

Large deviation properties of minimum spanning trees for random graphs

Mahdi Sarikhani ¹ and Alexander K. Hartmann ²

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

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



(Received 23 November 2025; accepted 22 July 2026; published 27 August 2026)

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 on being connected. By using large-deviation Markov-chain sampling, we can 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.

DOI: [10.1103/xc5g-t9dj](https://doi.org/10.1103/xc5g-t9dj)

I. INTRODUCTION

Graphs and networks [1–6] are ubiquitous in modeling systems such as chemical molecules, disordered materials, energy grids, spread of diseases, and social networks. Graphs can be analyzed in many ways, like the calculation of 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 are concerned with the fundamental analysis of ensembles of random graphs, which has been addressed to some extent in the literature. For example, for the distribution $P(S)$ of the size S of the largest component of random graphs, there exist analytical [7] and numerical results [8,9] covering the distribution over the full support, i.e., including the large-deviation properties, where the probabilities are exponentially small. In a similar way, the distributions of the graph diameter of random graphs have been studied analytically [10] and numerically [11]. There are also problems where, for given graphs, optimized subgraphs are constructed. For example, for graph matchings, subsets of pairs of connected nodes are considered, with a weight W given by the weights of the edges connecting the pairs. Here, results are known concerning the expectation value of the weight [12–14]. Also, the large-deviation behavior of the distribution $P(W)$ for complete graphs has been obtained from replica and cavity calculations [15]. With respect to numerical results, the high-probability part of the distribution $P(W)$ has been addressed [16].

Here, we consider a different type of subgraph, namely spanning trees, which are subsets of edges of a graph such that all vertices are connected to each other by exactly one path.

Being a tree, for N nodes the spanning tree has $N - 1$ edges. 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 [17] 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 [18], reconstruct evolutionary trees [19], cluster data [20], detect communities [21], determine patterns in images [22], and many more. In statistical physics, among several other applications, spanning trees have been analyzed by counting [23] or enumerating [24] them. Also, spanning trees have been used to analyze atomic deposition structures on surfaces [25], avalanches in sandpile models [26], transport [27] and other technical networks [28], stock markets [29,30], and fractal properties of percolation clusters [31–33]. Finally, spanning trees have also attracted the attention of mathematicians, where we use some of the work [34–40] for comparison with our results below. In particular, the behavior of the distribution $P(W)$ of weights has been addressed in a few cases [36,38], yet without explicitly calculating the distribution or the related rate function. Furthermore, we are not aware of numerical studies of $P(W)$ addressing the tails of the distribution.

In this work, we numerically study the distribution $P(W)$ of the weights of minimum spanning trees for two random ensembles, where we address the distribution over a large range of its support, down to exponentially small probabilities. The first ensemble 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 [35]; it is also known that the probability of deviating 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 on being connected; again, the weights are iid uniform in $[0,1]$. Here, to the best of our

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

knowledge, nothing is known about the distribution $P(W)$ at all, not even numerically. The approach we use utilizes a Markov-chain Monte Carlo large-deviation sampling [8,41] 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 its support. This ability is useful to verify the so-called large-deviation principle [42,43]. 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 that are relevant to our work. Next, in Sec. III, 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 all nodes in a graph 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

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

of the weights of the edges of the tree. The minimum spanning tree (MST) of graph G is a spanning tree whose weight $W(T)$ is minimal; we denote the MST's weight for this given graph as

$$W_G = \min_T W(T). \quad (2)$$

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.

Here, we do not consider just single specific given graphs, but ensembles of fixed or random connected graphs with, in any case, randomly chosen edge weights. This 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 of obtaining a certain graph G , which is meant to include the edge weights,

by $Q(G)$, one has

$$P(W) = \sum_G \delta_{W_G, W} Q(G). \quad (3)$$

Note that, since the weights are continuous variables, the $\sum_G$ actually also involves 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 iid uniform 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 [35] in the limit of infinite graph size $N \rightarrow \infty$ as

$$\langle W \rangle_N \rightarrow \zeta(3)/D \text{ where } \zeta(3) = \sum_{k=1}^{\infty} 1/k^3 = 1.202 \dots \quad (4)$$

This actually holds for arbitrary distribution functions $F(w_{ij})$ of non-negative edge weights, where $D = F'(0) > 0$ is the probability density at 0. In the present case with a uniform distribution on $[0, 1]$, we have $D = 1$. This is compatible with a previous result [34] that $\langle W \rangle_N \leq 2(1 + \log n/n)$. Recently, the finite-size corrections of $\langle W \rangle_N$ have also been calculated [40]. With respect to the distribution, the typical part becomes Gaussian in the limit of large N with variance about 1.6857 [37]. Concerning the tails, which are our main concern here, it has been shown [36,38] that the deviations from the typical value decrease exponentially with the graph size N , i.e., fulfill the large-deviation principle; see the end of this section.

Second, we study a special ER [44] 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 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 with $p = c/N$, where 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 on being connected. Technically, depth-first search [45] is used to verify connectedness. If a graph is not connected, it is discarded in the analysis. Note that for small values of c , typical ER graphs are not connected. Here, the Markov chain Monte Carlo (MCMC) approach presented below is more efficient for generating connected ER random graphs.

In large-deviation theory [42,43], 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

$$P(W) \simeq e^{-N\Phi(w)}, \quad (5)$$

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

that

$$P(W) = e^{-N\Phi(w)+o(N)}, \quad (N \rightarrow \infty). \quad (6)$$

The so-called large-deviation principle holds if, loosely speaking, the distribution has the form of Eq. (5) and the empirical rate function

$$\Phi_N(w) \equiv -\frac{1}{N} \ln P(wN), \quad (7)$$

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 [18,45]. 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 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 all other edges incident to l , 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 $O(M \ln N)$.

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 $Q(G)$, which includes the edge weights and, for the ER ensemble, the structure of the graph. 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 with 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, including very small probability densities of the order of 10^{-100} , we use a different approach [8,41,46]. Here, we display only the main ingredients of the algorithm; for a pedagogical introduction, see Ref. [47]. 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 that controls the weight. More precisely, using the natural probability density $Q(G)$ for a graph, the modified

density is given by

$$Q_\theta(G) = \frac{1}{Z(\theta)} Q(G) \exp(-W_G/\theta), \quad (8)$$

where $Z(\theta)$ is a suitable normalization that involves the integral over all possible assignments of the edge weights and, for ER graphs, the sum over all possible graphs. We denote this combined sum and integral for short as $Z(\theta) = \sum_G Q(G) \exp(-W_G/\theta)$.

