Partition functions of discrete coalescents: from Cayley's formula to Frieze's $\zeta(3)$ limit theorem

Louigi Addario-Berry

Abstract. In these expository notes, we describe some features of the multiplicative coalescent and its connection with random graphs and minimum spanning trees. We use Pitman's proof [13] of Cayley's formula, which proceeds via a calculation of the partition function of the additive coalescent, as motivation and as a launchpad. We define a random variable which may reasonably be called the empirical partition function of the multiplicative coalescent, and show that its typical value is exponentially smaller than its expected value. Our arguments lead us to an analysis of the susceptibility of the Erdős-Rényi random graph process, and thence to a novel proof of Frieze's $\zeta(3)$-limit theorem for the weight of a random minimum spanning tree.

1. Introduction

Consider a discrete time process $(P_i, 1 \le i \le n)$ of coalescing blocks, with the following dynamics. The process starts from the partition of $[n] = \{1, \ldots, n\}$ into singletons: $P_1 = \{\{1\}, \ldots, \{n\}\}$. To form P_{i+1} from P_i choose two parts P, P' from P_i and merge them. We assume there is a function κ such that the probability of choosing parts P, P' is proportional to $\kappa(|P|, |P'|)$; call κ a *rate kernel*.

Different rate kernels lead to different dynamics. Three kernels whose dynamics have been studied in detail are $\kappa(x, y) = 1$, $\kappa(x, y) = x + y$, and $\kappa(x, y) = xy$; these are often called Kingman's coalescent, the additive coalescent, and the multiplicative coalescent, respectively. In these cases there is a natural way to enrich the process and obtain a *forest-valued* coalescent.

These notes are primarily focussed on the properties of the forest-valued multiplicative coalescent. We proceed from a statistical physics perspective, and begin by analyzing the partition functions of the three coalescents. Here is what we mean by this. Say that a sequence $(P_1, \ldots, P_n)$ of partitions of $[n]$ is an *n-chain* if $P_1 = \{\{1\}, \ldots, \{n\}\}$ is the partition of n into singletons, and for $1 \le i < n$, P_{i+1} can be formed from P_i by merging two parts of P_i. Think of $\kappa(x, y)$ as the number

of possible ways to merge a block of size x with one of size y. Then corresponding to an n-chain $P = (P_1, \ldots, P_n)$ there are

$$\prod_{i=1}^{n-1} \kappa(|A_i(P)|, |B_i(P)|)$$

possible ways that the coalescent may have unfolded; here we write $A_i(P)$ and $B_i(P)$ for the blocks of P_i that are merged in P_{i+1}. Writing $\mathcal{P}_n$ for the set of n-chains, it follows that the total number of possibilities for the coalescent with rate kernel κ is

$$\sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \prod_{i=1}^{n-1} \kappa\left(|A_i(P)|, |B_i(P)|\right),$$

and we view this quantity as the partition function of the coalescent with kernel κ.

The partition functions of Kingman's coalescent and the additive and multiplicative coalescents have particularly simple forms: they are

$$Z_{\text{KC}}(n) = n!(n-1)!,$$
$$Z_{\text{AC}}(n) = n^{n-1}(n-1)!, \text{ and}$$
$$Z_{\text{MC}}(n) = n^{n-2}(n-1)!.$$

These formulae are proved in Section 2. A corollary of the formula for $Z_{\text{KC}}(n)$ is that the number of increasing trees with n vertices is $(n-1)!$; this easy fact is well-known. The formula for $Z_{\text{AC}}(n)$ is due to Pitman [13], who used it to give a beautiful proof of Cayley's formula; this is further detailed in Section 2.1.

It may seem surprising that the partition function of the multiplicative coalescent is so similar to that of the additive coalescent: near the start of the process, when most blocks have size 1, the additive coalescent has twice as many choices as the multiplicative coalescent. Later in the process, blocks should be larger, and one would guess that usually $xy > x + y$. Why these two effects should almost exactly cancel each other out is something of a mystery. On the other hand, the similarity of the partition functions may suggest that the additive and multiplicative coalescents have similar behaviour.

A more detailed investigation will reveal interesting behaviour whose subtleties are not captured by the above formulae. We will see in Section 2.3 that there is a naturally defined "empirical partition function" $\hat{Z}_{\text{MC}}(n)$ such that $Z_{\text{MC}}(n) = \mathbf{E}\left[\hat{Z}_{\text{MC}}(n)\right]$. However, $\hat{Z}_{\text{MC}}(n)$ is typically *exponentially smaller* than $Z_{\text{MC}}(n)$ (see Corollary 4.3), so in a quantifiable sense, the partition function $Z_{\text{MC}}(n)$ takes the value it does due to extremely rare events. Correspondingly, it turns out that the behaviour of the additive and multiplicative coalescents are typically quite different.

To analyze the typical value of $\hat{Z}_{\text{MC}}(n)$, we are led to develop the connection between the multiplicative coalescent and the classical Erdős-Rényi random graph process $(G(n, p), 0 \leq p \leq 1)$. The most technical part of the notes is the proof of

a concentration result for the susceptibility of $G(n,p)$; this is Theorem 4.4, below. Using a well-known coupling between the multiplicative coalescent and Kruskal's algorithm for the minimum weight spanning tree problem, our susceptibility bound leads easily to a novel proof of the $\zeta(3)$ limit for the total weight of the minimum spanning tree of the complete graph (this is stated in Theorem 5.1, below).[1]

Stylistic remarks

The primary purpose of these notes is expository (though there are some new results, notably Theorems 4.2 and 4.4). Accordingly, we have often opted for repetition over concision. We have also included plenty of exercises and open problems (the open problems are mostly listed in Section 7). Some exercises state facts which are required later in the text; these are distinguished by a ⊛.

2. A tale of three coalescents

2.1. Cayley's formula and Pitman's coalescent

We begin by describing the beautiful proof of Cayley's formula found by Jim Pitman, and its link with uniform spanning trees. Cayley's formula states that the number of trees with vertices $\{1, 2, \ldots, n\}$ is n^{n-2}, or equivalently that the number of *rooted* trees with vertices labeled by $[n] := \{1, 2, \ldots, n\}$ is n^{n-1}. To prove this formula, Pitman [13] analyzes a process we call Pitman's coalescent. To explain the process, we need some basic definitions. A *forest* is a graph with no cycles; its connected components are its trees. A *rooted forest* is a forest in which each tree t has a distinguished root vertex $r(t)$.

> **Pitman's Coalescent, Version 1.** The process has n steps, and at step i consists of a rooted forest $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$ with $n + 1 - i$ trees. (At step 1, these trees are simply isolated vertices with labels $1, \ldots, n$.) To obtain F_{i+1} from F_i, choose a pair (U_i, V_i), where $U_i \in [n]$ and V_i is the root of some tree of F_i not containing U_i, uniformly at random from among all such pairs. Add an edge from U_i to V_i, and root the resulting tree at the root of U_i's old tree. The forest F_{i+1} consists of this new tree together with the $n - i - 1$ unaltered trees from F_i.

The coalescents we consider all have the general form of Pitman's coalescent: they are forest-valued stochastic processes $(F_i, 1 \leq i \leq n)$, where $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$ is a forest with vertices labeled by $[n]$.

---

[1] We find this proof of the $\zeta(3)$ limit for the MST weight pleasing, as it avoids lemmas which involve estimating the number of unicyclic and complex components in $G(n,p)$; morally, the cycle structure of components of $G(n,p)$ should be unimportant, since cycles are never created in Kruskal's algorithm!

> **Pitman's Coalescent, Version 2.** Consider the directed graph $K_n^{\to}$ with vertices $\{1, \ldots, n\}$ and an oriented edge from k to ℓ for each $1 \leq k \neq \ell \leq n$. Let $\mathbf{W} = \{W_{(k,\ell)} : 1 \leq k \neq \ell \leq n\}$ be independent copies of a continuous random variable W, that weight the edges of $K_n^{\to}$. Let F_1 be as in Version 1. For $i \in \{1, \ldots, n-1\}$, form F_{i+1} from F_i by adding the smallest weight edge (k, ℓ) whose head ℓ is the root of one of the trees in F_i. (Each tree of F_i is rooted at its unique vertex having indegree zero in F_i.)

Note that in Version 2, for each $i \in \{1, \ldots, n\}$ and each tree T of F_i, all edges of T are oriented away from a single vertex of T; so, viewing this vertex as the root of T, the orientation of edges in T is fully specified by the location of its root.

Exercise 1. View the trees of Version 2 as rooted rather than oriented. Then the sequences of forests $(F_1, \ldots, F_n)$ described in Version 1 and Version 2 have the same distribution.

Say that a finite set $\{X_i, i \in I\}$ of random variables is *exchangeable* if for any two deterministic orderings of I as, say, $i_1, \ldots, i_k$ and $i_1', \ldots, i_i'$, the vectors $(X_{i_1}, \ldots, X_{i_k})$ and $(X_{i_1'}, \ldots, X_{i_k'})$ are identically distributed. In particular, if the elements of $\{X_i, i \in I\}$ are iid then the set is exchangeable.

Exercise 2. Suppose that the edge weights $\mathbf{W}$ are only assumed to be exchangeable and a.s. pairwise distinct. Show that the sequences of forests $(F_1, \ldots, F_n)$ described in Version 1 and Version 2 still have the same distribution.

To prove Cayley's formula, we compute the *partition function* of Pitman's coalescent: this is the total number of possibilities for its execution. (To do so, it's easiest to think about Version 1 of the procedure.) For example, when $n = 3$, there are 6 possibilities for the first step of the process: 3 choices for the first vertex, then 2 choices of a tree not containing the first vertex. For the second step, there are 3 choices for the first vertex; there is only one component not containing the chosen vertex, and we must choose it. Thus, for $n = 3$, the partition function has value $Z_{\mathrm{AC}}(3) = 6 \cdot 3 = 18$. More generally, for the n-vertex process, when adding the i'th edge we have n choices for the first vertex and $n - i$ choices of tree not containing the first vertex, so a total of $n(n - i)$ possibilities. Thus the partition function is

$$Z_{\mathrm{AC}}(n) = \prod_{i=1}^{n-1} n \cdot (n - i) = n^{n-1}(n - 1)! \tag{2.1}$$

It is not possible to recover the entire execution path of the additive coalescent from the final tree, since there is no way to tell in which order the edges were added. If we wish to retain this information, we may label each edge of $T_1^{(n)}$ with the step at which it was added. More precisely, $L(e)$ is the unique integer $i \in \{1, \ldots, n-1\}$
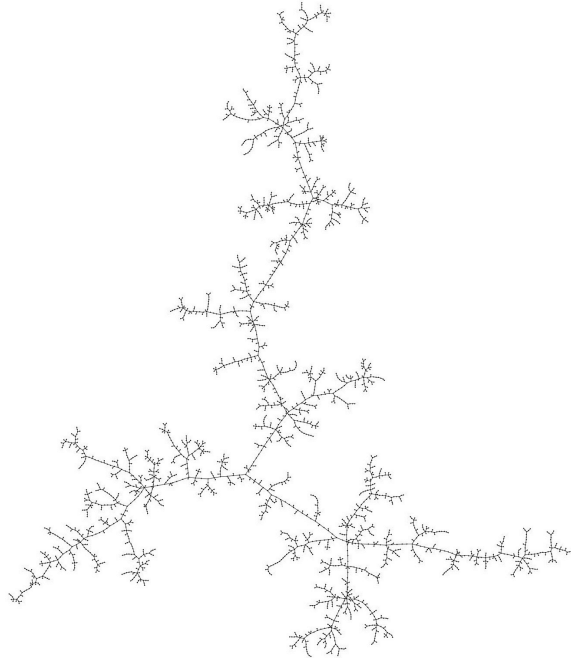
FIGURE 1. One of the 3000^{2998} labeled trees with 3000 vertices, selected uniformly at random.

such that e is not an edge of F_i but is an edge of F_{i+1}. It follows from the definition of the process that the edge labels are distinct, so $L : E(T_1^{(n)}) \to \{1, \ldots, n-1\}$ is a bijective map.

Now fix a rooted tree t with vertices $\{1, \ldots, n\}$, and consider the *restricted* partition function $Z_{\mathrm{AC},t}(n)$; this is simply the number of possibilities for the execution of the process for which the end result is the tree t. We claim that $Z_{\mathrm{AC},t}(n) = (n-1)!$. This is easy to see: for any labelling ℓ of the edges of t with integers $\{1, \ldots, n-1\}$, there is a unique execution path for which $(T_1^{(n)}, L) = (t, \ell)$, and there are $(n-1)!$ possible labellings ℓ. Thus, the probability of ending with the tree t is $Z_{\mathrm{AC},t}(n)/Z_{\mathrm{AC}}(n) = 1/n^{n-1}$. Since this number doesn't depend on t, only on n, it follows that every rooted labelled tree with n vertices is equally likely, and so there must be n^{n-1} such trees.

Note. *The preceding argument is correct, but treads lightly around an important point. When performing the process, the number of possibilities for the i'th edge does not depend on the first $i-1$ choices, so the probability of building a particular tree t by adding its edges in in a particular order is $[n^{n-1}(n-1)!]^{-1}$ regardless of the order. Of course, the* set *of possible choices at a given step must depend on the history of the process – for example, we must not add a single edge*

twice. More generally, thinking of Version 2, applying the procedure to a graph other than $K_n^{\rightarrow}$ need not yield a uniform spanning tree of the graph, and indeed may not even build a tree. (Consider, for example, applying the procedure to a two-edge path.)

By stopping Pitman's coalescent before the end, one can use a similar analysis to obtain counting formulae for forests. Write $Z_{\text{AC}}(n,k)$ for the total number of possibilities for Pitman's coalescent stopped at step k (so ending with $n+1-k$ forests). We write $(m)_\ell$ to denote the *falling factorial* $\prod_{i=0}^{\ell-1}(m-i)$.

Exercise 3.   (a) Show that $Z_{\text{AC}}(n,k) = n^{k-1}(n-1)_{k-1}$ for each for $1 \le k \le n$.
   (b) An *ordered labeled forest* is a sequence $(t_1, \ldots, t_\ell)$ where each t_i is a rooted labelled tree and all labels of vertices in the forest are distinct. Show that for each $1 \le \ell \le n$ the number of ordered labeled forests $(t_1, \ldots, t_\ell)$ with $\bigcup_{i=1}^\ell V(t_i) = [n]$, is $\ell \cdot n^{n-\ell-1} \cdot (n)_\ell$.

We briefly discuss a special case of Version 2. Suppose that $W_{(k,l)}$ is exponential with rate $X_{(k,\ell)}$, where $\mathbf{X} = \{X_{(k,\ell)} : 1 \le k \ne \ell \le n\}$ are independent copies of any non-negative random variable X. By standard properties of exponentials and the symmetry of the process, the dynamics in this case may be described as follows.

> **Pitman's Coalescent, Version 3.** Let F_1 be as in Version 1. For $i \in \{1, \ldots, n-1\}$, choose an edge whose head is the root of any one of the trees in F_i, each such edge (k,l) chosen with probability proportional to its weight $X_{(k,l)}$; add the chosen edge to create the forest F_{i+1}.

