By applying large-deviation approaches, we are able to obtain the distributions even in the tails, down to probability densities as small as 10^{-300} . For both ensembles, the results indicate that the large-deviation

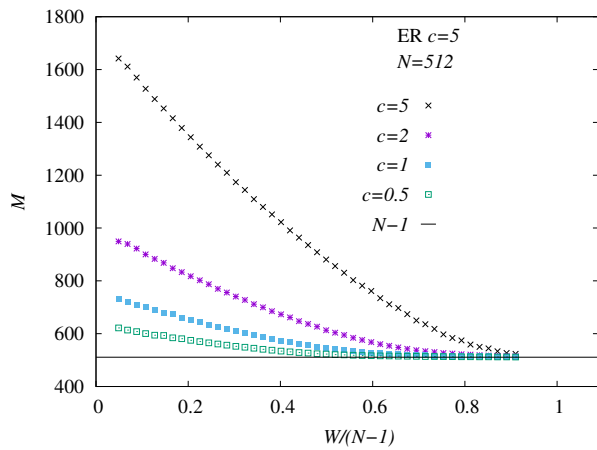


FIG. 12. Average of the number M of edges conditioned to the value W of the MST of G , shown as a function of $W/(N-1)$ for $N = 512$ and several values of c . The horizontal line indicates the minimum number $N - 1$ of edges needed to create a connected graph.

property is fulfilled, i.e., away from the typical values the probabilities become exponentially small with exponent linearly decreasing in N . For the complete graph, the behavior of the rate function of the right tails is compatible with a power-law behavior. Since for complete graphs, the graph structure is fixed, rare MST weights W are created by rare assignments of edge weights. But only the lowest edge weights have to be smaller or larger than typical to determine W , the overall mean weight in a graph, is only weakly affected.

For the ER case, for small values of c , the distributions

converge in the center to Gaussians, which makes sense as many independent weights contribute. Also the tails of the distribution are described well by a rate function describing the sum of $N - 1$ independently drawn numbers in $[0, 1]$. Beyond the value $c = 1$, which is the percolation threshold for the standard ER ensemble, the distributions become asymmetric and non-trivial tails develop. Here, the values of edge weights and the actual number of edges influence the MST weight and determine whether it is typical, atypically large or atypically small.

Clearly, one could analyse more properties, e.g., whether the variance of the node degrees also play a, likely minor, role. For future studies, one also could consider other ensembles like scale-free graphs, finite-dimensional disordered lattices, or general geometric graphs embedded in finite-dimensional spaces.

Also, it would be of particular interest if our work motivates other researchers to study the distribution of weights by analytical approaches, e.g., to obtain the rate functions and to investigate whether there is a “critical” value w_c where the rate function changes its behavior for the ER case.

ACKNOWLEDGMENTS

MS thanks Abolfazl Ramezanzpour for helpful discussions and assistance with debugging the code. The simulations were performed at the HPC cluster ROSA, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Program (INST 184/225-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

-
- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [2] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
 - [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, *Phys. Rep.* **424**, 175 (2006).
 - [4] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks: from biological nets to the Internet and WWW* (Oxford Univ. Press, 2006).
 - [5] M. Newman, *Networks: an Introduction* (Oxford University Press, 2010).
 - [6] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2012).
 - [7] R. L. Graham and P. Hell, *Ann. Hist. Comput.* **7**, 43 (1985).
 - [8] R. C. Prim, *The Bell System Technical Journal* **36**, 1389 (1957).
 - [9] L. L. Cavalli-Sforza and A. W. F. Edwards, *Evolution* **21**, 550 (1967).
 - [10] J. C. Gower and G. J. S. Ross, *J. Royal Stat. Soc. Ser. C* **18**, 54 (1969).
 - [11] J. Wu, X. Li, L. Jiao, X. Wang, and B. Sun, *Physica A* **392**, 2265 (2013).
 - [12] R. E. Osteen and P. P. Lin, *SIAM J. Comput.* **3**, 23 (1974).
 - [13] Z. Zhang, B. Wu, and Y. Lin, *Physica A* **391**, 3342 (2012).
 - [14] T. Li and W. Yan, *Physica A* **536**, 120877 (2019).
 - [15] C. Dussert, G. Rasigni, M. Rasigni, J. Palmari, and A. Llebaria, *Phys. Rev. B* **34**, 3528 (1986).
 - [16] D. V. Kvitarev, S. Lübeck, P. Grassberger, and V. B. Priezzhev, *Phys. Rev. E* **61**, 81 (2000).
 - [17] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, *Phys. Rev. Lett.* **96**, 148702 (2006).
 - [18] D.-H. Kim, J. D. Noh, and H. Jeong, *Phys. Rev. E* **70**, 046126 (2004).
 - [19] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, *Phys. Rev. E* **68**, 046130 (2003).
 - [20] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Phys. Rev. E* **68**, 056110 (2003).
 - [21] T. S. Jackson and N. Read, *Phys. Rev. E* **81**, 021130 (2010).
 - [22] T. S. Jackson and N. Read, *Phys. Rev. E* **81**, 021131 (2010).
 - [23] S. M. Sweeney and A. A. Middleton, *Phys. Rev. E* **88**, 032129 (2013).

- [24] T. I. Fenner and A. M. Frieze, *Combinatorica* **2**, 347 (1982).
- [25] A. Frieze, *Discrete Applied Mathematics* **10**, 47 (1985).
- [26] C. McDiarmid, *London Math. Soc. Lect. Note Ser.* **141**, 148 (2000).
- [27] S. Janson, *Rand. Struc. Algor.* **7**, 337 (1995).
- [28] A. D. Flaxman, *Electr. J. Comb.* **14**, N3 (2007).
- [29] C. Cooper, A. Frieze, N. Ince, S. Janson, and J. Spencer, *Comb. Prob. Comp.* **25**, 89–107 (2016).
- [30] J. A. Bucklew, *Introduction to rare event simulation* (Springer-Verlag, New York, 2004).
- [31] A. K. Hartmann, *Eur. Phys. J. B* **84**, 627 (2011).
- [32] F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2000).
- [33] H. Touchette, *Physics Reports* **478**, 1 (2009).
- [34] P. Erdős and A. Rényi, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17 (1960).
- [35] T. H. Cormen, S. Clifford, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms* (MIT Press, Cambridge (USA), 2001).
- [36] A. K. Hartmann, *Phys. Rev. E* **65**, 056102 (2002).
- [37] A. K. Hartmann, *SciPost Phys. Lect. Notes*, 100 (2025).
- [38] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [39] W. K. Hastings, *Biometrika* **57**, 97 (1970).
- [40] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
- [41] D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
- [42] The data is publicly available in the repository DARE of the University of Oldenburg [45].
- [43] J. M. Steele, in *Mathematics and Computer Science II*, edited by B. Chauvin, P. Flajolet, D. Gardy, and A. Mokkadem (Birkhäuser Basel, Basel, 2002) pp. 223–245.
- [44] A. K. Hartmann, P. L. Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, *Europhys. Lett.* **121**, 67004 (2018).
- [45] M. Sarikhani and A. K. Hartmann, “Data and gnuplot plot files,” DARE public repository, DOI: 10.57782/BKJRBF (2025).