Now, one needs to sample graphs G according to the modified density $Q_\theta(G)$. Since we are not aware of a method to do this directly, we generate Markov chains by using standard MCMC simulations. The Markov chain is given by graphs $G(0), G(1), \dots$, i.e., the current configuration at step t is given by an instance $G(t)$ of a graph, with weights. We use the Metropolis-Hastings algorithm [48–51], which is applied here as follows. At each step t , a candidate graph G^* is created from the current graph $G(t)$ by first copying $G(t)$. Next, the candidate graph G^* is modified slightly. Different techniques are used for the two different graph ensembles:

(1) 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 $[0, 1]$.

This weight change is performed in total for a certain number n_c of edges; for details on how n_c is chosen, see Sec. IV.

(2) 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 $[0, 1]$.

Here, for each graph, only one change is performed because each change involves $O(N)$ potential edges [52].

If, in the ER case, the candidate graph G^* is not connected, it is immediately rejected. This means that the initial graphs also need to be connected. Therefore, most of our simulations begin with a complete graph, even for the ER ensemble.

Note that these steps for generating the candidate graph G^* are done with respect to the original statistics of the two ensembles. This ensures that the sampling involves the factor $Q(G)$ in Eq. (8).

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

$$p = \min\{1, e^{-[W_{G^*} - W_{G(t)}]/\theta}\}. \quad (9)$$

In case the candidate graph is rejected, the current graph is kept, i.e., $G(t+1) = G(t)$. Thus, this Metropolis step ensures that the bias factor $\exp(-W_G/\theta)$ is included in the MCMC sampling. Since the construction of G^* is according to $Q(G)$, this completes the sampling according to Eq. (8).

In practice, 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) MST weights than the typical one.

By construction, the algorithm fulfills detailed balance. It is also ergodic, since within M or N steps, respectively, each

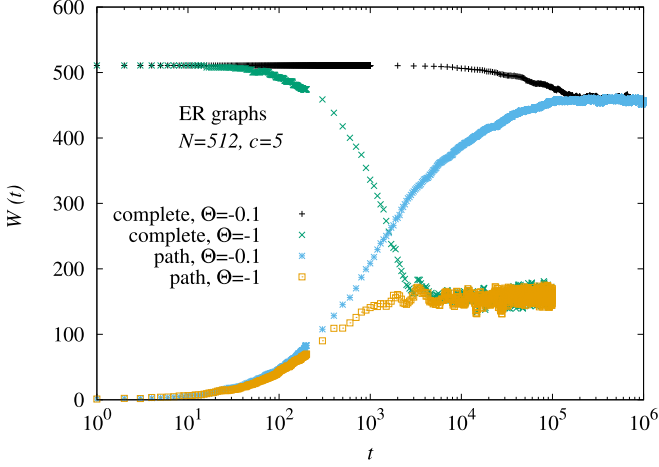


FIG. 1. Equilibration of the Markov chain: time series $W(t)$ for ER graphs with $N = 512$ and $c = 5$ for two different initial configurations, a complete graph (all $w_{ij} = 1$) and a path graph (all w_{ij} near 0).

possible graph may be constructed. The graphs $G(t)$ at the beginning of the Markov chain will resemble the initial graph $G(0)$ and not represent typical graphs with respect to $Q_\theta(G)$. Thus, one has to discard the first part of the Markov chain, i.e., equilibrate the simulation. To check equilibration, we monitor the time series $W(t)$. The simplest idea is to consider the time series equilibrated if no systematic trend is visible anymore, only fluctuations. This is particularly well visible if one starts two independent runs with initial configurations that lead to very different initial weights $W(0)$, 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 path graph of 511 edges where all weights are randomly close to zero. Here, $W(0)$ is about 0.2. The resulting time series $W(t)$ are shown [53] in Fig. 1 for two different temperatures, $\theta = -0.1$ and $\theta = -1$. In the first case, equilibration takes some time, more than 10^5 steps, while for $\theta = -1$, equilibration is reached faster.

Another way to investigate the equilibration progress of the MCMC simulations is to calculate correlations with the initial configuration of the form

$$C(t) = \frac{\langle W(0)W(t) \rangle_0 - \langle W(0) \rangle_0 \langle W(t) \rangle_0}{\langle W(0)^2 \rangle_0 - \langle W(0) \rangle_0^2}, \quad (10)$$

where the average $\langle \dots \rangle_0$ is over different initial graphs and weights, and over different stochastic histories of the MCMC simulation. The result for ER random graphs with $N = 512$ and $c = 5$, averaged over 100 independent runs, is shown in Fig. 2. After some time, the measured weight is practically decorrelated from the initial one. Again, it is visible that for the temperature parameter $\theta = -1$ the Markov chain evolves faster compared to $\theta = -0.1$. Note that, when looking at $C(t)$, the timescales to reach $C(t) \approx 0$ are somewhat smaller than those needed to observe equilibration when looking directly at $W(t)$, as shown in Fig. 1.

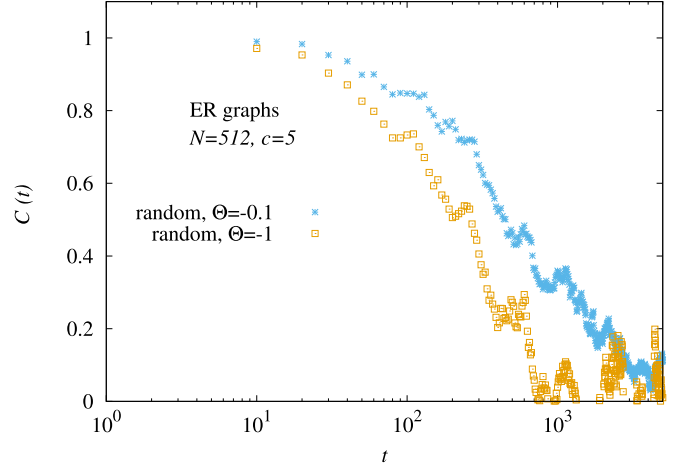


FIG. 2. Decorrelation of the Markov chain: time series $C(t)$ for ER graphs with $N = 512$ and $c = 5$ for random initial configurations, averaged over 100 runs.

Note that, for $\theta = -0.1$, the observed weights $W(t)$ are close to the maximum possible value $W = 511$, and exhibit only small fluctuations. Thus, for the simulations used to obtain the distributions, we attempted to reach such extreme values of the weight, i.e., the far tails, by suitably chosen temperature values.