Consider Version 3 of the procedure after $i-1$ edges have been added. Conditional on $\mathbf{X}$ and on the forest $(T_1^{(i)}, \ldots, T_{n-i+1}^{(i)})$, the probability of adding a particular edge (k, ℓ) whose head is a root, is proportional to $X_{(k,\ell)}$, so is equal to

$$\frac{X_{(k,\ell)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\ldots,n\} \setminus V(T_m^{(i)})} X_{(j, r(T_m^{(i)}))}}.$$

Now fix any sequence $f_1, \ldots, f_n$ of forests that can arise in the process. Write $f_i = (t_k^{(i)}, 1 \le k \le n+1-i)$ and for $i = 1, \ldots, n-1$ write (k_i, ℓ_i) for the unique edge of f_{i+1} not in f_i. Then by the above,

$$\mathbf{P}\{F_i = f_i, 1 \le i \le n \mid \mathbf{X}\} = \prod_{i=1}^{n-1} \frac{X_{(k_i, \ell_i)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\ldots,n\} \setminus V(t_m^{(i)})} X_{(j, r(t_m^{(i)}))}}.$$

By Exercise 1 and the above analysis, it follows that for any such sequence $f_1, \ldots, f_n$,

$$\mathbf{E}\left[\prod_{i=1}^{n-1} \frac{X_{(k_i, \ell_i)}}{\sum_{m=1}^{n-i+1} \sum_{j \in \{1,\ldots,n\} \setminus V(t_m^{(i)})} X_{(j, r(t_m^{(i)}))}}\right] = \frac{1}{n^{n-1}(n-1)!}.$$

It is by no means obvious at first glance that this expectation should not depend on law of X, let alone that it should have such a simple form.

2.2. Kingman's coalescent and random recursive trees

Pitman's coalescent starts from isolated vertices labeled from $\{1, \ldots, n\}$, and builds a rooted tree by successive edge addition. At each step, an edge is added *to* some vertex, *from* some root (of a component not containing the chosen vertex). When we calculated $Z_{\mathrm{AC}}(n)$, it was important that the number of possibilities at each step depended only on the *number* of trees in the current forest and not, say, their sizes, or some other feature.

Pitman's merging rule (*to* any vertex, *from* a root) yielded a beautiful proof of Cayley's formula. It is natural to ask what other rules exist, and what information may be gleaned from them. Of course, *from* any vertex, *to* a root just yields the additive coalescent, with edges of the resulting tree oriented towards the root rather than towards the leaves. What about *from* any root, *to* any (other) root, as in the following procedure? In a very slight abuse of terminology, we call this rule *Kingman's coalescent*. We again start from a rooted forest F_1 of n isolated vertices $\{1, \ldots, n\}$. Recall that we write $F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$.

> **Kingman's Coalescent.** At step i, choose an ordered pair (U_i, V_i) of distinct roots from $\{r(T_1^{(i)}), \ldots, r(T_{n+1-i}^{(i)})\}$, uniformly at random from among the $(n+1-i)(n-i)$ such pairs. Add an edge from U_i to V_i, and root the resulting tree at U_i. The forest F_{i+1} consists of this new tree together with the $n-i-1$ unaltered trees from F_i.

Our convention is that when an edge is added from u to v, the root of the resulting tree is u; this maintains that edges are always oriented towards the leaves. For Kingman's coalescent, when i trees remain there are $i(i-1)$ possibilities for which oriented edge to add. Like for Pitman's coalescent, this number depends only on the number of trees, and it follows that the total number of possible execution paths for the process is

$$Z_{\mathrm{KC}}(n) = \prod_{i=2}^{n} i(i-1) = n!(n-1)! \,. \tag{2.2}$$

What does this number count?

To answer the preceding question, as in the additive coalescent let $L : E(T_1^{(n)}) \to \{1, \ldots, n-1\}$ label the edges of $T_1^{(n)}$ in their order of addition. It is easily seen that for Kingman's coalescent, the edge labels decrease along any root-to-leaf path; we call such a labelling a *decreasing edge labelling*.[2] Furthermore, any decreasing edge labelling of $T_1^{(n)}$ can arise. Once again, the full behaviour of the coalescent is described by pair $(T_1^{(n)}, L)$, and conversely, the coalescent determines $T_1^{(n)}$ and

---

[2] It is more common to order by *reverse* order of addition, so that labels increase along root-to-leaf paths; this change of perspective may help with Exercise 4.
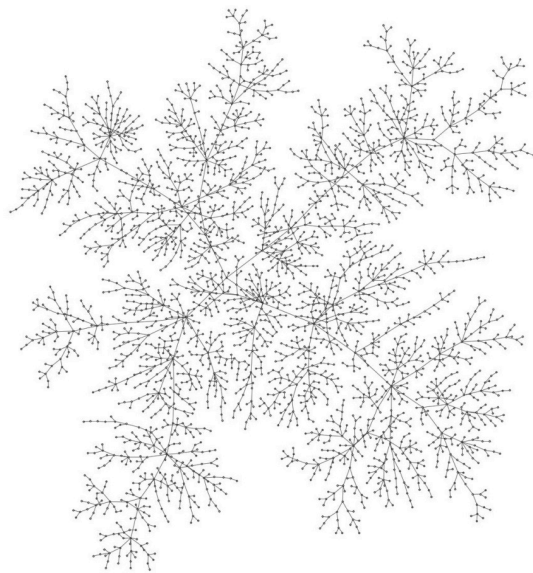
FIGURE 2. One of the 2999! rooted trees on 3000 vertices with a decreasing edge labelling (labels suppressed).

L. These observations yield that the number of rooted trees with vertices labelled $\{1, \ldots, n\}$, additionally equipped with a decreasing edge labelling, is $n!(n-1)!$. The factor $n!$ simply counts the number of ways to assign the labels $\{1, \ldots, n\}$ to the vertices. By symmetry, each vertex labelling of a given tree is equally likely to arise, and so we have the following.

Proposition 2.1. *The number of pairs (T, L), where T is a rooted tree with n vertices and L is a decreasing edge labelling of T, is $(n-1)!$.*

Exercise 4 (Random recursive trees). Prove Proposition 2.1 by introducing and analyzing an n-step procedure that at step i constructs a rooted tree with i vertices.

Before the next exercise, we state a few definitions. For a graph G, write $|G|$ for the number of vertices of G. If T is a rooted tree and u is a vertex of T, write T_u for the subtree of T consisting of u together with its descendants in T (we call T_u the subtree of T rooted at u). Also, if u is not the root, write $p(u)$ for the parent of u in T. Finally, write $\mathrm{aut}(T)$ for the number of rooted automorphisms of T.

Exercise 5. Show that for a fixed rooted tree T, the number of decreasing edge labellings of T with labels $1, 2, \ldots, |V(T)| - 1$ is

$$\frac{1}{\mathrm{aut}(T)} \cdot \prod_{v \in V(T)} \frac{(|T_v| - 1)!}{\prod_{\{u \in V(T) : p(u) = v\}} |T_u|!} \,.$$

Our convention is that an empty product equals 1; a special case is that $0! = 1$. It follows from the preceding exercise that, writing $\mathcal{T}_n$ for the set of rooted trees with n vertices,

$$\sum_{T \in \mathcal{T}_n} \frac{1}{\mathrm{aut}(T)} \cdot \prod_{v \in V(T)} \frac{|E(T_u)|!}{\prod_{\{u \in V(T) : p(u) = v\}} |V(T_u)|!} = (n-1)! \,;$$

another formula that one may not find obvious at first glance.

To finish the section, note that just like for Pitman's coalescent, we might well consider a version of this procedure that is "driven by" iid non-negative weights $\mathbf{X} = \{X_{(k,\ell)} : 1 \leq k \neq \ell \leq n\}$. (Recall that we viewed these weights as exponential *rates*, then used the resulting exponential clocks at each step to determine which edge to add.) At each step, add an oriented edge whose tail and head are both the roots of some tree of the current forest, each such edge chosen with probability proportional to its weight. For this procedure, conditional on $\mathbf{X}$, after adding the first $i - 1$ edges, the conditional probability of adding a particular edge (k, ℓ) is

$$\frac{X_{(k,\ell)}}{\sum_{1 \leq j \neq m \leq n} X_{(r(T_j^{(i)}), r(T_m^{(i)}))}} \,.$$

Now fix any sequence $f_1, \ldots, f_n$ of forests that can arise in the process, write $f_i = (t_k^{(i)}, 1 \leq k \leq n + 1 - i)$, and for $i = 1, \ldots, n - 1$ write (k_i, ℓ_i) for the unique edge of f_{i+1} not in f_i. Then we have

$$\mathbf{P}\{F_i = f_i, 1 \leq i \leq n \mid \mathbf{X}\} = \prod_{i=1}^{n-1} \frac{X_{(k_i, \ell_i)}}{\sum_{1 \leq m \neq j \leq n} X_{(r(t_m^{(i)}), r(t_j^{(i)}))}} \,.$$

It follows from the above analysis that for any such sequence $f_1, \ldots, f_n$,

$$\mathbf{E}\left[\prod_{i=1}^{n-1} \frac{X_{(k_i, \ell_i)}}{\sum_{1 \leq m \neq j \leq n} X_{(r(t_m^{(i)}), r(t_j^{(i)}))}}\right] = \frac{1}{n!(n-1)!} \,.$$

Once again, it is not even *a priori* clear that this expectation should not depend on the law of X.

Exercise 6 (First-passage percolation). Develop and analyze a "Version 3" variant of the tree growth procedure from Exercise 4, using exponential edge weights.
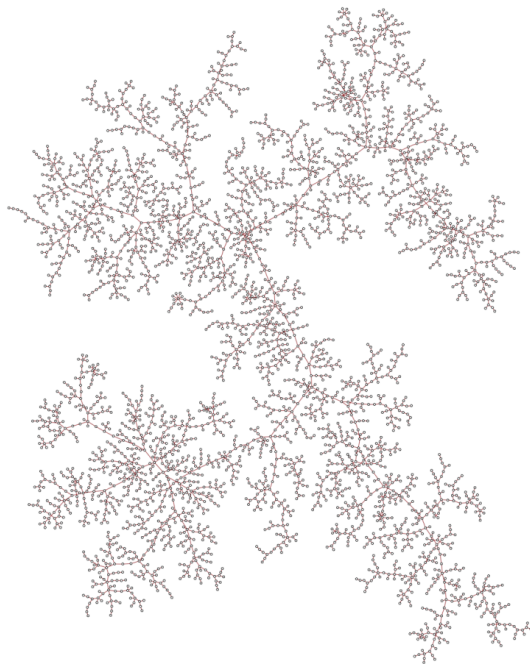
FIGURE 3. The tree resulting from the multiplicative coalescent
on 3000 points.

2.3. The multiplicative coalescent and minimum spanning trees

The previous two sections considered merging rules of the form any-to-root and
root-to-root, and obtained Pitman's coalescent and Kingman's coalescent, respec-
tively. We now take up the "any-to-any" merging rule. This is arguably the most
basic of the three rules, but its behaviour is arguably the hardest to analyze.
We begin as usual from a forest F_1 of n isolated vertices $\{1, \ldots, n\}$, and write
$F_i = \{T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}\}$. In the multiplicative coalescent there is no natural way
to maintain the property that edges are oriented toward some root vertex, so we
view the trees of the forests as unrooted, and their edges as unoriented. Given a
set S, write $\binom{S}{k}$ for the set of k-element subsets of S.

> **The multiplicative coalescent.** To obtain F_{i+1} from F_i, choose an pair
> $\{U_i, V_i\}$ uniformly at random from the set of pairs $\{u, v\} \in \binom{[n]}{2}$ for
> which u and v are different trees of F_i. Add an edge from U_i to V_i to
> form the forest F_{i+1}.

This is known as the *multiplicative* coalescent, because the number of possible
choices of an edge joining trees $T_j^{(i)}$ and $T_k^{(i)}$ is $|T_j^{(i)}||T_k^{(i)}|$. It follows that the

number of possible edges that may be added to the forest F_i is

$$\sum_{1 \leq j \neq k \leq n+1-i} |T_j^{(i)}||T_k^{(i)}| = \frac{1}{2}\left(n^2 - \sum_{T \in F_i} |T|^2\right).$$

The above expression is more complicated than for the additive coalescent or Kingman's coalescent: it depends on the forest F_i, for one.

In much of the remainder of these notes, we investigate an expression for the partition function $Z_{\mathrm{MC}}(n)$ of the multiplicative coalescent that arises from the preceding formula. To obtain this expression, recall the definition of an n-chain from Section 1, and that $\mathcal{P}_n$ is the set of n-chains.

Exercise 7. Show that $|\mathcal{P}_n| = \frac{(n!)^2}{n \cdot 2^{n-1}}$.

The multiplicative coalescent determines an n-chain in which the i'th partition is simply $P(F_i) := \{V(T_j^{(i)}), 1 \leq j \leq n + 1 - i\}$. It is straightforward to see that the number of possibilities for the multiplicative coalescent that give rise to a particular n-chain $P = (P_1, \ldots, P_n)$ is simply

$$\prod_{i=1}^{n-1} |A_i(P)||B_i(P)|,$$

where $A_i(P)$ and $B_i(P)$ are the parts of P_i that are combined in P_{i+1}. It follows that

$$Z_{\mathrm{MC}}(n) = \sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \prod_{i=1}^{n-1} (|A_i(P)||B_i(P)|).$$

This certainly looks more complicated than in the previous two cases. However, there is an exact formula for $Z_{\mathrm{MC}}(n)$ whose derivation is perhaps easier than for either $Z_{\mathrm{AC}}(n)$ or $Z_{\mathrm{KC}}(n)$ (though it does rely on Cayley's formula).

Proposition 2.2. $Z_{\mathrm{MC}}(n) = n^{n-2}(n-1)!$

Proof. Let $\mathcal{S}$ be the set of pairs (t, ℓ) where t is an unrooted tree with $V(t) = [n]$ and $\ell : E(t) \to [n-1]$ is a bijection. By Cayley's formula, the number of trees t with $V(t) = [n]$ is n^{n-2}, so $\mathcal{S} = n^{n-2}(n-1)!$.

For $e \in E(T_n^{(1)})$, let $L(e) = \sup\{i : e \notin E(F_i)\}$. Then $L : E(T_n^{(1)}) \to [n-1]$ is a bijection. Thus the pair $(T_n^{(1)}, L)$ is an element of $\mathcal{S}$. To see this map is bijective, note that if $(T_n^{(1)}, L) = (t, \ell)$ then for each $1 \leq i \leq n$, F_i is the forest on $[n]$ with edges $\{\ell^{-1}(j), 1 \leq j < i\}$. The result follows. $\square$

The above proposition yields that $Z_{\mathrm{MC}}(n) = Z_{\mathrm{AC}}(n)/n$. If we were to additionally choose a root for $T_n^{(1)}$, we would obtain identical partition functions. This suggests that perhaps the additive and multiplicative coalescents have similar structures. One might even be tempted to believe that the trees built by the

two coalescents are identically distributed; the following exercise (an observation of Aldous [3]), will disabuse you of that notion.

> **Exercise 8.** Let T be built by the multiplicative coalescent, and let T' be obtained from the additive coalescent by unrooting the final tree. Show that if $n \geq 4$ then T and T' are not identically distributed.

Despite the preceding exercise, it is tempting to guess that the two trees are still similar in structure; this was conjectured by Aldous [3], and only recently disproved [2]. In the remainder of the section, we begin to argue for the difference between the two coalescents, from the perspective of their partition functions. For $1 \leq k \leq n$, write $Z_{\mathrm{MC}}(n,k)$ for the partition function of the first k steps of the multiplicative coalescent,

$$Z_{\mathrm{MC}}(n,k) = \sum_{P=(P_1,\ldots,P_k)\in\mathcal{P}_{n,k}} \prod_{i=1}^{k-1} \left(|A_i(P)||B_i(P)|\right),$$

where $\mathcal{P}_{n,k}$ is the set of length-k initial segments of n-chains. We have, e.g., $Z_{\mathrm{MC}}(n,1) = 1$, $Z_{\mathrm{MC}}(n,2) = \binom{n}{2}$, and $Z_{\mathrm{MC}}(n,n) = Z_{\mathrm{MC}}(n)$.