For equilibrated Markov chains, in the limit of infinitely long, or at least long enough, simulation time, the distribution of graphs follows the density as given by Eq. (8).

The probability distribution for W at temperature θ is the quantity that is actually estimated when sampling the MST weight W by the MCMC simulations. It is given by a sum over all graph structures and an integral over all weight assignments leading to the MST weight $W_G = W$ (again using $\sum_G$ as an abbreviation),

$$\begin{aligned} P_\theta(W) &= \sum_G \delta_{W_G, W} Q_\theta(G) \\ &= \frac{1}{Z(\theta)} \sum_G \delta_{W_G, W} e^{-W_G/\theta} Q(G) \\ &= \frac{e^{-W/\theta}}{Z(\theta)} \sum_G \delta_{W_G, W} Q(G) \\ &= \frac{e^{-W/\theta}}{Z(\theta)} P(W) \\ \Rightarrow P(W) &= e^{W/\theta} Z(\theta) P_\theta(W). \end{aligned} \quad (11)$$

Thus, the target distribution $P(W)$ can be estimated from $P_\theta(W)$ at finite temperatures, up to the 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 with large absolute value $|\theta|$ cause sampling of the distribution close to its typical value, while temperatures with small absolute value are used to access the tails of the distribution. Thus, performing the simulations for

a suitably chosen set $\{\theta_1, \dots, \theta_{n_\theta}\}$ of n_θ temperatures allows us to estimate $P(W)$ over a large range, possibly over its full support.

The normalization constants $Z(\theta)$ can be computed by comparing histograms from neighboring temperatures. The basic idea is that Eq. (11) holds for arbitrary temperatures θ . Thus, for two temperatures θ_1 and θ_2 , one has

$$e^{W/\theta_1} Z(\theta_1) P_{\theta_1}(W) = e^{W/\theta_2} Z(\theta_2) P_{\theta_2}(W), \quad (12)$$

which allows one to determine the ratio $Z(\theta_1)/Z(\theta_2)$. In practice, this can be done if one has pairs of close-by temperatures chosen such that the obtained histograms overlap sufficiently, by a couple of bins. Since Eq. (12) is not fulfilled exactly by the data, the relative normalization constants $Z(\theta_1)/Z(\theta_2)$ can be obtained by requiring that the rescaled histograms for close-by temperatures exhibit a minimum squared difference for those bins of W where both histograms have sufficient statistics. 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 rescaled histograms to each other. For details, see Refs. [8,46].

IV. RESULTS

We have performed simulations for complete graphs and connected ER graphs with varying connectivity c for different numbers 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, we consider the ER graphs.

For both ensembles, we have empirically determined suitable sets of temperatures such that for neighboring temperatures the resulting histograms of W overlap in at least four bins. The temperatures with the lowest absolute value $|\theta|$ were chosen such that equilibration of the MCMC was possible within a few hours. More details are given in the corresponding subsections.

A. Complete graphs

For complete graphs, the number of temperatures for all sizes $N = 32, 64, 128, 256$, and 512 ranged from 22 to 24, i.e., did not change much with system size. For the sampling with basically no bias, we have chosen $\theta = 10^5$ for all sizes N . For positive temperatures, addressing the left tails of the distributions, the values were selected in the range $[0.1, 4]$. To address the right tails of $P(W)$, the temperature range for the smallest size $N = 32$ was $[-4, -1]$. Thus, $\theta = -1$ is the upper negative temperature we could equilibrate within a reasonable time. This value decreases with system size. For the largest size $N = 512$, only the range $[-4, -2]$ could be addressed. Note that the number n_c of weights that are changed in each MCMC step to generate a candidate graph G^* depends on both the temperature θ and the size N , where generally the number of changed entries grows with size N . We have chosen n_c such that about 50% of all candidate graphs were accepted, respectively. For example, for the largest size $N = 512$, we changed $n_c = 8192$ entries in each MCMC step at $\theta = 10^5$, while for $\theta = -2$, we changed only $n_c = 64$ entries.

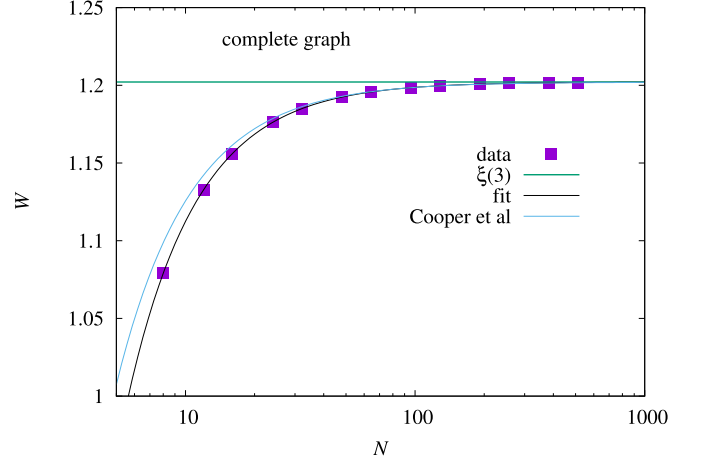


FIG. 3. Average spanning tree weight W as a function of the number N of nodes for the complete graph. The horizontal line indicates the limiting value obtained analytically; see Eq. (4), while the upper solid curve represents Eq. (14). The lower solid curve indicates a fit to Eq. (13).

The average spanning-tree weight for the complete graph is shown as a function of the graph size in Fig. 3. A convergence to the analytical result equation (4) is well visible, confirming our simulations. Note that there exist exact results [54], and generalizations to other edge-weight distributions [39], for small graphs $N \leq 9$. The result, $\frac{199462271}{184848378} \approx 1.07906$, agrees with the five leading digits of our simulation data. We have fitted a power-law function,

$$W_{\text{avg}}(N) = W_{\text{avg}}^{\infty} + bN^{-c}, \quad (13)$$

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. It has been predicted by Cooper *et al.* [40] that the finite-size behavior of the expectation value $W(N)$ exhibits the form

$$W(N) = \zeta(3) + \frac{c_1}{N} + \frac{c_2 + o(1)}{N^{4/3}}. \quad (14)$$

Thus, the first two corrections are of order N^{-1} and $N^{-4/3}$, and there are higher-order corrections that decrease more quickly than $N^{-4/3}$, but it is unknown how fast. The constants were determined [40] as $c_1 = 0.0384956 \dots$ and $c_2 = -1.7295$. We have included this function without a free parameter in Fig. 3. Here, the agreement is worse, showing that the additional corrections $o(1)N^{-4/3}$ are significant.