The argument of Proposition 2.2 shows that $Z_{\mathrm{MC}}(n,k) = u_{n,k} \cdot (k-1)!$, where $u_{n,k}$ is the number of unrooted forests with vertices $[n]$ and $k-1$ total edges. The identity

$$u_{n,k} = \binom{n}{n+1-k} n^{k-2} \sum_{i=0}^{n+1-k} \left(\frac{-1}{2n}\right)^i \binom{n+1-k}{i}(n+1-k+i)\cdot(k-1)_i,$$

was derived by Rényi [14], and I do not know of an exact formula that simplifies the above expression. We begin to see that there is more to the multiplicative coalescent than first meets the eye.

If we can't have a nice, simple identity, what about bounds? Of course, there is the trivial upper bound $Z_{\mathrm{MC}}(n,k) \leq (n(n-1)/2)^{k-1}$, since at each step there are at most $\binom{n}{2}$ pairs to choose from; similar bounds hold for the other two coalescents. To improve this bound, and more generally to develop a deeper understanding of the dynamics of the multiplicative coalescent, our starting point is the following observation.

Given an n-chain $P = (P_1,\ldots,P_n)$, for the multiplicative coalescent we have

$$\mathbf{P}\left\{(P(F_i), 1 \leq i \leq n) = P\right\} = \prod_{i=1}^{n-1} \frac{2|A_i(P)||B_i(P)|}{n^2 - \sum_{\pi\in P_i}|\pi|^2}.$$

This holds since for $1 \leq i \leq n-1$, given that $P(F_j) = P_j$ for $1 \leq j \leq i$, there are $(n^2 - \sum_{\pi\in P_i}|\pi|^2)/2$ choices for which oriented edge to add to form F_{i+1}, and

$P(F_{i+1}) = P_{i+1}$ for precisely $|A_i(P)||B_i(P)|$ of these. It follows that

$$Z_{\mathrm{MC}}(n) = \sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \mathbf{P}\left\{(P(F_i), 1 \le i \le n) = P\right\} \cdot \prod_{i=1}^{n-1} \frac{n^2 - \sum_{\pi\in P_i}|\pi|^2}{2}$$

$$= \sum_{P=(P_1,\ldots,P_n)\in\mathcal{P}_n} \mathbf{P}\left\{(P(F_i), 1 \le i \le n) = P\right\} \cdot 2^{-(n-1)}$$

$$\cdot \mathbf{E}\left\{ \prod_{i=1}^{n-1} \left(n^2 - \sum_{T\in F_i}|T|^2 \right) \,\middle|\, (P(F_i), 1 \le i \le n) = P \right\}$$

$$= 2^{-(n-1)} \cdot \mathbf{E}\left[\prod_{i=1}^{n-1} \left(n^2 - \sum_{T\in F_i}|T|^2 \right) \right]. \tag{2.3}$$

A mechanical modification of the logic leading to (2.3) yields the following expression, valid for each $1 \le k \le n$:

$$Z_{\mathrm{MC}}(n,k) = 2^{-(k-1)}\mathbf{E}\left[\prod_{i=1}^{k-1} \left(n^2 - \sum_{T\in F_i}|T|^2 \right) \right]. \tag{2.4}$$

Write

$$\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k) = \prod_{i=1}^{k-1} \left(n^2 - \sum_{T\in F_i}|T|^2 \right),$$

let $\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n) = \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,1)$, and let $\hat{Z}_{\mathrm{MC}}(n,k) = 2^{-(k-1)}\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k)$ and $\hat{Z}_{\mathrm{MC}}(n) = \hat{Z}_{\mathrm{MC}}(n,n)$. With this notation, (2.3) and the subsequent equation state that

$$\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n,k)\right] = Z_{\mathrm{MC}}(n,k) = \frac{1}{2^{k-1}}\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k)\right]. \tag{2.5}$$

The random variable $\hat{Z}_{\mathrm{MC}}(n)$ is a sort of *empirical partition function* of the multiplicative coalescent. The superscript arrow on $\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n,k)$ is because the factor 2^{k-1} may be viewed as corresponding to a choice of orientation for each edge of F_k. The random variable $\hat{Z}_{\mathrm{MC}}(n)$ of course contains more information than simply its expected value, so by studying it we might hope to gain a greater insight into the behaviour of the coalescent. Much of the remainder of these notes is devoted to showing that $\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n)\right] = Z_{\mathrm{MC}}(n)$ is a *terrible* predictor of the typical value of $\hat{Z}_{\mathrm{MC}}(n)$. More precisely, there are unlikely execution paths along which the multiplicative coalescent has many more possibilities than along a typical path; such paths swell the *expected* value of $\hat{Z}_{\mathrm{MC}}(n)$ to exponentially larger than its *typical* size.

The logic leading to (2.3) and (2.4) may also be applied to the additive coalescent; the result is boring but instructive. First note that

$$Z_{\mathrm{AC}}(n,k) = \sum_{P=(P_1,\ldots,P_k)\in\mathcal{P}_{n,k}} \prod_{i=1}^{k-1} \left(|A_i(P)| + |B_i(P)| \right).$$

For the additive coalescent, the total number of choices at step i is $n(n-i)$, and given that $P(F_i) = P_i$, the number of choices which yield $P(F_{i+1}) = P_{i+1}$ is $A_i(P) + B_i(P)$. Writing $\mathbf{P}_{\mathrm{AC}}$ for probabilities under the additive coalescent, we thus have

$$\mathbf{P}_{\mathrm{AC}} \left\{ (P(F_i), 1 \le i \le k) = (P_1, \dots, P_k) \right\} = \prod_{i=1}^{k-1} \frac{|A_i(P)| + |B_i(P)|}{n(n-i)}$$

Following the logic through yields

$$Z_{\mathrm{AC}}(n, k) = \mathbf{E}_{\mathrm{AC}} \left[\prod_{i=1}^{k-1} n(n-i) \right] = \mathbf{E}_{\mathrm{AC}} \left[n^{k-1}(n-1)_{k-1} \right].$$

Thus, the "empirical partition function" of the additive coalescent is a constant, so contains no information beyond its expected value. (This fact is essentially the key to Pitman's proof of Cayley's formula.)

The terms of the products (2.3) and (2.4), though random, turn out to behave in a very regular manner (but proving this will take some work). Through a study of these terms, we will obtain control of $\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n) \right]$, and thereby justify the above assertion that $\hat{Z}_{\mathrm{MC}}(n)$ is typically very different from its mean.

2.3.1. The growth rate of $Z_{\mathbf{mc}}(n, \lfloor n/2 \rfloor)$. As a warmup, and to introduce a key tool, we approximate the value of $Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)$ using a connection between the multiplicative coalescent and a process we call (once again with a very slight abuse of terminology) the Erdős-Rényi coalescent. Write K_n for the *complete graph*, i.e. the graph with vertices $[n]$ and edges $(\{i,j\}, 1 \le i < j \le n)$.

> **The Erdős-Rényi coalescent.** Choose a uniformly random permutation $e_1, \dots, e_{\binom{n}{2}}$ of $E(K_n)$. For $0 \le i \le \binom{n}{2}$, let $G_i^{(n)}$ have vertices $[n]$ and edges $\{e_1, \dots, e_i\}$.

Our indexing here starts at zero, unlike in the multiplicative coalescent; this is slightly unfortunate, but it is standard for the Erdős-Rényi graph process to index so that $G_i^{(n)}$ has i edges. This process is different from the previous coalescent processes, most notably because it creates graphs with cycles.

Note that we can recover the multiplicative coalescent from the Erdős-Rényi coalescent in the following way. Informally, simply ignore any edges added by the Erdős-Rényi coalescent that fail to join distinct components. More precisely, for each $0 \le m \le \binom{n}{2}$, let τ_m be the number of edges $\{U_i, V_i\}$, $0 < i \le m$ such that U_i and V_i lie in different components of $G_{i-1}^{(n)}$. (See Figure 4 for an example.) Observe that

$$\tau_m + 1 = \begin{cases} \tau_m & \text{if } G_{m+1}^{(n)} \text{ and } G_m^{(n)} \text{ have the same number of components} \\ \tau_m + 1 & \text{if } G_{m+1}^{(n)} \text{ has one fewer component than } G_m^{(n)}. \end{cases}$$
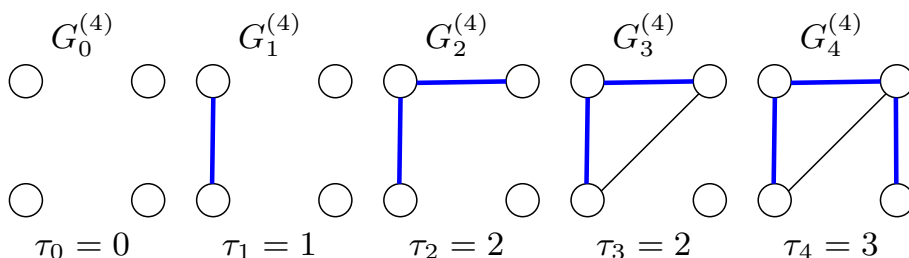
FIGURE 4. An example of the first steps of the Erdős-Rényi coalescent. The multiplicative coalescent is obtained by keeping only the thicker, blue edges.

In other words, τ_m increases precisely when the the endpoints of the edge added to $G_m^{(n)}$ are in different components. Further, the set

$$\{e_m : m \geq 1, \tau_m > \tau_{m-1}\}$$

contains $n-1$ edges, since $G_0^{(n)}$ has n components and $G_{\binom{n}{2}}^{(n)}$ almost surely has only one component.

Set $I_1 = 0$ and for $1 < k \leq n$ let

$$I_k = \inf\{m \geq 1 : \tau_m = k - 1\}.$$

Then for $1 < k \leq n$, the edge e_{I_k} joins distinct components of $G_{I_k-1}^{(n)}$, and by symmetry is equally likely to be any such edge. Thus, letting F_k be the graph with edges $\{e_{I_j} : 1 \leq j \leq k\}$ for $1 \leq k \leq n$, the process $\{F_k, 1 \leq k \leq n\}$ is precisely distributed as the multiplicative coalescent. This is a *coupling* between the Erdős-Rényi graph process and the multiplicative coalescent; its key property is that for all $1 \leq k \leq n$, the vertex sets of the trees of F_k are the same as those of the components of $G_{I_k}^{(n)}$.

Having found the multiplicative coalescent within the Erdős-Rényi coalescent, we can now use known results about the latter process to study the former. For a graph G, and $v \in V(G)$, we write $N(v) = N_G(v)$ for the set of nodes adjacent to v (the *neighbours* of v), and write $C(v) = C_G(v)$ for the connected component of G containing v. We will use the results of the following exercise.[3]

> **Exercise 9.** ⊛
>
> (a) Show that in the Erdős-Rényi coalescent, if all components have size at most s then the probability a uniformly random edge from among the remaining edges has both endpoints in the same component is at most $(s-1)/(n-1)$.
> (b) Show that for all $0 \leq m \leq n/2$, in $G_m^{(n)}$, $\mathbf{E}|N(v)| \leq 2m/n$.

---

[3]Until further notice, we omit ceilings and floors for readability.

(c) Prove by induction that for all $0 \leq m < n/2$, in $G_m^{(n)}$, $\mathbf{E}\left[|C(1)|\right] \leq n/(n - 2m)$.

(**Hint.** First condition on $N(1)$, then average.)

(d) Prove that for all $\epsilon > 0$,

$$\limsup_{n \to \infty} \mathbb{P}\left(G_{(1-\epsilon)n/2}^{(n)} \text{ has a component of size } > \epsilon n\right) = 0.$$

(**Hint.** Given that the largest component of $G_m^{(n)}$ has size s, with probability at least s/n vertex 1 is in such a component.)

Using the above exercise, we now fairly easily prove a lower bound on the partition function of the first half of the multiplicative coalescent.

Proposition 2.3. *For all $\beta > 0$,*

$$\mathbf{P}\left\{\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n, \lfloor n/2 \rfloor) \geq n^{(1-\beta)n}\right\} \to 1 \ \ as \ n \to \infty.$$

We begin by showing that typically $I_t = (1 + o(1))t$ until $t \geq n/2$.

Lemma 2.4. *For all $\epsilon > 0$, $\limsup_{n \to \infty} \mathbb{P}\left(I_{(1-\epsilon)n/2} \geq n/2\right) = 0$.*

Proof. Fix $\epsilon > 0$, let $\delta = \epsilon/3$, and let E be the event that all components of $G_{(1-\delta)n/2}$ have size at most δn. For $m \geq 0$, conditional on $G_m^{(n)}$, by Exercise 9 (a), $\tau_{m+1} - \tau_m$ stochastically dominates a Bernoulli$(1-(s-1)/(n-1))$ random variable, where s is the size of the largest component of $G_m^{(n)}$.

For n large and $s \leq \delta n$ we have $1 - (s - 1)/(n - 1) \geq 1 - \epsilon/2$. Therefore, on E and for large n the sequence $(\tau_{m+1} - \tau_m, 0 \leq m < (1 - \delta)n/2)$ stochastically dominates a sequence $(B_m, 0 \leq m < (1 - \delta)n/2)$ of iid Bernoulli$(1 - \epsilon/2)$ random variables. It follows that

$$\mathbf{P}\left\{\tau_{(1-\delta)n/2} \leq (1 - \epsilon)n/2\right\} \leq \mathbf{P}\left\{E^c\right\} + \mathbf{P}\left\{\tau_{(1-\delta)n/2} \leq (1 - \epsilon)n/2 \mid E\right\}$$
$$\leq \mathbf{P}\left\{E^c\right\} + \mathbf{P}\left\{\mathrm{Bin}((1 - \delta)n/2, 1 - \epsilon/2) < (1 - \epsilon)n/2\right\}$$
$$= o(1),$$

the last line Exercise 9 (d) and Chebyshev's inequality (note that $(1 - \delta)(1 - \epsilon/2)n/2 > (1 - 5\epsilon/6)n/2$). On the other hand, if $\tau_{(1-\delta)n/2} > (1 - \epsilon)n/2$ then $I_{(1-\epsilon)n/2} \leq (1 - \delta)n/2 < n/2$. $\qquad\square$

Proof of Proposition 2.3. View $(F_1, \ldots, F_n)$ as coupled with the Erdős-Rényi coalescent as above, so that F_k and $G_{I_k}^{(n)}$ have the same components. Fix $\delta \in (0, 1/4)$ and let $k = k(n) = n/2 - 2\delta n$. Let E_1 be the event that $I_{n/2-\delta n} < n/2$.[4] Since $I_{m+1} \geq I_m + 1$ for all m, we have

$$I_k \leq I_{n/2-\delta n} - ((n/2 - \delta n) - k) = I_{n/2-\delta n} - \delta n.$$

Thus, on E_1 we have $I_k \leq (1 - 2\delta)n/2$.

---

[4]We omit the dependence on n in the notation for E_1; similar infractions occur later in the proof.

Next let E_2 be the event that all component sizes in $G^{(n)}_{(1-2\delta)n/2}$ are at most δn. The components of F_k are precisely the components of $G^{(n)}_{I_k}$, so if $E_1 \cap E_2$ occurs then since on E_1 we have $I_k \leq (1-2\delta)n/2$, all components of F_k have size at most δn. In this case, for all $i \leq k$ the components of F_i clearly also have size at most δn.