Next, we consider the full distribution $P(W)$. In Fig. 4, it is shown for the complete graph with $N = 256$ nodes. By using the large-deviation approach, we are able to reach the tails of the distribution to values of the probability density as small as 10^{-300} . Because the distribution is supported only on positive values of W , we attempted to fit it using 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 $\beta = 1.22(2)$.

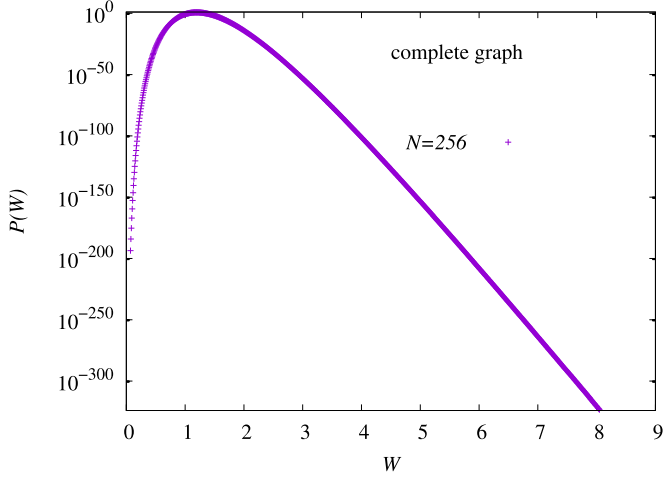


FIG. 4. Distribution of spanning tree weight W for the complete graph with $N = 256$ nodes.

Next, we analyze the rate function equation (7), which is shown for different numbers N of nodes in Fig. 5. 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 the right part is compatible with a power law,

$$\tilde{\phi}(W) = l(W - W_0)^\beta, \quad (15)$$

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

Now, we want to understand what makes rare graph realizations exhibit particularly large or small MST weights W . For the complete graph, the graph structure is fixed, so it must be

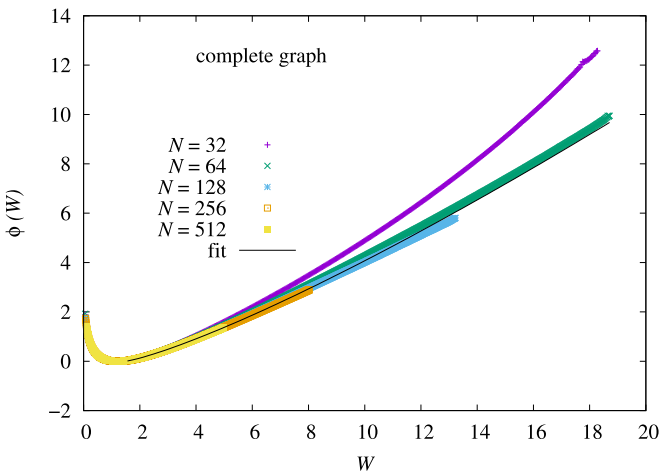


FIG. 5. Rate functions $\phi(W)$ for the complete graph for different values of the number N of nodes. The line shows the result of a fit to the data for $N = 256$ to the power-law equation (15).

due to the specific edge weights, i.e., the $N(N - 1)/2$ weights of each graph G . For this purpose, we have evaluated several quantities for the largest considered size $N = 512$ and for each statistically independent graph $G = (V, E)$ encountered in the MCMC simulation:

- (1) The mean weight

$$W_G^{\text{mean}} = \frac{2}{N(N - 1)} \sum_{\{i,j\} \in V} w_{i,j}, \quad (16)$$

where the mean is taken over all edges of the graph.

- (2) We have for each node $i \in V$ determined the minimum weight $w_{\min}(i) = \min_j w_{i,j}$ among all edges that are incident to the node. We take the average over all nodes, arriving at

$$W_G^{\min} = \frac{1}{N} \sum_i w_{\min}(i). \quad (17)$$

Since each node is connected to all other nodes by at least one edge, $(N - 1)W_G^{\min}$ forms a lower bound and, as we will see, also provides a good approximation of the MST weight W_G .

- (3) Since each spanning tree consists of $N - 1$ edges, we have stored the weight W_G^{N-1} of the $(N - 1)$ th smallest weight in each graph G . This is obtained by first sorting the $N(N - 1)/2$ weights $\{w_{i,j}\}$ in ascending order $w_1^* < w_2^* < \dots < w_{N(N-1)/2}^*$ and then taking

$$W_G^{N-1} = w_{N-1}^*, \quad (18)$$

$(N - 1)W_G^{N-1}$ might serve as a good upper approximation for W_G , but it is not a strict upper bound.

- (4) Finally, we also store the variance

$$W_G^{\text{var}} = \frac{2}{N(N - 1)} \sum_{\{i,j\} \in V} (w_{i,j} - W_G^{\text{mean}})^2, \quad (19)$$

of the weights of each graph to see whether all edge weights get concentrated around certain values.

We now analyze these quantities as a function of the MST weight W of each graph G . For this purpose, we show the averages of $W_G^{\min}$, W_G^{N-1} , and W_G^{mean} conditioned on $W_G = W$, by using bins, in Fig. 6. For better comparison, these averages are displayed as a function of $w = W/(N - 1)$, 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 their typical values, conditioned on W . For 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, most of which do not contribute to the MST. As shown in the inset of Fig. 6, a typical edge weight is always close to 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 on large MST weights, is 0.509, only slightly larger than the

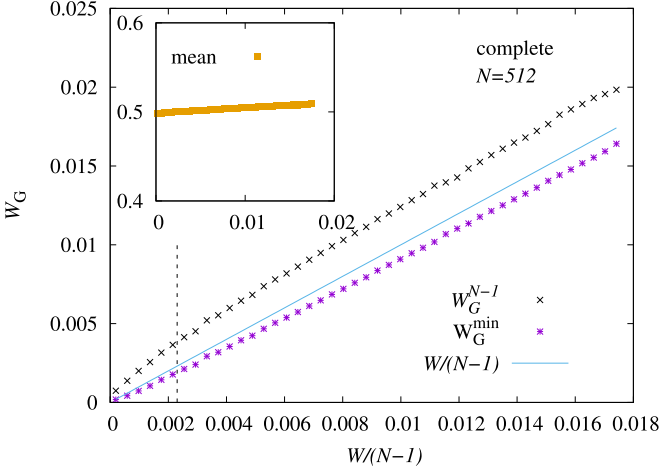


FIG. 6. Average of $W_G^{\min}$, W_G^{N-1} , and (inset) W_G^{mean} , each conditioned on the value W of the MST of G (for space-saving reasons, we do not repeat the labels for the inset). The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.002$ indicates the typical edge weight of an MST, corresponding to $P(W)$ exhibiting a peak near $W = 1$ in Fig. 4 for $N = 512$.

unconditional mean of 0.5. A similar weak effect can be seen for the empirical variance (not shown), 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 0.0805 for $W/(N-1) = 0.018$.