It follows[5] that on $E_1 \cap E_2$, for all $i \leq k$,

$$\sum_{T \in F_i} |T|^2 \leq \delta n^2$$

so on $E_1 \cap E_2$,

$$\hat{Z}^{\rightarrow}_{\mathrm{MC}}(n, k+1) = \prod_{i=1}^{k} \left(n^2 - \sum_{T \in F_i} |T|^2 \right)$$
$$\geq n^{2k}(1-\delta)^k \qquad\qquad (2.6)$$
$$= n^{n(1-4\delta)} \cdot (1-\delta)^{n/2} \qquad\qquad > n^{n(1-5\delta)},$$

the last inequality holding for n large. By Exercise 9 (d) and Lemma 2.4, $\mathbb{P}(E_1 \cap E_2) \rightarrow 1$ as $n \rightarrow \infty$. Since $\hat{Z}^{\rightarrow}_{\mathrm{MC}}(n, \lfloor n/2 \rfloor) \geq \hat{Z}^{\rightarrow}_{\mathrm{MC}}(n, k+1)$ for n large, the result follows. □

The following exercise is to test whether you are awake.

Exercise 10. Prove that
$$\frac{\log Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)}{n \log n} \rightarrow 1 \,,$$
as $n \rightarrow \infty$.

We next use Proposition 2.3 (more precisely, the inequality (2.6) obtained in the course of its proof) to obtain a first lower bound on $Z_{\mathrm{MC}}(n)$.

Corollary 2.5. *It holds that*
$$\frac{Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)}{Z_{\mathrm{AC}}(n, \lfloor n/2 \rfloor)} = \left(\frac{e}{4} \right)^{(1+o(1))n/2} \,.$$

Proof. By Proposition 2.3 and (2.5), we have
$$Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor) \geq 2^{-(\lfloor n/2 \rfloor - 1)} n^{(1+o(1))n},$$

so by Exercise 3,
$$\frac{Z_{\mathrm{MC}}(n, \lfloor n/2 \rfloor)}{Z_{\mathrm{AC}}(n, \lfloor n/2 \rfloor)} = \frac{n^{(1+o(1))n}}{2^{\lfloor n/2 \rfloor - 1} n^{\lfloor n/2 \rfloor - 1}(n-1)_{\lfloor n/2 \rfloor - 1}} = \frac{n^{(1+o(1))n}(n/2)!}{2^{n/2} n^{n/2} n!} \,.$$

---

[5]To maximize $\sum_j x_j^2$ subject to the conditions that $\sum_j x_j = 1$ and that $\max_j x_j \leq \delta$, take $x_j = \delta$ for $1 \leq j \leq \delta^{-1}$.

Using Stirling's approximation[6], it follows easily that
$$\frac{Z_{\text{MC}}(n, \lfloor n/2 \rfloor)}{Z_{\text{AC}}(n, \lfloor n/2 \rfloor)} \geq \left(\frac{e}{4}\right)^{(1+o(1))n/2}.$$
The corresponding upper bound follows similarly, using that $Z_{\text{MC}}(n, \lfloor n/2 \rfloor) \leq (n(n-1)/2)^{\lfloor n/2 \rfloor - 1} = n^{n(1+o(1))}/2^{n/2}$. □

Exercise 11. Perform the omitted calculation using Stirling's formula from the proof of Corollary 2.5.

The preceding corollary is evidence that despite the similarity of the partition functions $Z_{\text{MC}}(n)$ and $Z_{\text{AC}}(n)$, the fine structure of the multiplicative coalescent is may be interestingly different from that of the additive coalescent.

2.3.2. The multiplicative coalescent and Kruskal's algorithm. There is a pleasing interpretation of "Version 2" of the multiplicative coalescent, which is driven by exchangeable distinct edge weights $\mathbf{W} = \{W_{\{j,k\}}, 1 \leq j < k \leq n\} = \{W_e, e \in E(K_n)\}$. (A special case is that the elements of $\mathbf{W}$ are iid continuous random variables.). The symmetry of the model makes it straightforward to verify that this results in a sequence $(F_1, \ldots, F_n)$ with the same distribution as the multiplicative coalescent.

Multiplicative Coalescent Version 2: Kruskal's algorithm.
Let F_1 be a forest of n isolated vertices $1, \ldots, n$.
For $1 \leq i < n$:
⋆ Let $\{j, k\} \in E(K_n)$ minimize $\{W_{\{j,k\}} : j, k$ in distinct trees of $F_i\}$.
⋆ Form F_{i+1} from F_i by adding $\{j, k\}$.

Exercise 12. ⊛ Prove that any exchangeable, distinct edge weights $\mathbf{W} = \{W_e, e \in E(K_n)\}$ again yield a process with the law of the multiplicative coalescent.

At step i, the edge-weight driven multiplicative coalescent simply adds the smallest weight edge whose endpoints lie in distinct components of F_i. In other words, it adds the smallest weight edge whose addition will not create a *cycle* in the growing graph. This is simply *Kruskal's algorithm* for building the minimum weight spanning tree. When the weights $W_{\{j,k\}}$ are all non-negative, the tree obtained at the end of the Version 2 multiplicative coalescent, $T_1^{(n)}$, is the minimum weight spanning tree of K_n with weights $\mathbf{W}$. We denote it $\text{MST}(K_n, \mathbf{W})$, and refer to it as the *random MST of K_n*.

Order $E(K_n)$ by increasing order of $\mathbf{W}$-weight as $e_1, \ldots, e_{\binom{n}{2}}$. The exchangeability of $\mathbf{W}$ implies this is a uniformly random permutation of $E(K_n)$. Letting

---
[6]Stirling's approximation says that $m!/(\sqrt{2\pi m}(m/e)^m) \to 1$ as $m \to \infty$; in fact the (much less precise) fact that $\log(m!) = m \log m - m + o(m)$ is enough for the current situation.

$G_k^{(n)}$ have edges $e_1, \ldots, e_k$ thus yields an important instantiation of our coupling of the Erdős-Rényi coalescent and the multiplicative coalescent; we return to this in Section 5.

2.3.3. Other features of the multiplicative coalescent. The remainder of the section is not essential to the main development. The following exercise was inspired by a discussion with Remco van der Hofstad.

> **Exercise 13 (First-passage percolation).** Consider the multiplicative coalescent driven by exchangeable, distinct edge weights $\mathbf{W}$ and for $1 \le i < j \le n$, let $d(i,j) = \min \sum_{e \in \gamma} W_e$, the minimum taken over paths from i to j in K_n. Show that the minimum is attained by a unique path $\gamma_{i,j}$. Find exchangeable edge weights $\{W_e, e \in E(K_n)\}$ for which, for each for each $1 \le i < j \le n$, $\gamma_{i,j}$ is a path of $T_1^{(n)}$.

Finally, we turn to Version 3 of the process, in which we view arbitrary iid non-negative weights $\mathbf{X} = \{X_{i,j}, 1 \le i < j \le n\}$ as *rates* for edge addition. In view of the preceding paragraph, this gives a process that results in a tree with the same *distribution* as the random MST of K_n, but which is not necessarily equal to the MST. In particular, the tree is *not* a deterministic function of the edge weights; for example, we may take $\mathbf{X}$ to be a deterministic vector such as the all-ones vector, whereas the resulting tree always is random.

> **Exercise 14.** Find (iid random) *rates* $\mathbf{X}$ for which, in version 3 of the process, the resulting tree $T_1^{(n)}$ is equal to the random MST of K_n with *weights* $\mathbf{X}$, with probability tending to one as $n \to \infty$.

3. Intermezzo: The heights of the three coalescent trees

To date we have been primarily studying the partition functions of the coalescent processes. The processes have many other interesting features, however. In this section we discuss differences between the structures of the trees formed by the three coalescents.

Write $T_{\mathrm{KC}}^{(n)}, T_{\mathrm{AC}}^{(n)}$, and $T_{\mathrm{MC}}^{(n)}$, respectively, for the trees formed by Kingman's coalescent, the additive coalescent, and the multiplicative coalescent. In each case the coalescent starts from n isolated vertices $\{1, \ldots, n\}$, so each of these trees has vertices $\{1, \ldots, n\}$. If T is any of these trees and e is an edge of T, we write $L(e) = i$ if e was the i'th edge added during the execution of the coalescent. Above, we established the following facts about the distributions of these random trees.

1. Ignoring vertex labels, $(T_{\mathrm{KC}}^{(n)}, L)$ is uniformly distributed over pairs (t, ℓ), where t is a rooted tree with n vertices and ℓ is a decreasing edge labelling

of t. (We simply refer to such pairs as *decreasing trees with n vertices*, for short.)

2. $T_{\mathrm{AC}}^{(n)}$ is uniformly distributed over the set of rooted trees with vertices $\{1, \ldots, n\}$. (We refer to such trees as *rooted labeled trees with n vertices*.)

3. $T_{\mathrm{MC}}^{(n)}$ is distributed as the minimum weight spanning tree of the complete graph K_n, with iid continuous edge weights $\mathbf{W} = \{W_{i,j}, 1 \leq i < j \leq n\}$.

What is known about these three distributions? To illustrate the difference between them, we consider a fundamental tree parameter, the *height*: this is simply the greatest number of edges in any path starting from the root.[7] The third tree, $T_{\mathrm{MC}}^{(n)}$ is not naturally rooted, but one may check that any choice of root will yield the same height up to a multiplicative factor of two; we root $T_{\mathrm{MC}}^{(n)}$ at vertex 1 by convention. Given a rooted tree t, we write $r(t)$ for its root and $h(t)$ for its height. The following exercise develops a fairly straightforward route to upper bounds on $h(T_{\mathrm{KC}}^{(n)})$ that are tight, at least to first order.

> **Exercise 15.** Let D_i be the number of edges on the path from vertex i to $r(T_{\mathrm{KC}}^{(n)})$.
> (a) Show that $\{D_1, \ldots, D_n\}$ are exchangeable random variables.
> (b) (i) Show that D_1 can be written as a sum of independent Bernoulli random variables.
>   (ii) For $\alpha > 0$, find λ such that Bernoulli$(\alpha) \preceq_{\mathrm{st}}$ Poisson(λ).
>   (iii) Show that D_1 is stochastically dominated by a Poisson$(\log n)$ random variable.
> (c) Show that for X a Poisson(μ) random variable, for integer $x > \mu$,
> $$\mathbf{P}\{X > x\} \leq \frac{e^{-\mu}\mu^x}{(x-\mu)(x-1)!}.$$
> (d) Show that $\mathbf{P}\{\max_{1 \leq i \leq n} D_i \geq e \log n\} \to 0$ as $n \to \infty$.
> (e) Show that $\limsup_{n \to \infty}(\max_{1 \leq i \leq n} D_i - e \log n) \to -\infty$ in probability.

We next turn to $T_{\mathrm{AC}}^{(n)}$. I am not aware of an easy way to directly use the additive coalescent to analyze the height of $T_{\mathrm{AC}}^{(n)}$. However, one can use the additive coalescent to derive combinatorial results which, together with exchangeability, yields lower bounds of the correct order of magnitude, and upper bounds that are tight up to poly-logarithmic corrections; such bounds are the content of the following exercise. A non-negative random variable R has the *standard Rayleigh* distribution if it has density $f(x) = xe^{-x^2/2}$ on $[0, \infty)$.

> **Exercise 16.** Let D_i be the number of edges on the path from vertex i to $r(T_{\mathrm{AC}}^{(n)})$.
> (a) Show that $\{D_1, \ldots, D_n\}$ are exchangeable random variables.
> (b) Show that the number of pairs (t, i), where t is a rooted labeled tree with $V(t) = [n]$ and $i \in V(t)$ has $d(r(t), i) = k - 1$, is $k \cdot (n)_k \cdot n^{n-k-1}$.

---

[7]A glance back at Figures 1, 2 and 3 gives a hint as to the relative heights of the three trees.

(c) Show that for $1 \le k \le n$, $\mathbf{P}\{D_1 = k - 1\} = \frac{k}{n} \prod_{i=1}^{k-1}\left(1 - \frac{i}{n}\right)$. Conclude that $D_1/\sqrt{n}$ converges in distribution to a standard Rayleigh.

(d) Using (c) and a union bound, show that if $(c_n, n \ge 1)$ are constants with $c_n \to \infty$ then $\mathbf{P}\left\{\max_{1 \le i \le n} D_i > c_n \sqrt{n \log n}\right\} \to 0$.

(e) Use the exchangeability of the trees in a uniformly random ordered labeled forest to prove that $\mathbf{P}\{|\{i : D_i \ge k/2\}| \ge n/2 \mid D_1 = k\} \ge 1/2$ for all $1 \le k \le n$.

(f) Use (c) and (e) to show that if $(c_n, n \ge 1)$ are constants with $c_n \to \infty$ then $\mathbf{P}\left\{\max_{1 \le i \le n} D_i > c_n \sqrt{n}\right\} \to 0$, strengthening the result from (d).

From the preceding exercise, we see immediately that $T_{\mathrm{AC}}^{(n)}$ has a very different structure from $T_{\mathrm{KC}}^{(n)}$, which had logarithmic height. Moreover, the heights of the two trees are *qualitatively* different. The height of $T_{\mathrm{KC}}^{(n)}$ is concentrated: $h(T_{\mathrm{KC}}^{(n)})/\log n \to e$ in probability. On the other hand, $h(T_{\mathrm{AC}}^{(n)})$ is diffuse: $h(T_{\mathrm{AC}}^{(n)})/n^{1/2}$ converges in distribution to a non-negative random variable with a density.[8] [9]

What about the tree $T_{\mathrm{MC}}^{(n)}$ built by the multiplicative coalescent? Probabilistically, this is the most challenging of the three to study. For $T_{\mathrm{KC}}^{(n)}$ and $T_{\mathrm{AC}}^{(n)}$, Exercises 15 and 16 yielded exact or nearly exact expressions for the distance between the root and a fixed vertex (by exchangeability, this is equivalent to the distance between the root and a *uniformly random* vertex). The partition function $Z_{\mathrm{MC}}(n)$ seems too complex for such a direct argument to be feasible.

The coalescent procedure can be used to obtain lower bounds on the height, but with greater effort than in the two preceding cases. Our approach is elucidated by the following somewhat challenging exercise. Let K_n have iid Exponential$[0, 1]$ edge weights, and let H be the subgraph of K_n with the same vertices, but containing only edges of weight at most $1/n$. A *tree component* of H is a connected component of H that is a tree.

Exercise 17.

(a) Let N be the number of vertices in tree components of H of size at most $\lfloor n^{1/4} \rfloor$. Using Chebyshev's inequality, show that $\mathbf{P}\{N = 0\} \to 0$ as $n \to \infty$.

(b) Fix $S \subset \{1, \ldots, n\}$. Show that, given that H contains a tree component whose vertices are precisely S, then such a component is uniformly distributed over labeled trees with vertices S.

(c) Use Kruskal's algorithm to show that any tree component of H is a subtree of the minimum weight spanning tree of $T_{\mathrm{MC}}^{(n)}$.

---

[8]Neither of these convergence statements follows from the exercises, and both require some work to prove. The fact that $h(T_{\mathrm{KC}}^{(n)})/\log n \to e$ in probability was first shown by Devroye [8]. The distributional convergence of $h(T_{\mathrm{AC}}^{(n)})/n^{1/2}$ is a result of Rényi and Szekeres [15].