Note that such a large range of values $W/(N-1) \in [0, 0.018]$ can only be accessed with a large-deviation approach. With direct sampling of, say, 10^6 graphs, 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, we studied the same set of system sizes $N \in \{32, 64, 128, 256, 512\}$ as for the complete graphs. We have investigated five different values of the connectivity, namely, $c = 0.5, 0.7, 1.0, 2.0$, and 5.0 . This includes two values in the nonpercolating regime $c < 1$, one at the critical connectivity, and two in the percolating regime $c > 1$, for the standard ER graphs. We again had to choose the number n_θ of temperatures and the actual values, which were chosen in the same way as described in Sec. IV A. For convenience, we used the same temperature parameters for all five values of c . For $N \leq 256$, $n_\theta = 19$ temperatures were sufficient, while for $N = 512$, we needed $n_\theta = 30$ different values. The range of temperatures where equilibration was possible within a reasonable time again depended on the size N . Clearly, for all system sizes we again considered a large temperature $\theta = 10^5$ for basically unbiased sampling. For the smallest size $N = 32$, the positive temperatures were selected as $\theta \in [0.01, 1]$, while the negative temperatures were taken from the interval $[-1, -0.01]$. These intervals gradually became smaller with growing size N . For $N = 512$, finally, we considered $\theta \in [0.05, 1]$ and $\theta \in [-1, -0.1]$. As mentioned, in each MCMC step, exactly one node was selected randomly, and all adjacent edges were randomly recreated, including the weights.

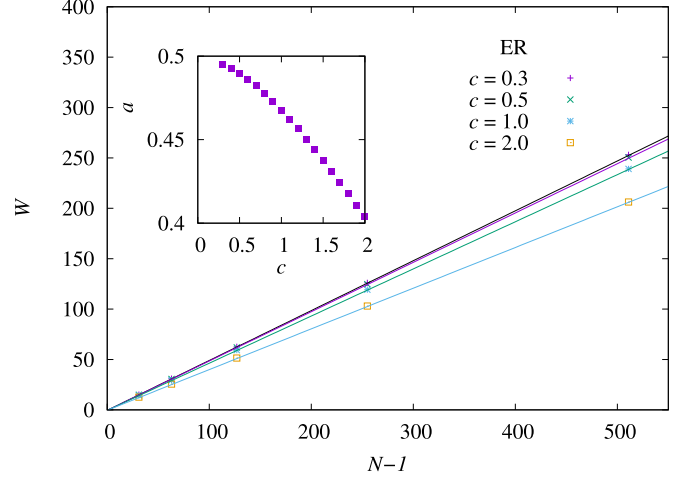


FIG. 7. Average spanning tree weight W as a function of the number N of nodes for the connected ER ensemble. The lines indicate fits to a linear function with slope a . The inset shows the slope a as a function of the connectivity c .

The resulting behavior in the ER ensemble can be richer than in complete graphs, since not only the edge weights but also the graph structure vary. We first analyze the mean MST weight, denoted 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. 7. 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. 7, 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, 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. 8. 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 principle is fulfilled.

As mentioned above, the restriction to connected graphs means we expect that 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 higher probabilities. From the central limit theorem, we

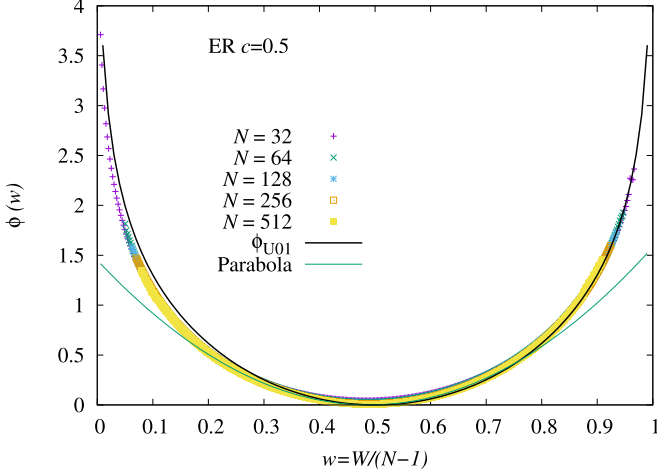


FIG. 8. Rate functions $\phi(w)$ for the connected ER graph with $c = 0.5$ for different values of the number N of nodes. The black line shows the rate function ϕ_{U01} for the average of $N - 1$ iid random numbers distributed uniformly in $[0, 1]$. The green line shows the result of a fit of the data for $N = 512$ in the high-probability region to a parabola equation (20).

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 clearly reflected in the shape of the rate function. Indeed, a fit to a parabola

$$\tilde{\phi}(w) = l(w - w_0)^2, \quad (20)$$

in the range $w \in [0.3, 0.7]$ yields a good fit; see Fig. 8. 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. For example, 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 also compare 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 [43], 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 taken 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 variable 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. 8. A very good agreement with the numerical data is visible; the small differences are likely because graphs with more than

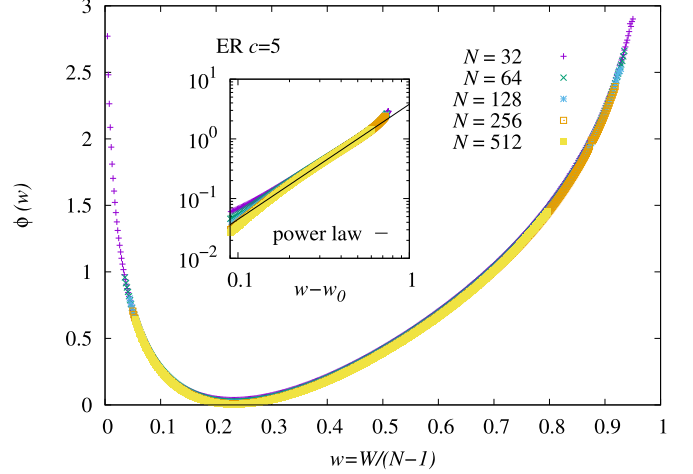


FIG. 9. Rate functions $\phi(w)$ for the connected ER graph with $c = 5$ for different values of the number N of nodes. In the inset, the same data is shown on a double log scale. The line shows the result of a fit of the data for $N = 512$ to a power-law equation (15).