[9]In fact, if edge lengths in $T_{\mathrm{AC}}^{(n)}$ are multiplied by $n^{-1/2}$ then the resulting object converges in distribution to a random compact metric space called the *Brownian continuum random tree* (or *CRT*), and $h(T_{\mathrm{AC}})/n^{1/2}$ converges in distribution to the height of the CRT. For more on this important result, we refer the reader to [4, 10]

(d) Use Exercise 16 (c) to conclude that, as $n \to \infty$,

$$\mathbf{P}\left\{T_{\mathrm{MC}}^{(n)} \text{ has height at most } \frac{n^{1/8}}{\log^2 n}\right\} \to 0\,.$$

This shows that $T_{\mathrm{MC}}^{(n)}$ is quite different from $T_{\mathrm{KC}}^{(n)}$.[10] It is not as straightforward to bound the height of $T_{\mathrm{MC}}^{(n)}$ away from $n^{1/2}$ using the tools currently at our disposal. It turns out that $T_{\mathrm{MC}}^{(n)}$ has height of order $n^{1/3}$ (and has non-trivial fluctuations on this scale), but proving this takes a fair amount of work [1] and is beyond the scope of these notes.

Exercise 18 (Open problem – two point function of random MSTs). Let D_n be the distance from vertex n to vertex 1 in $T_{\mathrm{MC}}^{(n)}$. Obtain an explicit expression for the distributional limit of $D_n/n^{1/3}$.

4. The susceptibility process.

The remainder of the paper focusses exclusively on the multiplicative coalescent, which we continue to denote $(F_1, \ldots, F_n)$. Recall that $\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n, k) = \prod_{i=1}^{k-1}\left(n^2 - \sum_{T \in F_i} |T|^2\right)$. The terms in the preceding product are not independent; linearity of expectation makes the "empirical entropy" $\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)$ easier to study.

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)\right] = \sum_{k=1}^{n-1} \mathbf{E}\left[\log(n^2 - \sum_{T \in F_k} |T|^2)\right]. \tag{4.1}$$

The expectation in the latter sum is closely related to the *susceptibility* of the forest F_i. More precisely, given a finite graph G, write $\mathcal{C}(G)$ for the set of connected components of G. The susceptibility of G is the quantity

$$\chi(G) = \frac{1}{|G|} \sum_{C \in \mathcal{C}(G)} |C|^2.$$

Recalling that $C(v) = C_G(v)$ is the component of G containing v, we may also write $\chi(G) = |G|^{-1} \sum_{v \in V(G)} |C(v)|$, so $\chi(G)$ is the expected size of the component containing a uniformly random vertex from G.

Exercise 19. Let G be any graph, write L and S for the number of vertices in the largest and second-largest components of G, respectively. Then

$$\frac{L^2}{|G|} \le \chi(G) \le \frac{L^2}{|G|} + S.$$

---

[10]With more care, one can show that with high probability H contains tree components containing around $n^{2/3}$ vertices and with height around $n^{1/3}$, which yields that with high probability $T_{\mathrm{MC}}^{(n)}$ has height of order at least $n^{1/3}$.

Viewing F_i as a graph with vertices $\{1, \ldots, n\}$, (4.1) becomes

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)\right] = 2(n-1)\log n + \mathbf{E}\left[\sum_{k=1}^{n-1}\log\left(1 - \frac{\chi(F_k)}{n}\right)\right]. \qquad (4.2)$$

In order to analyze this expression, we use the connection with the Erdős-Rényi coalescent $(G_m^{(n)}, 0 \le m \le \binom{n}{2})$, which we described in Section 2.3.1; in brief, we coupled to $(F_k, 1 \le k \le n)$ by letting F_k have edges $\{e_{I_j}, 1 \le j \le k\}$, where I_k was the first time m that $G_m^{(n)}$ had $n+1-k$ components. In the next proposition and throughout the remainder of the paper, we interpret $x\log x$ as 0 when $x = 1$.

Proposition 4.1.

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)\right] = 2(n-1)\log n +$$

$$\sum_{m=0}^{\binom{n}{2}-1}\left(1 - \frac{n+2m}{n^2}\right)^{-1}\mathbf{E}\left[\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\cdot\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\right].$$

Proof. In the coupling with the Erdős-Rényi coalescent, F_k and $G_{I_k}^{(n)}$ have the same connected components, so $\chi(F_k) = \chi(G_{I_k}^{(n)})$. We obtain the identity

$$\sum_{k=1}^{n-1}\log\left(1 - \frac{\chi(F_k)}{n}\right) = \sum_{k=1}^{n-1}\log\left(1 - \frac{\chi(G_{I_k}^{(n)})}{n}\right)$$

$$= \sum_{m=0}^{\binom{n}{2}-1}\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\mathbf{1}_{[\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})]}$$

Using the tower law for conditional expectations, we thus have

$$\mathbf{E}\left[\sum_{k=1}^{n-1}\log\left(1 - \frac{\chi(F_k)}{n}\right)\right]$$

$$= \sum_{m=0}^{\binom{n}{2}-1}\mathbf{E}\left[\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\mathbf{1}_{[\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})]}\right].$$

$$= \sum_{m=0}^{\binom{n}{2}-1}\mathbf{E}\left[\mathbf{E}\left[\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\mathbf{1}_{[\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})]} \mid G_m^{(n)}\right]\right]$$

$$= \sum_{m=0}^{\binom{n}{2}-1}\mathbf{E}\left[\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\cdot\mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)}) \mid G_m^{(n)}\right\}\right]$$

For any finite graph G, the quantity $\chi(G)/|G| = \sum_{C \in \mathcal{C}(G)}|C|^2/|G|^2$ is simply the probability that a pair (U, V) of independent, uniformly random vertices of G lie in

the same component of G. Let (U, V) be independent, uniformly random elements of $[n] = V(G^{(n)}_{m-1})$. Then

$$\mathbf{P}\left\{U \neq V, \{U, V\} \notin E(G^{(n)}_m) \mid G^{(n)}_m\right\} = 1 - \frac{1}{n} - \frac{2|E(G^{(n)}_m)|}{n^2} = 1 - \frac{n + 2m}{n^2}$$

Conditionally given that $U \neq V$ and $\{U, V\} \notin E(G^{(n)}_m)$, the pair $\{U, V\}$ has the same law as e_{m+1}. It follows that

$$\mathbf{P}\left\{\chi(G^{(n)}_{m+1}) > \chi(G^{(n)}_m) \mid G^{(n)}_m\right\}$$

$$= \mathbf{P}\left\{C_{G^{(n)}_m}(U) \neq C_{G^{(n)}_m}(V) \mid G^{(n)}_m, U \neq V, \{U, V\} \notin E(G^{(n)}_m)\right\}$$

$$= \left(1 - \frac{\chi(G^{(n)}_{m-1})}{n}\right)\left(1 - \frac{n + 2m}{n^2}\right)^{-1}, \tag{4.3}$$

so

$$\mathbf{E}\left[\sum_{k=1}^{n-1} \log\left(1 - \frac{\chi(F_k)}{n}\right)\right]$$

$$= \sum_{m=0}^{\binom{n}{2}-1} \left(1 - \frac{n + 2m}{n^2}\right)^{-1} \mathbf{E}\left[\log\left(1 - \frac{\chi(G^{(n)}_m)}{n}\right) \cdot \left(1 - \frac{\chi(G^{(n)}_m)}{n}\right)\right]. \tag{4.4}$$

The proposition now follows from (4.2). $\qquad\qquad\qquad\square$

It turns out that there is a deterministic, increasing function $f : [0, \infty) \to [0, 1]$ such that $\sup_{0 \leq m < \binom{n}{2}} |\chi(G^{(n)}_m)/n - f(m/n)| \to 0$ in probability, as $n \to \infty$. Much of the rest of the paper is devoted to explaining this fact in more detail. However, imagine for the moment that such a function f exists and, moreover, that terms in the sum with $m \gg n$ have an insignificant total contribution. With these assumptions, the sum in (4.4) looks like a Riemann approximation for $\int_0^\infty (1 - f(x)) \log(1 - f(x))\mathrm{d}x$ with mesh $1/n$. We should then expect that

$$\mathbf{E}\left[\log \hat{Z}^{\to}_{\mathrm{MC}}(n)\right] = 2(n - 1)\log n - (1 + o(1))n \cdot \int_0^\infty (1 - f(x))\log(1 - f(x))\mathrm{d}x\,.$$

This is indeed the case. Furthermore, enough is known about f that explicit evaluation of the integral is possible, and we obtain the following theorem.

Theorem 4.2. *Let*

$$\zeta_{\mathrm{MC}} = \zeta(2) - 3 + \log 2 - \log^2 2\,. \tag{4.5}$$

Then

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n)\right] = n \cdot (2\log n + \zeta_{\mathrm{MC}} + o(1)).$$

Numerically, ζ_{MC} is around -1.14237.

Corollary 4.3. *There is $c > 0$ such that* $\mathbf{P}\left\{\hat{Z}_{\mathrm{MC}}(n)/\mathbf{E}\hat{Z}_{\mathrm{MC}}(n) < e^{-cn}\right\} \to 1$ *as* $n \to \infty$.

Proof. Fix $c \in \mathbb{R}$ and suppose that $\mathbf{P}\left\{\hat{Z}_{\mathrm{MC}}(n) \geq n^{2n}e^{cn}\right\} > \epsilon > 0$. Then

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n) - 2n \log n - cn + 1\right]$$

$$\geq \epsilon \mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n) - 2n \log n + cn + 1 \mid \hat{Z}_{\mathrm{MC}}(n) \geq n^{2n}e^{-cn}\right]$$

$$\geq \epsilon.$$

Thus, if $\liminf_{n\to\infty} \mathbf{P}\left\{\hat{Z}_{\mathrm{MC}}(n) \geq n^{2n}e^{cn}\right\} > 0$ then for all n large enough,

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n)\right] > 2n \log n + cn - 1\,.$$

It thus follows from Theorem 4.2 that for all $\epsilon > 0$,

$$\mathbf{P}\left\{\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n) \geq n^{2n}e^{(\zeta_{\mathrm{MC}}^{\rightarrow}+\epsilon)n}\right\} \to 0$$

as $n \to \infty$. On the other hand,

$$\mathbf{E}\hat{Z}_{\mathrm{MC}}^{\rightarrow}(n) = n^{n-2}(n-1)! = n^{2n}e^{-n(1+o(1))}\,.$$

Since $\zeta_{\mathrm{MC}} < -1$, the result follows. $\square$

The form of the constant ζ_{MC} is unimportant, though intriguing. What is clear from the above is that information about the susceptibility process of the multiplicative coalescent yields bounds for $\hat{Z}_{\mathrm{MC}}(n)$. The aim of the next section is thus to understand the susceptibility process in more detail.

4.1. Bounding χ using a graph exploration

The coupling between the "Version 2" multiplicative coalescent (Kruskal's algorithm) and the Erdős-Rényi coalescent from Section 2.3.2 applied to arbitrary exchangeable, distinct edge weights $\mathbf{W}$. In this coupling, for $m \in [\binom{n}{2}]$, we took $G_m^{(n)}$ to be the subgraph of K_n consisting of the m edges of smallest $\mathbf{W}$-weight.

In the current section, it is useful to be more specific. We suppose the entries of $\mathbf{W}$ are iid Uniform$[0,1]$ random variables. Write $G(n,p)$ for the graph with vertices $[n]$ and edges $\{e_j : W_{e_j} \leq p\}$. In $G(n,p)$, each edge of K_n is independently present with probability p. Furthermore, we have $G(n,p) = G_{m_p}^{(n)}$, where $m_p = \max\{i : W_{e_i} \leq p\}$, so this also couples the Erdős-Rényi coalescent with the process $(G(n,p), 0 \leq p \leq 1)$. The next exercise is standard, but important.

Exercise 20. ⊛ Show that for any $p \in (0,1)$ and $m \in \binom{n}{2}$, given that $|E(G(n,p))| = m$, the conditional distribution of $G(n,p)$ is the same as that of $G_m^{(n)}$.

For $c > 0$, let $\alpha = \alpha(c)$ be the largest real solution of $e^{-cx} = 1 - x$. The aim of this section is to prove the following result.

Theorem 4.4. *For all $n \geq 1$ and $0 \leq p \leq n^{-19/20}$,*

$$\mathbf{P}\left\{|\chi(G(n,p)) - \alpha(np)^2 n| > 22n^{4/5}\right\} < 6ne^{-12n^{1/10}}$$

The coupling with the Erdős-Rényi coalescent will allow us to derive corresponding results for $G_m^{(n)}$. While the ingredients for the proof are all in the literature, and closely related results have certainly appeared in many places, we were unable to find a reference for the form we require. Some of the basic calculations required for the proof appear as exercises; the first such exercise relates to properties of the function α.

> **Exercise 21.** ⊛
> (a) Show that α is continuous and that α is concave and strictly positive on $(1, \infty)$.
> (b) Show that for $0 < c \leq 1$, $\alpha(c) = 0$, and for $c \geq 2$, $1 - 2e^{-c} \leq \alpha(c) \leq 1 - e^{-c}$.
> (c) Show that $\alpha(c)$ is increasing and $c(1 - \alpha(c))$ is decreasing.
> (d) Show that $\alpha'(c) < 2$ for all $c > 1$ and that $\frac{\mathrm{d}}{\mathrm{d}c}\alpha(c) \uparrow 2$ as $c \downarrow 1$. Conclude that $2\epsilon(1 - o(1)) \leq \alpha(1 + \epsilon) \leq 2\epsilon$, the first inequality holding as $\epsilon \downarrow 0$.
> (e) Show that $\alpha(c)$ is the survival probability of a Poisson(c) branching process. *(This exercise is not used directly.)*

Our proof of Theorem 4.4 hinges on a variant of the well-known and well-used depth-first search exploration procedure. In depth-first search, at each step one vertex is "explored": its neighbours are revealed, and those neighbours lying in the undiscovered region of the graph are added to the "depth-first search queue" for later exploration. In our variant, if the queue is ever empty, in the next step we add each undiscovered vertex to the queue independently with probability p. (It is more standard to add a *single* undiscovered vertex, but adding randomness turns out to simplify the formula for the expected number of unexplored vertices.)

We now formally state our search procedure for $G(n, p)$. At step i the vertex set $[n]$ is partitioned into sets E_i, D_i and U_i, respectively containing *explored, discovered*, and *undiscovered* vertices. We always begin with $E_0 = \emptyset$, $D_0 = \{1\}$, and $U_0 = [n] \setminus \{1\}$. For a set S, we write $\mathrm{Bin}(S, p)$ to denote a random subset of S which contains each element of S independently with probability p. For $v \in [n]$ we write $N(v)$ for the neighbours of v in $G(n, p)$. Finally, we define the *priority* of a vertex $v \in [n]$ to be its time of discovery $\inf\{j : v \in D_j\}$, so that vertices that are discovered later have higher priority.

> **Search process for $G(n, p)$.**
> STEP i:
> ⋆ If $D_i \neq \emptyset$ then choose $v \in D_i$ with highest priority (if there is a tie, pick the vertex with smallest label among highest-priority vertices). Let $E_{i+1} = E_i \cup \{v\}$, let $D_{i+1} = (D_i \cup (N(v) \cap U_i)) \setminus \{v\}$ and let $U_{i+1} = U_i \setminus (N(v) \cap U_i)$.

> ⋆ If $D_i = \emptyset$ then let $D_{i+1} = \mathrm{Bin}(U_i, p)$, independently of all previous steps. Let $E_{i+1} = E_i$ and let $U_{i+1} = U_i \setminus D_{i+1}$.

Observe that the sequence $((D_i, E_i, U_i), i \geq 0)$ describing the process may be recovered from either $(D_i, i \geq 0)$ or $(U_i, i \geq 0)$. The order of exploration yields the following property of the search process. Suppose $D_i = \emptyset$ for a given i. Then D_{i+1} may contain several nodes, all of which have priority $(i+1)$. Starting at step $(i+1)$, the search process will fully explore the component containing the smallest labelled vertex of D_{i+1} before exploring any vertex in any other component. More strongly, the search process will explore the components that intersect D_{i+1} in order of their smallest labeled vertices.

For $i > 0$ such that $E_i \neq E_{i-1}$, write v_i for the unique element of $E_i \setminus E_{i-1}$. Say that a component exploration concludes at time t if v_{t+1} and v_t are in distinct components of $G(n, p)$. The observation of the preceding paragraph implies the following fact about the search process. Set $D_0 = \emptyset$ for convenience.

Fact 4.5. Fix $t > 0$ and let $i = i(t) = \max\{j < t : D_j = \emptyset\}$. If a component exploration concludes at time t then $|D_{i+1}| \geq n - t - |U_t|$.

Proof. Since a component exploration concludes at time t we have $D_t \subset D_{i+1}$. Furthermore, $|E_t| \leq t$ because $|E_0| = \emptyset$ and $|E_{j+1} \setminus E_j| \leq 1$ for all $j \geq 0$. As D_t, E_t and U_t partition $[n]$, we thus have

$$|U_t| = n - |E_t| - |D_t| \geq n - t - |D_{i+1}|. \qquad \square$$

In proving Theorem 4.4 we use a concentration inequality due to McDiarmid [11]. Let $\mathrm{X} = (X_i, 1 \leq i \leq m)$ be independent Bernoulli(q) random variables. Suppose that $f : \{0, 1\}^m$ is such that for all $1 \leq k \leq m$, for all $(x_1, \ldots, x_k) \in \{0, 1\}^k$,

$$|\mathbf{E}\left[f(x_1, \ldots, x_k, X_{k+1}, \ldots, X_m)\right] - \mathbf{E}\left[f(x_1, \ldots, 1 - x_k, X_{k+1}, \ldots, X_m)\right]| \leq 1.$$

In other words, given the values of the first $k-1$ variables, knowledge of the k'th variable changes the conditional expectation by at most one.

Theorem 4.6 (McDiarmid's inequality). *Let* X *and* f *be as above. Write* $\mu = \mathbf{E}\left[f(\mathrm{X})\right]$. *Then for* $x > 0$,

$$\mathbf{P}\left\{f(X) \geq \mu + x\right\} \leq e^{-x^2/(2mq + 2x/3)}, \quad \mathbf{P}\left\{f(X) \leq \mu - x\right\} \leq e^{-x^2/(2mq + 2x/3)}.$$

Our probabilistic analysis of the search process begins with the following observation. For each $i \geq 0$, the set $U_i \setminus U_{i+1} = D_{i+1} \setminus D_i$ of vertices discovered at step i has law $\mathrm{Bin}(U_i, p)$. This observation also allows us to couple the search process with a family $\mathrm{B} = (B_{i,j}, i \geq 1, j \geq 1)$ of iid Bernoulli(p) random variables, by inductively letting $U_i \setminus U_{i+1} = D_{i+1} \setminus D_i$ equal $\{j \in U_i : B_{i,j} = 1\}$, for each $i \geq 1$. The coupling shows that for all $i \geq 1$, $|U_i|$ satisfies the hypotheses of Theorem 4.6, with $m = ni$ and $q = p$. Also, using the preceding coupling, the next exercise is an easy calculation.

Exercise 22. ⊛ Show that for $i \geq 0$, $\mathbf{E}\left[|U_{i+1}| \mid (U_j, j \leq i)\right] = |U_i|(1-p)$; conclude that $\mathbf{E}|U_i| = (n-1)(1-p)^i$ for all $i \geq 0$.

The exploration of the component $C(1)$ is completed precisely at the first time j that $D_j = \emptyset$; this is also the first time j that $|U_j| = n - j$, and for earlier times k we have $|U_k| < n - k$. If we had $|U_i| = \mathbf{E}|U_i|$ for all i then the above exercise would imply that $|C(1)| = \min\{t \in \mathbb{N} : (n-1)(1-p)^t \geq n - t\}$. Of course, $|U_i|$ does not equal $\mathbf{E}|U_i|$ for all i. However, $|U_i|$ does track its expectation closely enough that a consideration of the expectation yields an accurate prediction of the first-order behaviour of $\chi(G(n,p))$. We next explain this in more detail, then proceed to the proof of Theorem 4.4. Write $t(n,p)$ for the largest real solution of $n(1-p)^x = n - x$. We will use the next exercise, the first part of which which gives an idea of how $t(n,p)$ behaves when p is moderately small.

Exercise 23. ⊛
(a) Show that $t(n,p) = n \cdot \alpha(n \log(1/(1-p)))$. Conclude that if $p \leq n^{-3/4}$ then with $\mu = np$, we have

$$\alpha(\mu)n \leq t(n,p) \leq \alpha(\mu)n + \frac{2n^{1/2}}{1-p}.$$

(**Hint.** Use Exercise 21 (d).)
(b) Show that

$$n(1-p)^s \geq (n-s) + (s-t)(1 + (n-t)\log(1-p)) \text{ for } s > t$$

Write L and S for the sizes of the largest and second largest components of $G(n,p)$, respectively. From time 0 to time $t = t(n,p)$, the search process essentially explores a single component. We thus expect that $L \geq t$. Next, since $n(1-p)^{t+1} > n - (t+1)$ and $n(1-p)^t = n - t$, by the convexity of $(1-p)^s$ we have $n(1-p)^{s+1} \geq n(1-p)^s - 1$ for all $s \geq t$. Exercise 22 then implies that $\mathbf{E}|U_{s+1}| \geq \mathbf{E}|U_s| - 1$ for all integer $s \geq t$. In other words, when exploring a component after time t, the search process on average discovers less than one new vertex in each step. Such an exploration should quickly die out and, indeed, after time t the components uncovered by the search process typically all have size $o(n)$. Together with the first point, this suggests that $L \leq t + o(n)$ and $S = o(n)$. Using the bounds on t from Exercise 23 (a) and the bounds on $\chi(G)$ from Exercise 19, we are led to predict that

$$\alpha(np)^2 n + o(n) = \frac{L^2}{n} \leq \chi(G(n,p)) \leq \frac{L^2}{n} + S = \alpha(np)^2 n + o(n).$$

Theorem 4.4 formalizes and sharpens this prediction, and we now proceed to its proof.

Proof of Theorem 4.4. Throughout the proof we assume n is large (which is required for some of the inequalities), and write $t = t(n,p)$, $\alpha = \alpha(np)$.

Case 1: $p \leq 1/n + 6/n^{6/5}$ **("subcritical p").**

Recall that exploration of $C(1)$ concludes the first time i that $|U_i| \geq n - i$. Letting $t^+ = 21n^{4/5}$, we have $(1 - p)^{t^+} \geq 1 - t^+ p + (t^+ p)^2/2 - (t^+ p)^3/6 > 1 - t^+ p + (t^+ p)^2/3$, and it follows straightforwardly that

$$\mathbf{E}|U_{t^+}| = n(1 - p)^{t^+} \geq n(1 - t^+ p + (t^+ p)^2/3) \geq n - t^+ + 3n^{3/5}.$$

Applying the bound for the lower tail from Theorem 4.6 to $|U_t|$, it follows that

$$\mathbf{P}\left\{|U_{t^+}| \leq n - t^+\right\} \leq \mathbf{P}\left\{|U_{t^+}| \leq \mathbf{E}|U_{t^+}| - 3n^{3/5}\right\} \leq e^{-(9/2)n^{1/5}}.$$

At all times i before exploration of the first component concludes we have $|U_i| < n - i$, so the preceding bound yields

$$\mathbf{P}\left\{|C(1)| \geq 21n^{4/5}\right\} \leq e^{-(9/2)n^{1/5}},$$

We always have $\chi(G(n,p)) \leq \max_{i \in [n]} |C(i)|$ so, by a union bound,

$$\mathbf{P}\left\{\chi(G(n,p)) \geq 21n^{4/5}\right\} \leq \mathbf{P}\left\{\max_{i \in [n]} |C(i)| \geq 21n^{4/5}\right\} \leq ne^{-(9/2)n^{1/5}}.$$

For this range of p, by Exercise 21 (d) we also have $n\alpha(np)^2 \leq 17n^{3/5}$, and so the bound in Theorem 4.4 follows.

Case 2: $1/n + 6/n^{6/5} < p \leq 1/n^{19/20}$ **("supercritical p").**

We begin by explaining the steps of the proof. (I) First, logic similar to that in case 1 shows that the largest component of $G(n,p)$ is unlikely to have size much larger than t. (II) Next, we need a corresponding *lower* tail bound on the size of the largest component; the proof of this relies on Fact 4.5. (III) Finally, we need to know that with high probability there is only one component of large size; after ruling out one or two potential pathologies, this follows from the subcritical case. We treat the three steps in this order. Write $\Delta = n^{3/4}$ and $t^{\pm} = t(n,p) \pm \Delta$.

(I) We claim that

$$n(1 - p)^{t^+} \geq n - t^+ + 5n^{11/20}. \tag{4.6}$$

To see this, first use Exercise 23 (b) to obtain

$$n(1 - p)^{t^+} \geq n - t^+ + \Delta(1 + (n - t)\log(1 - p)).$$

Let $c = n\log(1/(1 - p))$. By Exercise 23 (a),

$$1 + (n - t)\log(1 - p) = 1 - c(1 - \alpha(c)).$$

Next, as $p \geq 1/n + 6/n^{6/5}$ we have $n\log(1/(1 - p)) \geq np \geq 1 + 6/n^{1/5} =: c^*$. By Exercise 21 (c) and (d), it follows that

$$c(1 - \alpha(c)) \leq c^*(1 - \alpha(c^*)) = \left(1 + \frac{6}{n^{1/5}}\right)\left(1 - \frac{(2 + o(1))6}{n^{1/5}}\right) \leq 1 - \frac{5}{n^{1/5}},$$

so $1 + (n-t)\log(1-p) \geq 5/n^{1/5}$. (Similar bounds using Exercise 21 (c) and (d) crop up again later in the proof). Since $\Delta/n^{1/5} = n^{11/20}$, (4.6) follows. Having established (4.6), essentially the same logic as in Case 1 yields

$$\mathbf{P}\left\{\max_{i \in [n]} |C(i)| \geq t^+\right\} \leq n\mathbf{P}\left\{|C(1)| \geq t^+\right\} \leq ne^{-(25/2)n^{1/10}}. \qquad (4.7)$$

(II) We now turn to the lower tail of $\max_{i \in [n]} |C(i)|$. The calculations are similar but slightly more involved. Since $p = o(1)$ and $p\Delta \leq n^{-1/5} = o(1)$, for n large $(1-p)^{-\Delta} \leq 1 + p\Delta + (p\Delta)^2$, so

$$n(1-p)^{t^-} = (n-t)(1-p)^{-\Delta} \leq (n-t)(1 + p\Delta + (p\Delta)^2). \qquad (4.8)$$

Since $t \geq n\alpha(n,p)$, it follows easily from Exercises 21 (c) and (d) that $p(n-t) \leq 1 - \frac{5}{n^{1/5}}$. Using (4.8) and the bound $p\Delta \leq n^{-1/5}$, we thus have

$$n(1-p)^{t^-} \leq (n-t) + \Delta\left(1 - \frac{5}{n^{1/5}}\right)(1 + p\Delta)$$

$$\leq n - t^- - \frac{4\Delta}{n^{1/5}}$$

$$= n - t^- - 4n^{3/5}. \qquad (4.9)$$

Next, basic arithmetic shows that if $m \geq (n + p^{-1})/2$ then $m(1-p) \leq m - 1 - (np-1)/2 \leq m - (1 + 6n^{-1/5})$. Furthermore, for p in the range under consideration, $(n + p^{-1})/2 \leq n - 2n^{4/5}$, so

$$n(1-p)^{\Delta} \leq \max(n - 2n^{4/5}, n - \Delta - 6\Delta n^{-1/5}) = n - \Delta - 6n^{11/20}.$$

Since $n(1-p)^t$ is concave as a function of t, this bound and (4.9) together imply that $n(1-p)^x \leq n - x - 6n^{11/20}$ for all $x \in [\Delta, t^-]$. Applying Theorem 4.6 for $t \in [\Delta, t^-]$, and a union bound, yields

$$\mathbf{P}\left\{\exists t \in [\Delta, t^-] : |U_t| \geq n - t - n^{11/20}\right\} \leq (t^- - \Delta)e^{-(25/2)n^{1/10}}.$$

Now suppose that $|U_t| < n - t - n^{11/20}$ for all $t \in [\Delta, t^-]$. In this case, if a component exploration concludes at some time $t \in [\Delta, t^-]$ then by Fact 4.5 there is $i < t$ such that $D_i = \emptyset$ and $|D_{i+1}| > n - t - |U_t| > n^{11/20}$. On the other hand, for all $i \geq 0$, $|D_{i+1} \setminus D_i|$ is stochastically dominated by $\text{Bin}(n,p)$, so by a union bound followed by a Chernoff bound (or an application of Theorem 4.6),

$$\mathbf{P}\left\{\exists i < t^- : D_i = \emptyset, |D_{i+1}| > n^{11/20}\right\} \leq t^-\mathbf{P}\left\{\text{Bin}(n,p) > n^{11/20}\right\} \leq t^- e^{-n^{11/20}}.$$

It follows that

$$\mathbf{P}\{\text{A component exploration concludes between times } \Delta \text{ and } t^-\}$$

$$\leq (t^- - \Delta)e^{-(25/2)n^{1/10}} + t^- e^{-n^{11/20}}$$

$$\leq 2ne^{-(25/2)n^{1/10}}. \qquad (4.10)$$

(III) Let N be the number of vertices which remain undiscovered the first time after time t^- that the search process finishes exploring a component, and write B for the event that some component whose exploration starts *after* time t^- has size greater than $21n^{4/5}$. Then

$$\mathbf{P}\{B\} = \mathbf{P}\{B, N > n - (t^- - \Delta)\} + \sum_{m \le n - (t^- - \Delta)} \mathbf{P}\{B, N = m\}$$

$$\le \mathbf{P}\{N > n - (t^- - \Delta)\} + \sup_{m \le n - (t^- - \Delta)} \mathbf{P}\{B \mid N = m\}.$$

The first probability is at most $2ne^{-(25/2)n^{1/10}}$ by (4.10). To bound the second, note that

$$n - (t^- - \Delta) \le n + 2n^{3/4} - t \le n + 3n^{3/4} - n\alpha(np) = n(1 - \alpha(np) + 3n^{-1/4}).$$

By Exercise 21 (c) and (d), and since $p \le n^{-19/20}$, for $m \le n - (t^- - \Delta)$ we therefore have

$$mp \le np(1 - \alpha(np)) + 3n^{3/4}p < 1.$$

For such m, the bound for "subcritical p" from Case 1 thus yields

$$\mathbf{P}\{B \mid N = m\} \le me^{-(9/2)m^{1/5}}.$$

This is less than $ne^{-(9/2)n^{4/25}}$ for $m \ge n^{4/5}$. If $m \le n^{4/5}$ then the largest component explored after time m also has size $\le n^{4/5}$, so $\mathbf{P}\{B \mid N = m\} = 0$. We conclude that

$$\mathbf{P}\{B\} \le 2ne^{-(25/2)n^{1/10}} + ne^{-(9/2)n^{4/25}} \le 3ne^{-(25/2)n^{1/10}}. \tag{4.11}$$

(IV) Now to put the the pieces together. The lower bound is easier: by the first inequality from Exercise 19, we have $\chi(G(n,p)) \ge n^{-1}\max_{i \in [n]}|C(i)|^2$. From (4.10) it then follows that

$$\mathbf{P}\left\{\chi(G(n,p)) < \frac{(t^- - \Delta)^2}{n}\right\} \le \mathbf{P}\left\{\max_{i \in [n]}|C(i)| < \Delta - t^-\right\} \le 2ne^{-(25/2)n^{1/10}}. \tag{4.12}$$

By Exercise 23 (a),

$$\frac{(t^- - \Delta)^2}{n} = \frac{(t - 2n^{3/4})^2}{n} \ge \frac{(n\alpha - 3n^{3/4})^2}{n} \ge n\alpha^2 - 9n^{1/2},$$

and the lower bound then follows from (4.12).