$N - 1$ edges also contribute, in particular for small values of w .

We have also considered $c = 0.7$ and $c = 1$ in the (for the original ensemble) nonpercolating and critical region; here, the behavior is very similar.

For larger values of c , the shape of the distribution changes. In Fig. 9, 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 as $N \rightarrow \infty$ exists and therefore the large-deviation principle also holds here. In the intermediate regime, $w \in [0.3, 0.7]$, we have fitted a power-law equation (15) 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.

Note that for larger values of w , the data bend upward; thus, a growth with a larger exponent is possible. Such a change in the behavior of rate functions, sometimes termed “phase transitions,” has also been observed for the rate functions of other random systems like those described by the Kardar-Parisi-Zhang equation [55].

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

$$\gamma_1 = \frac{\langle (W - \langle W \rangle)^3 \rangle}{\sigma^3}, \quad (21)$$

where σ^2 is the variance $\langle (W - \langle W \rangle)^2 \rangle$. The result is shown for different numbers N of nodes in Fig. 10. For values $c \leq 1$, the skewness is very close to zero, indicating a very symmetric distribution. Only from $c = 1$ onward, 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.

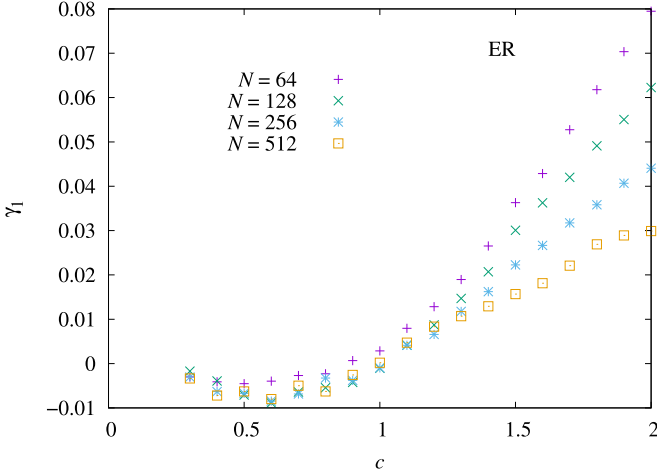


FIG. 10. Skewness γ_1 as a function of the connectivity parameter c for the connected ER ensemble, for different system sizes N .

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. 11 the average graph weights $W_G^{\min}$, W_G^{N-1} , and W_G^{mean} , for the case $c = 0.5$, conditioned on 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. 6, 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. 11. Only for atypically 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

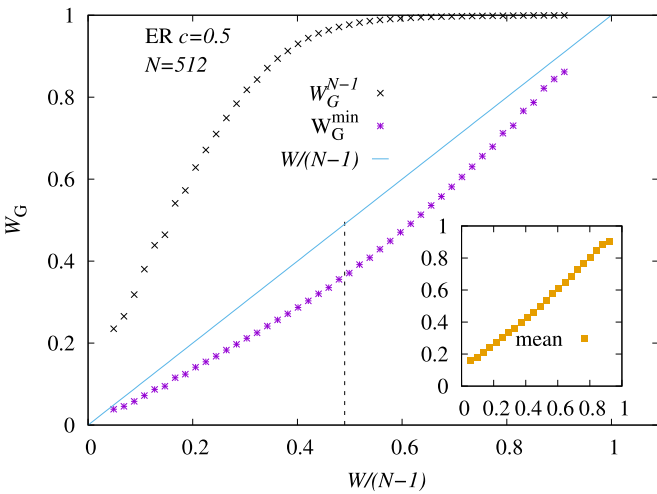


FIG. 11. Average of $W_G^{\min}$, W_G^{N-1} , and (inset) W_G^{mean} , each conditioned on the value W of the MST of G for ER random graphs with $c = 0.5$ (for space-saving reasons, we do not repeat the labels for the inset). The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.49$ indicates the typical edge weight in an MST for this value of c .

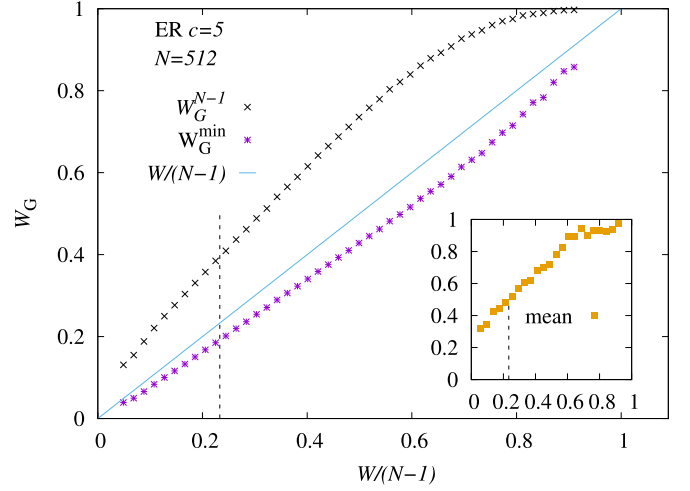


FIG. 12. Average of $W_G^{\min}$, W_G^{N-1} , and (inset) W_G^{mean} , each conditioned on the value W of the MST of G , for ER random graphs with $c = 5$ (for space-saving reasons, we do not repeat the labels for the inset). The line indicates the diagonal $W_G = W/(N-1)$. The vertical line near $W/(N-1) = 0.23$ indicates the typical edge weight in an MST for this value of c .

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. 12. Since more edges are present in the graph, but the graph is still sparse, the results look somewhat in between the results for $c = 0.5$ and for the complete graph. Again, the sparseness here 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 particularly small or large MST weights. In the ER ensemble, the number of edges also fluctuates. For any value of c , the typical number of edges, due to the restriction to connected graphs, is about $\max\{N-1, cN/2\}$. However, there exist rare graphs that are much denser or, for $c > 2$, much sparser. 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. 13, where the average number M of edges, conditioned on 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 , typical connected graphs also contain only a 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 their edge weights, as discussed above. For small values of $W/(N-1)$, conversely, the number of edges is atypically larger, so the minimization can lead to lower MST weights.