For the upper bound, any component of $G(n,p)$ whose exploration concludes before step $n^{3/4}$ of the search process has size at most $n^{3/4}$. Write S for the number of vertices of the second-largest component of $G(n,p)$. By (4.11), we then have

$$\mathbf{P}\left\{S \ge 21n^{4/5}\right\} \le 3ne^{-(25/2)n^{1/10}}.$$

Combined with the second inequality from Exercise 19 and with (4.7), we obtain

$$\mathbf{P}\left\{\chi(G(n,p)) \ge \frac{(t + n^{3/4})^2}{n} + 21n^{4/5}\right\} \le 4ne^{-(25/2)n^{1/10}}. \tag{4.13}$$

An easy calculation using Exercise 23 (a) shows that $(t + n^{3/4})^2/n + 21n^{4/5} \leq n\alpha^2 + 22n^{4/5}$, and the theorem then follows from (4.12) and (4.13). $\qquad\square$

To conclude the section, we use Theorem 4.4 to show that $\mathbf{E}\left[\chi(G_m^{(n)})\right]$ is well-approximated by $\alpha(2m/n)$ in a range which covers the most important values of m. (Exercise 24, below, extends this to all $0 \leq m \leq \binom{n}{2}$.)

Lemma 4.7. *For n large, for all $m \leq n^{22/21}$,*

$$\left|\mathbf{E}\left[\chi(G_m^{(n)})\right] - \alpha^2(2m/n)n\right| \leq 23n^{4/5}.$$

Proof. Write $x_m = \inf\{p : |E(G(n,p))| = m\}$. Since $\mathbf{E}|E(G(n,p))| = p\binom{n}{2}$, we expect x_m to be near $p_m := m/\binom{n}{2}$. Write $\hat{\alpha} = \alpha(2m/(n-1)) = \alpha(np_m)$, let $\delta = n^{-4/3}$, and let $p_m^{\pm} = p_m \pm \delta$.

In the coupling of $(G(n,p), 0 \leq p \leq 1)$ and $(G_m^{(n)}, 0 \leq m \leq \binom{n}{2})$, if $x_m > p_m^-$ then $G(n, p_m^-)$ is a subgraph of $G_m^{(n)}$ and so $\chi(G_m^{(n)}) \geq \chi(G(n, p_m^-))$. Likewise, if $x_m < p_m^+$ then $\chi(G_m^{(n)}) \leq \chi(G(n, p_m^+))$ We thus have

$$\chi(G_m^{(n)}) \geq \chi(G(n, p_m^-))\mathbf{1}_{[x_m > p_m^-]}$$
$$\geq \chi(G(n, p_m^-)) - n\mathbf{1}_{[x_m \leq p_m^-]}, \qquad \text{and}$$
$$\chi(G_m^{(n)}) \leq \chi(G(n, p_m^+))\mathbf{1}_{[x_m < p_m^+]} + n\mathbf{1}_{[x_m \geq p_m^+]}$$
$$\leq \chi(G(n, p_m^+)) + n\mathbf{1}_{[x_m \geq p_m^+]}.$$

Since α is 2-Lipschitz, $\hat{\alpha} - 2/n^{1/3} \leq \alpha(np_m^-) \leq \alpha(np_m^+) \leq \hat{\alpha} + 2/n^{1/3}$, from which it follows that both $\alpha(np_m^-)^2n$ and $\alpha(np_m^+)^2n$ are within $5n^{2/3}$ of $\hat{\alpha}^2n$. By the preceding lower bound on $\chi(G_m^{(n)})$ and Theorem 4.4 we thus have

$$\mathbf{E}\left[\chi(G_m^{(n)})\right] \geq \mathbf{E}\left[\chi(G(n, p_m^-))\right] - n\mathbf{P}\left\{x_m \leq p_m^-\right\}$$
$$\geq \hat{\alpha}^2n - 5n^{2/3} - 22n^{4/5} - n\mathbf{P}\left\{\mathrm{Bin}\left(\binom{n}{2}, p_m^-\right) \geq m\right\}$$
$$\geq \hat{\alpha}^2n - 5n^{2/3} - 22n^{4/5} - 1$$

the last inequality holding straightforwardly by a Chernoff bound (note that $\binom{n}{2}p_m^- = m - (n-1)/(2n^{1/3}) \leq m - m^{3/5}/3$). We likewise have

$$\mathbf{E}\left[\chi(G_m^{(n)})\right] \leq \hat{\alpha}^2n + 5n^{2/3} + 22n^{4/5} + 1.$$

Finally, $2m/(n-1) - 2m/n = 2m/(n(n-1)) = O(n^{-8/9})$, so since α is 2-Lipschitz we have $\alpha(2m/n)^2 = \hat{\alpha}^2 + O(n^{-8/9})$, and the result follows. $\qquad\square$

5. Frieze's $\zeta(3)$ limit for the MST weight

Before proving Theorem 4.2, we warm up by using the same approach to study the total weight of random MSTs. Throughout the section, $\mathbf{W} = (W_e, e \in E(K_n))$ are exchangeable, distinct, non-negative edge weights. Recall from Section 2.3.2 that "Version 2" of the multiplicative coalescent (aka Kruskal's algorithm) considers edges one-by-one in increasing order of weight, adding only edges which connect distinct trees in the forest, and that the result is the minimum spanning tree $T = \mathrm{MST}(K_n, \mathbf{W})$.

Write $w(T) = \sum_{e \in E(T)} W_e$ for the total weight of T. We use susceptibility bounds to approximate $w(T)$ and derive a version of Frieze's famous $\zeta(3)$ limit.

Theorem 5.1 (Frieze [9]). *Write $X_1, \ldots, X_{\binom{n}{2}}$ for the increasing ordering of $\mathbf{W}$. If $\mathbf{E}X_m = (1 + o(1))m$, then $2\mathbf{E}w(MST(K_n, \mathbf{W}))/n^2 \to \zeta(3)$ as $n \to \infty$.*

By $\mathbf{E}X_m = (1 + o(1))m$ we mean that $\lim_{n \to \infty} \sup_{m \in [\binom{n}{2}]} |1 - \mathbf{E}X_m/m| = 0$. This condition can be relaxed, and the proof can be modified to obtain convergence in probability under suitable hypotheses, but for expository reasons we have opted for simplicity over full generality. Before beginning the proof, we first note a special case. Suppose that the weights W_e are independent Uniform$[0, 1]$ random variables. Then $\mathbf{E}[X_k] = k/(\binom{n}{2} + 1)$, $\mathbf{E}[X_k \cdot n^2/2] = (1 + o(1))k$. The theorem thus implies that for such uniform edge weights, the total weight of the random MST of K_n converges to $\zeta(3)$ without renormalization. This is the most often quoted special case of Frieze's result.

Our proof is based on the following identity for $\mathbf{E}[w(T)]$.

Proposition 5.2. *Write $X_1, \ldots, X_{\binom{n}{2}}$ for the increasing ordering of $\mathbf{W}$. Then*

$$\mathbf{E}[w(T)] = \sum_{m=0}^{\binom{n}{2}-1} \mathbf{E}[X_{m+1}] \cdot \mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\}. \tag{5.1}$$

Proof. Let $e_1, \ldots, e_{\binom{n}{2}}$ be the ordering of $E(K_n)$ by increasing weight, so e_m has weight X_m. In the coupling with the Erdős-Rényi coalescent, Kruskal's algorithm adds edge e_k precisely if e_k joins distinct components of $G_{k-1}^{(n)}$, which occurs if and only if $\chi(G_k^{(n)}) > \chi(G_{k-1}^{(n)})$. For this coupling we thus have

$$w(T) = \sum_{m=0}^{\binom{n}{2}-1} X_{m+1} \cdot \mathbf{1}_{[\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})]}.$$

By the exchangeability of $\mathbf{W}$, the vector $(X_1, \ldots, X_{\binom{n}{2}})$ is independent of the ordering of $E(K_n)$. The event that $\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})$ is measurable with respect to the ordering of $E(K_n)$, so is independent of $(X_1, \ldots, X_{\binom{n}{2}})$. The proposition follows on taking expectations. $\qquad\square$

We use the result of the following exercise to deduce that terms with $m \geq 5n \log n$ play an unimportant role in the summation (5.1). Fix $1 \leq k \leq \lfloor n/2 \rfloor$ and let N_k be the number of sets $A \subset [n]$ of size k such that, in $G_m^{(n)}$, there are no edges from A to $[n] \setminus A$. Note that $G_m^{(n)}$ is connected precisely if $N_k = 0$ for all $1 \leq k \leq \lfloor n/2 \rfloor$.

Exercise 24. ⊛

(a) Let E_k be the event that there are no edges from $[k]$ to $[n] \setminus [k]$ in $G_m^{(n)}$. With $p = m/\binom{n}{2}$, show that $\mathbf{P}\{E_k\} \leq (1-p)^{k(n-k)}$. Deduce that
$$\mathbf{P}\{N_k > 0\} \leq n^k(1-p)^{k(n-k)} \leq (ne^{-p(n-k)})^k .$$

(b) Show that $\mathbf{P}\left\{G_{\lceil 5n \log n \rceil}^{(n)} \text{ is not connected}\right\} \leq n^{-4}$.

(c) Show that the bound in Lemma 4.7 in fact holds for all $m \in [\binom{n}{2}]$.

Corollary 5.3. *With the notation of Proposition 5.2, we have*

$$\mathbf{E}\left[w(T)\right] \geq \sum_{m=0}^{5n \log n} \mathbf{E}\left[X_{m+1}\right]\left(1 - \frac{\mathbf{E}\chi(G_m^{(n)})}{n}\right) \qquad and$$

$$\mathbf{E}\left[w(T)\right] \leq \left(1 + \frac{12 \log n}{n}\right)\sum_{m=0}^{5n \log n} \mathbf{E}X_{m+1}\left(1 - \frac{\mathbf{E}\chi(G_m^{(n)})}{n}\right) + \frac{1}{2n^2}\mathbf{E}\left[X_{\binom{n}{2}}\right] .$$

Proof. Write

$$\mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\} = \mathbf{E}\left[\mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)}) \mid G_m^{(n)}\right\}\right] .$$

We derived an identity for the inner conditional probability in (4.3); using that identity and linearity of expectation, we obtain

$$\mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\} = \left(1 - \frac{\mathbf{E}\chi(G_m^{(n)})}{n}\right)\left(1 - \frac{n+2m}{n^2}\right)^{-1} . \qquad (5.2)$$

The latter is always at least $1 - \mathbf{E}\chi(G_m^{(n)})/n$, and the lower bound then follows from Proposition 5.2 by truncating the sum at $m = 5n \log n$.

For the upper bound, note that if $G_m^{(n)}$ is connected then $\chi(G_m^{(n)}) = n$, so

$$\mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\} \leq \mathbf{P}\left\{G_m^{(n)} \text{ is not connected}\right\} .$$

Using Exercise 24 (b) and the fact that the X_i are increasing, it follows that

$$\mathbf{E}\left[w(T)\right] \leq \sum_{m=0}^{5n \log n} \mathbf{E}X_{m+1} \cdot \mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\}$$

$$+ \sum_{m=5n \log n + 1}^{\binom{n}{2}-1} \mathbf{E}X_{m+1} \cdot \mathbf{P}\left\{G_m^{(n)} \text{ is not connected}\right\}$$

$$\leq \sum_{m=0}^{5n \log n} \mathbf{E}X_{m+1} \cdot \mathbf{P}\left\{\chi(G_{m+1}^{(n)}) > \chi(G_m^{(n)})\right\} + \binom{n}{2} \cdot \mathbf{E}\left[X_{\binom{n}{2}}\right] \cdot \frac{1}{n^4}\,.$$

For $m \leq 5n \log n$, $(1 - \frac{n+2m}{n^2})^{-1} \leq 1 + 12 \log n / n$, and the result follows from (5.2). $\qquad\square$

To prove Theorem 5.1, we use Lemma 4.7 and Corollary 5.3 to show that after appropriate rescaling, the sum in Proposition 5.2 is essentially a Riemann sum approximating an appropriate integral. The value of that integral is derived in the following lemma.

Lemma 5.4.
$$\int_0^\infty \lambda \cdot (1 - \alpha^2(\lambda))\mathrm{d}\lambda = 2\zeta(3)\,.$$

Proof. Aldous and Steele [5] write that "calculation of this integral is quite a pleasing experience"; though the calculation appears in that work, why should we deprive ourselves of the pleasure? Anyway, the proof is short. First, use integration by parts to write

$$\int_0^\infty \lambda \cdot (1 - \alpha^2(\lambda))\mathrm{d}\lambda = \int_0^\infty \alpha(\lambda)\alpha'(\lambda)\lambda^2 \mathrm{d}\lambda = \int_1^\infty \alpha(\lambda)\alpha'(\lambda)\lambda^2 \mathrm{d}\lambda\,,$$

the second equality since $\alpha(\lambda) = 0$ for $\lambda < 1$. The identity $1 - \alpha(c) = e^{-c\alpha(c)}$ (this is how we *defined* α) implies that $\lambda^2 = (\alpha(\lambda)^{-1}\log(1-\alpha(\lambda)))^2$, so we may rewrite the latter integral as

$$\int_1^\infty \frac{\log^2(1 - \alpha(\lambda))}{\alpha(\lambda)} \cdot \alpha'(\lambda)\mathrm{d}\lambda = \int_0^1 \frac{\log^2(1 - \alpha)}{\alpha}\mathrm{d}\alpha\,,$$

where we used the obvious change of variables $\alpha = \alpha(\lambda)$. Now a final change of variables: $u = -\log(1 - \alpha)$ transforms this into

$$\int_0^\infty u^2 \frac{e^{-u}}{1 - e^{-u}}\mathrm{d}u = \int_0^\infty u^2 \sum_{k=1}^\infty e^{-ku}\mathrm{d}u\,.$$

Since $\int_0^\infty u^2 e^{-ku} = 2/k^3$, the final expression equals $\sum_{k=1}^\infty 2/k^3 = 2\zeta(3)$. $\qquad\square$

Our final step before the proof is to show that the integrand is well-behaved on the region of integration; the straightforward bound we require is stated in the following exercise. Recall that α is continuous on $[0, \infty)$ and is differentiable except at $x = 1$.

Exercise 25. ⊛ Let $f(x) = x(1 - \alpha^2(x))$, where α is as above. Show there exists $C < \infty$ such that $|f'(x)| \leq C$ for all $x \neq 1$. (In fact we can take $C = 2$.)