Also, not only the number of edges, i.e., the mean degree of the nodes, might matter, but also the actual degree

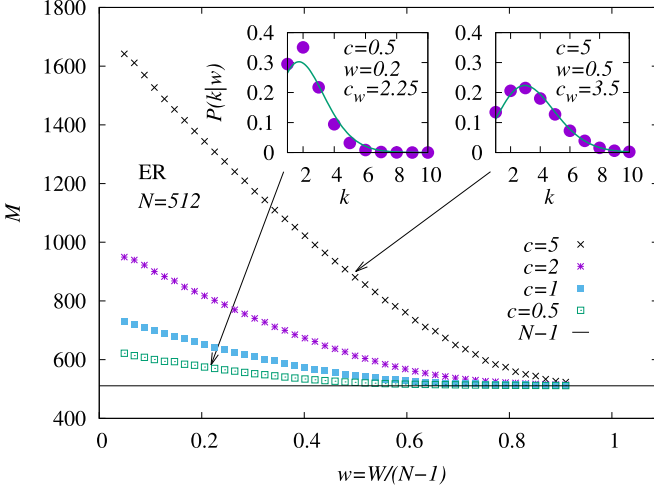


FIG. 13. Average of the number M of edges conditioned on the value W of the MST of G , shown as a function of $w = W/(N-1)$ for $N = 512$ and several values of c . The horizontal line indicates the minimum number $N-1$ of edges needed to create a connected graph. The inset plots show the degree distributions $P(k|w)$ of graphs conditioned on the values w for (left) $c = 0.5$, $w = 0.2$ and (right) $c = 5$ and $w = 0.5$. The lines show the restricted Poisson distribution equation (22) for $c = c_w = 2.25$, corresponding to $M = 576$ edges, and $c = c_w = 3.5$, corresponding to $M = 896$ edges, respectively.

distributions $P(k)$. In principle, graphs with extreme values of W could exhibit nodes with atypically high degrees k , i.e., hubs, or conversely an atypical number of nodes with very small degrees like $k = 1$. Nevertheless, we have found that the degree distributions are still basically Poisson, as in the standard ER graphs, with the restriction that nodes of degree $k = 0$ do not exist because of the condition that all graphs are connected. Since the standard Poisson distribution for degree c is given by $p(k, c) = e^{-c} c^k / k!$, and because the probability for a nonzero degree is $1 - p(0, c)$, the restricted Poisson distribution is given by

$$q(k, c) = \frac{p(k, c)}{1 - p(0, c)} = \frac{e^{-c}}{1 - e^{-c}} \frac{c^k}{k!} \quad (k \geq 1). \quad (22)$$

Clearly, this is only an approximation; the real degree distributions of ER graphs restricted to connected graphs will have a more complex structure.

If the degree distribution $P(k|w)$ for graphs conditioned on specific values of $w = W/(N-1)$ follows the restricted Poisson, then the mean degree c_w observed for these graphs should be sufficient to determine the corresponding degree distribution according to Eq. (22). This is indeed the case, as the two insets in Fig. 13 show. First, for $c = 0.5$, the typical MST weight w per edge is 0.5, as visible in Fig. 8. Here, we consider graphs with atypically small MST weight densities $w = 0.2$. This leads to an average conditioned degree $c_w = 2.25$, which is larger than the typical degree $c = 2$ needed for the graph to be connected. In the left inset, one sees that the measured degree distribution follows the restricted Poisson for $c = c_w = 2.25$ rather well. Second, for $c = 5$, the typical MST weight w per edge is near 0.2, as visible in Fig. 9. Here, we consider graphs conditioned on having an atypically large MST weight per edge of $w = 0.5$. These graphs exhibit an

average degree of $c_w = 3.5$, corresponding to $M = 896$ edges. Again, the restricted Poisson for $c = c_w = 3.5$ is rather close to the observed degree distribution $P(k|w)$. Thus, the average degree of the sampled graphs, as shown in the main plot of Fig. 9, is sufficient to at least approximately determine the actual degree distributions.

In summary, both atypical graph structure and atypical edge weights influence the observed atypical MST weights in weighted connected ER graphs.

V. DISCUSSION

We have studied the distribution of the weight of minimum spanning trees for complete graphs and for connected ER random graphs with connectivity c . With respect to simple properties such as the expectation value, our results are compatible with previous analytical studies. By applying large-deviation approaches, we can 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 property is fulfilled, i.e., away from the typical values the probabilities become exponentially small with an exponent proportional to N .

For the complete graph, the behavior of the rate function of the right tail is compatible with a power-law behavior. Since the graph structure is fixed, rare MST weights W are created by rare assignments of edge weights. However, only the lowest edge weights need to deviate from typical values to determine W , so the overall mean weight of the graph is only weakly affected.

In 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 of the standard ER ensemble, the distributions become asymmetric and nontrivial tails develop. Here, the values of the 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 analyze more properties, e.g., whether the variance of the node degrees also plays a likely minor role. For future studies, one could also 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 using 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 in the ER case.

ACKNOWLEDGMENTS

M.S. thanks A. Ramezanzpour for helpful discussions and assistance with debugging the code. The simulations were performed on the HPC cluster ROSA, located at the University of Oldenburg (Germany), and funded by the DFG through its Major Research Instrumentation Program (INST Grant No.

184/225-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State. ChatGPT was used to perform a spelling and grammar check of the manuscript.

DATA AVAILABILITY

The data generated in this work are publicly available in the DARE repository of the University of Oldenburg [56].

-
- [1] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [2] M. E. J. Newman, The structure and function of complex networks, *SIAM Rev.* **45**, 167 (2003).
 - [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Complex networks: Structure and dynamics, *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 University Press, Oxford, 2003).
 - [5] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
 - [6] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2012).
 - [7] M. Biskup, L. Chayes, and S. A. Smith, Large-deviations/thermodynamic approach to percolation on the complete graph, *Random Struct. Algorithms* **31**, 354 (2007).
 - [8] A. K. Hartmann, Large-deviation properties of largest component for random graphs, *Eur. Phys. J. B* **84**, 627 (2011).
 - [9] H. Schawe and A. K. Hartmann, Large deviations of connected components in the stochastic block model, *Phys. Rev. E* **102**, 052108 (2020).
 - [10] T. Łuczak, Random trees and random graphs, *Random Struct. Algorithms* **13**, 485 (1998).
 - [11] A. K. Hartmann and M. Mézard, Distribution of diameters for Erdős-Rényi random graphs, *Phys. Rev. E* **97**, 032128 (2018).
 - [12] M. Mézard and G. Parisi, On the solution of the random link matching problems, *J. Phys.* **48**, 1451 (1987).
 - [13] G.-Y. Shi, Y.-X. Kong, H. Liao, and Y.-C. Zhang, Analysis of ground state in random bipartite matching, *Physica A* **444**, 397 (2016).
 - [14] C. Lucibello, E. M. Malatesta, G. Parisi, and G. Sicuro, The random fractional matching problem, *J. Stat. Mech.* (2018) 053301.
 - [15] E. M. Malatesta, G. Parisi, and G. Sicuro, Fluctuations in the random-link matching problem, *Phys. Rev. E* **100**, 032102 (2019).
 - [16] J. Houdayer, J. H. Boutet de Monvel, and O. C. Martin, Comparing mean field and Euclidean matching problems, *Eur. Phys. J. B* **6**, 383 (1998).
 - [17] R. L. Graham and P. Hell, On the History of the Minimum Spanning Tree Problem, *IEEE Ann. Hist. Comput.* **7**, 43 (1985).
 - [18] R. C. Prim, Shortest connection networks and some generalizations, *Bell Syst. Tech. J.* **36**, 1389 (1957).
 - [19] L. L. Cavalli-Sforza and A. W. F. Edwards, Phylogenetic analysis: Models and estimation procedures, *Evolution* **21**, 550 (1967).
 - [20] J. C. Gower and G. J. S. Ross, Minimum spanning trees and single linkage cluster analysis, *J. R. Stat. Soc. C: Appl. Stat.* **18**, 54 (1969).
 - [21] J. Wu, X. Li, L. Jiao, X. Wang, and B. Sun, Minimum spanning trees for community detection, *Physica A* **392**, 2265 (2013).
 - [22] R. E. Osteen and P. P. Lin, Picture skeletons based on eccentricities of points of minimum spanning trees, *SIAM J. Comput.* **3**, 23 (1974).
 - [23] Z. Zhang, B. Wu, and Y. Lin, Counting spanning trees in a small-world Farey graph, *Physica A* **391**, 3342 (2012).
 - [24] T. Li and W. Yan, Enumeration of spanning trees of 2-separable networks, *Physica A* **536**, 120877 (2019).
 - [25] C. Dussert, G. Rasigni, M. Rasigni, J. Palmari, and A. Llebaria, Minimal spanning tree: A new approach for studying order and disorder, *Phys. Rev. B* **34**, 3528 (1986).
 - [26] D. V. Ktitarov, S. Lübeck, P. Grassberger, and V. B. Priezzhev, Scaling of waves in the Bak-Tang-Wiesenfeld sandpile model, *Phys. Rev. E* **61**, 81 (2000).
 - [27] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, Transport in weighted networks: Partition into superhighways and roads, *Phys. Rev. Lett.* **96**, 148702 (2006).
 - [28] D.-H. Kim, J. D. Noh, and H. Jeong, Scale-free trees: The skeletons of complex networks, *Phys. Rev. E* **70**, 046126 (2004).
 - [29] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, Topology of correlation-based minimal spanning trees in real and model markets, *Phys. Rev. E* **68**, 046130 (2003).
 - [30] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, *Phys. Rev. E* **68**, 056110 (2003).
 - [31] T. S. Jackson and N. Read, Theory of minimum spanning trees. I. Mean-field theory and strongly disordered spin-glass model, *Phys. Rev. E* **81**, 021130 (2010).
 - [32] T. S. Jackson and N. Read, Theory of minimum spanning trees. II. Exact graphical methods and perturbation expansion at the percolation threshold, *Phys. Rev. E* **81**, 021131 (2010).
 - [33] S. M. Sweeney and A. A. Middleton, Minimal spanning trees at the percolation threshold: A numerical calculation, *Phys. Rev. E* **88**, 032129 (2013).
 - [34] T. I. Fenner and A. M. Frieze, On the connectivity of random-orientable graphs and digraphs, *Combinatorica* **2**, 347 (1982).
 - [35] A. M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Appl. Math.* **10**, 47 (1985).
 - [36] C. McDiarmid, in *Surveys in Combinatorics*, edited by J. Siemons (Cambridge University Press, Cambridge, 1989), pp. 148–188.
 - [37] S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, *Random Struct Algorithms* **7**, 337 (1995).
 - [38] A. D. Flaxman, The lower tail of the random minimum spanning tree, *Electron. J. Combin.* **14**, N3 (2007).
 - [39] W. V. Li and X. Zhang, On the difference of expected lengths of minimum spanning trees, *Comb. Probab. Comput.* **18**, 423 (2009).
 - [40] C. Cooper, A. Frieze, N. Ince, S. Janson, and J. Spencer, On the length of a random minimum spanning tree, *Comb. Probab. Comput.* **25**, 89 (2016).

- [41] J. A. Bucklew, *Introduction to Rare Event Simulation* (Springer New York, 2004).
- [42] F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2008).
- [43] H. Touchette, The large deviation approach to statistical mechanics, *Phys. Rep.* **478**, 1 (2009).
- [44] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17 (1960).
- [45] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (MIT Press, Cambridge, 2001).
- [46] A. K. Hartmann, Sampling rare events: Statistics of local sequence alignments, *Phys. Rev. E* **65**, 056102 (2002).
- [47] A. K. Hartmann, Numerical aspects of large deviations, *SciPost Phys. Lect. Notes* **100** (2025).
- [48] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [49] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97 (1970).
- [50] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, Oxford, 1999).
- [51] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2005).
- [52] One could optimize the performance by changing more entries, with the number of changes depending on N and θ , but with our simple choice we could obtain sufficient performance.
- [53] The data generated in this work are publicly available in the DARE repository of the University of Oldenburg [56].
- [54] 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.
- [55] A. K. Hartmann, P. Le Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, High-precision simulation of the height distribution for the KPZ equation, *Europhys. Lett.* **121**, 67004 (2018).
- [56] M. Sarikhani and A. K. Hartmann, Data and gnuplot plot files, DARE public repository, doi: [10.57782/BKJRBF](https://doi.org/10.57782/BKJRBF) (2025).