Proof of Theorem 5.1. By the preceding exercise, for all $0 < \epsilon \leq x$,

$$\left| \int_{x-\epsilon}^{x} \lambda(1 - \alpha^2(\lambda))\mathrm{d}\lambda - \epsilon x(1 - \alpha^2(x)) \right| \leq C\epsilon^2 \,.$$

Taking $\epsilon = 2/n$, $x = 2m/n$ and summing over $m \in \{1, \ldots, 5n\log n\}$ we obtain in particular that

$$\sum_{m=1}^{5n\log n} \frac{4m}{n^2}(1 - \alpha^2(2m/n)) = \int_0^{10\log n} \lambda(1 - \alpha^2(\lambda))\mathrm{d}\lambda + O\left(\frac{\log n}{n^2}\right)$$
$$= 2\zeta(3) - o(1) \,,$$

the second equality by Lemma 5.4. If $\mathbf{E}[X_m] = (1 + o(1))m$ then by the preceding equation, Lemma 4.7, and the lower bound in Corollary 5.3, we have

$$\mathbf{E}[w(T)] \geq (1 + o(1))\frac{n^2}{2} \cdot \zeta(3) \,,$$

and likewise (this time using the upper bound in Corollary 5.3)

$$\mathbf{E}[w(T)] \leq (1 + o(1))\frac{n^2}{2} \cdot \zeta(3) + \frac{1}{2n^2}\mathbf{E}\left[X_{\binom{n}{2}}\right]$$
$$= (1 + o(1))\frac{n^2}{2} \cdot \zeta(3) + O(1) \,,$$

which completes the proof. ☐

6. Estimating the empirical entropy

We already know the broad strokes of the argument, since they are the same as for our proof of Theorem 5.1. Recall that we are trying to approximate $\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n)\right] = \mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)\right] - (n-1)\log 2$. Proposition 4.1 reduces this to the study of the sum

$$\Xi = \sum_{m=0}^{\binom{n}{2}-1} \left(1 - \frac{n+2m}{n^2}\right)^{-1} \mathbf{E}\left[\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\right]. \quad (6.1)$$

We use Theorem 4.4 and Exercise 24 to approximate this sum by an integral. Before proceeding to the ϵ's and δ's, we evaluate the integral.

Proposition 6.1. *We have*

$$\int_0^\infty (1 - \alpha^2(\lambda)) \cdot \log(1 - \alpha^2(\lambda))\mathrm{d}\lambda = 2(\zeta_{\mathrm{MC}} + \log 2) \,. \quad (6.2)$$

Proof. A similar calculation to that of Lemma 5.4, though decidedly less pleasing. Since $\alpha(\lambda) = 0$ for $\lambda \leq 1$, we may change the domain of integration to $[1, \infty)$. Then use the identity

$$\alpha'(\lambda) = \frac{\alpha(\lambda)^2(1 - \alpha(\lambda))}{\alpha(\lambda) + (1 - \alpha(\lambda))\log(1 - \alpha(\lambda))},$$

which follows from the fact that $1 - \alpha(\lambda) = e^{-\lambda\alpha(\lambda)}$ by differentiation. The integral under consideration thus equals

$$\int_1^\infty (1 - \alpha^2(\lambda))\log(1 - \alpha^2(\lambda)) \cdot \frac{\alpha(\lambda) + (1 - \alpha(\lambda))\log(1 - \alpha(\lambda))}{\alpha(\lambda)^2(1 - \alpha(\lambda))} \cdot \alpha'(\lambda)\mathrm{d}\lambda,$$

from which the substitution $\alpha = \alpha(\lambda)$ gives

$$\int_0^1 \frac{(1 + \alpha)\log(1 - \alpha^2)(\alpha + (1 - \alpha)\log(1 - \alpha))}{\alpha^2}\mathrm{d}\alpha.$$

Substituting $u = -\log(1 - \alpha)$, we have $1 - \alpha = e^{-u}$, $1 + \alpha = 2 - e^{-u}$ and $\log(1 - \alpha^2) = \log(2 - e^{-u}) - u$, and the above integral becomes

$$\int_0^1 \frac{(2 - e^{-u})(\log(2 - e^{-u}) - u) \cdot (1 - e^{-u} - ue^{-u})}{(1 - e^{-u})^2} \cdot e^{-u}\mathrm{d}u.$$

This integral can be calculated with a little effort (or easily, for those who accept computer assisted proofs), and equals

$$\frac{\pi^2}{3} - 6 + 4\log 2 - 2\log^2(2)$$

Comparing with (4.5) completes the proof (recall that $\zeta(2) = \pi^2/6$). □

The next lemma generalizes Lemma 4.7, at the cost of obtaining a non-explicit error bound. We use a slightly different proof technique than for Lemma 4.7; this time we exploit that a binomial random variable is reasonably likely to take values close to its mean (see the following exercise).

Exercise 26. ❋ Show that

$$\mathbf{P}\left\{\mathrm{Bin}\left(\binom{n}{2}, \frac{2m}{n^2}\right) = m\right\} = \Omega\left(\frac{1}{n}\right)$$

uniformly in $0 \leq m \leq n^2/4$, in that

$$\liminf_{n \to \infty} \inf_{m \in \{0, 1, \ldots, \lfloor n^2/4 \rfloor\}} n \cdot \mathbf{P}\left\{\mathrm{Bin}\left(\binom{n}{2}, \frac{2m}{n^2}\right) = m\right\} > 0.$$

For a continuous function $f : [0, 1] \to \mathbb{R}$, and $\epsilon \in [0, 1]$, write $\gamma_f(\epsilon) = \sup_{|x-y| \leq \epsilon} |f(x) - f(y)|$, and $\|f\| = \sup_{x \in [0,1]} |f(x)|$.

Lemma 6.2. *Let $f : [0, 1] \to \mathbb{R}$ be continuous. Then*

$$\sup_{m \in [\binom{n}{2}]} \left|\mathbf{E}\left[f(\chi(G_m^{(n)})/n)\right] - f(\alpha(2m/n)^2)\right| = O(\gamma_f(n^{-1/5}))) + O(n^{-4}).$$

Proof. First suppose $m \geq n^2/4$. Then $\alpha(2m/n) \geq \alpha(n/2) \geq 1 - 2e^{-n/2}$ by Exercise 21 (b), so we have $|f(\alpha^2(2m/n)) - f(1)| \leq \gamma_f(2e^{-n/2}) \leq \gamma_f(n^{-1/5})$, the latter for n large. Next, since $\chi(G) = |G|$ whenever G is connected, by Exercise 24 (b) we have

$$\left|\mathbf{E}\left[f(\chi(G_m^{(n)})/n)\right] - f(1)\right| \leq \|f\|\mathbf{P}\left\{\chi(G_m^{(n)}) \neq n\right\} = O(n^{-4}).$$

This handles the case $m \geq n^2/4$, so we now assume $0 \leq m \leq n^2/4$.

Let $p = 2m/n^2$, so $np = 2m/n$. By Exercise 20, we have

$$\mathbf{E}\left[f(\chi(G_m^{(n)})/n)\right] = \mathbf{E}\left[f(\chi(G(n,p))/n) \mid |E(G(n,p))| = m\right].$$

By Exercise 26, there is $C > 0$ such that for all $m \leq n^2/4$,

$$\mathbf{P}\left\{|\chi(G(n,p)) - n\alpha(2m/n)^2| > 22n^{4/5} \mid |E(G(n,p))| = m\right\}$$
$$\leq Cn\mathbf{P}\left\{|\chi(G(n,p)) - n\alpha(2m/n)^2| > 22n^{4/5}\right\}$$
$$= O(n^2 e^{-12n^{1/10}}),$$

the last line by Theorem 4.4. It follows that

$$\mathbf{E}\left[f(\chi(G_m^{(n)})/n)\right]$$
$$\geq \inf_{|a - n\alpha(2m/n)^2| \leq 22n^{4/5}} f(a/n) \cdot \mathbf{P}\left\{|\chi(G(n,p)) - n\alpha(2m/n)^2| \leq 22n^{4/5} \mid |E(G(n,p))| = m\right\}$$
$$\quad - \|f\| \cdot \mathbf{P}\left\{|\chi(G(n,p)) - n\alpha(2m/n)^2| > 22n^{4/5} \mid |E(G(n,p))| = m\right\}$$
$$= \inf_{|a - n\alpha(2m/n)^2| \leq 22n^{4/5}} f(a/n) - O(n^2 e^{-12n^{1/10}})$$
$$\geq f(\alpha(2m/n)^2) - \beta(f, 22n^{-1/5}) - O(n^2 e^{-12n^{1/10}}).$$

We likewise have $\mathbf{E}\left[f(\chi(G_m^{(n)})/n)\right] \leq f(\alpha(2m/n)^2) + \beta(f, 22n^{-1/5}) + O(n^2 e^{-12n^{1/10}})$. Since $\gamma_f(22n^{-1/5}) \leq 22\gamma_f(n^{-1/5})$ and $n^2 e^{-12n^{1/10}} = o(n^{-4})$, the result follows. $\square$

In what follows we only apply the preceding lemma with $m = o(\binom{n}{2})$, but it seems more satisfying to prove the estimate over the full range of possibilities; handling larger m only added a few lines to the proof. We are now ready to wrap things up. Define a function $f : [0,1] \to \mathbb{R}$ by $f(x) = (1-x)\log(1-x)$ for $x \in [0,1)$; recalling the convention that $0\log 0 = 0$, we see that f is a continuous function. The next exercise, the last of the notes, asks the reader to establish two basic properties of f.

Exercise 27.   (a) Show that $\|f\| = 1/e$.
   (b) Show that for all $x, y \in [0,1]$, $|f(x) - f(y)| \leq |x - y|\log(1/|x - y|)$.

Proof of Theorem 4.2. By Exercises 24 (b) and 27 (a), for $m \geq 5n \log n$ we have

$$\mathbf{E}\left[\left(1 - \frac{\chi(G_m^{(n)})}{n}\right) \log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\right] \leq \|f\| \cdot \mathbf{P}\left\{G_m^{(n)} \text{ is not connected}\right\} \leq \frac{1}{en^4}.$$

Summing in (6.1) over indices $m \geq 5n \log n$ and using that $1 \leq (1 - (n + 2m)/n^2)^{-1} = 1 + O(\log n/n)$ for $m \leq 5n \log n$, this implies that

$$\left| \Xi - \sum_{m=0}^{5n \log n} \mathbf{E}\left[\left(1 - \frac{\chi(G_m^{(n)})}{n}\right) \log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\right]\right| = O\left(\log^2 n\right).$$

By Lemma 6.2 we have

$$\sum_{m=0}^{5n \log n} \mathbf{E}\left[\left(1 - \frac{\chi(G_m^{(n)})}{n}\right) \log\left(1 - \frac{\chi(G_m^{(n)})}{n}\right)\right]$$

$$= \sum_{m=0}^{5n \log n} \left(1 - \alpha^2(2m/n)\right) \log\left(1 - \alpha^2(2m/n)\right) + O\left(n \log n \cdot \left(\gamma_f(n^{-1/5}) + n^{-4}\right)\right).$$

Exercise 27 (b) implies that $\gamma_f(n^{-1/5}) \leq (\log n)/(5n^{1/5})$, so the preceding two bounds yield

$$\left| \Xi - \sum_{m=0}^{5n \log n} \left(1 - \alpha^2(2m/n)\right) \log\left(1 - \alpha^2(2m/n)\right)\right| = O(n^{4/5} \log^2 n). \qquad (6.3)$$

Next, by Exercise 21 (d) we know that $\alpha'(x) < 2$ for $x > 1$. Also, $\alpha(x) = 0$ for $x \leq 1$. It follows that for $x, y \geq 0$ and $\epsilon \in (0,1)$, if $|x - y| < \epsilon$ then $\alpha^2(x) - \alpha^2(y) \leq 4\epsilon(\alpha(x) + \epsilon) < 8\epsilon$. By Exercise 27 (b), this implies that for $x \in [0, \infty)$ and $\epsilon \in (0,1)$,

$$\left| \epsilon \cdot f(\alpha^2(x)) - \int_{x-\epsilon}^{x} f(\alpha^2(\lambda))\mathrm{d}\lambda\right| \leq 8\epsilon^2 \log(1/8\epsilon).$$

We then have, as in Theorem 5.1, that

$$\sum_{m=1}^{5n \log n} \left(1 - \alpha^2(2m/n)\right) \log\left(1 - \alpha^2(2m/n)\right)$$

$$= \frac{n}{2} \int_0^{10 \log n} (1 - \alpha^2(\lambda)) \log(1 - \alpha^2(\lambda))\mathrm{d}\lambda + O\left(\frac{\log^2 n}{n}\right)$$

$$= n(\zeta_{\mathrm{MC}} + \log 2) + o(n),$$

The second equality holds since $\int_0^\infty (1 - \alpha^2(\lambda)) \log(1 - \alpha^2(\lambda))\mathrm{d}\lambda = 2(\zeta_{\mathrm{MC}} + \log 2)$ and the integrand is negative, so $\lim_{x \to \infty} \int_x^\infty (1 - \alpha^2(\lambda)) \log(1 - \alpha^2(\lambda))\mathrm{d}\lambda = 0$.

Combining this with (6.3) and Proposition 4.1, we obtain

$$\mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}(n)\right] = \mathbf{E}\left[\log \hat{Z}_{\mathrm{MC}}^{\rightarrow}(n)\right] - (n-1)\log 2 = n \cdot (2 \log n + \zeta_{\mathrm{MC}}^{\rightarrow} + o(1)),$$

which is the assertion of the theorem. $\qquad \square$

7. Unanswered questions

The partition function of the multiplicative coalescent provides an interesting avenue by which to approach the probabilistic study of the process. It connects up nicely with other perspectives, and offers its own insights and challenges. We saw above that the empirical partition function of the multiplicative coalescent is a subtle and interesting random variable. Here are a few questions related to $\hat{Z}_{\mathrm{MC}}(n)$, and more generally to the multiplicative coalescent, that occurred to me in the course of writing these notes and which I believe deserve investigation.

- The large deviations of $\log \hat{Z}_{\mathrm{MC}}(n)$ should be interestingly non-trivial. Can a large deviations rate function be derived? This should be related to large deviations for component sizes in the random graph process. Such results exist for fixed p [7, 12], but not (so far as I am aware) for the process as p varies. (Considering the following sort of problem would be a step in the right direction. Let E_n be the event that the largest component of $G(n,p)$ has at least $0.1n$ fewer vertices than average, for all $p \in [2/n, 3/n]$. Find a law of large numbers for $\log \mathbf{P}\{E_n\}$.)

- Relatedly, what partition chains are responsible for the large value of $\mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n)\right]$?

  It is not too hard to show the following: to maximize $\prod_{i=1}^{n-1} n^2(1 - \chi(F_i)^2/n)$ one should keep the component sizes as small as possible. In particular, if $n = 2^p$ then one maximizes this product by first pairing all singletons to form trees of size two, then pairing these trees to form trees of size 4, etcetera. This shows that for $n = 2^p$,

$$\operatorname{ess\,sup} \hat{Z}_{\mathrm{MC}}(n) = 2^{-(n-1)} \prod_{k=1}^{p=1} \prod_{j=0}^{n/2^k - 1} \left(n^2 - 2^{k-1}(n + j \cdot 2^k)\right) .$$

  which is within a factor 4 of $2^{-(n-1)}n^{2(n-1)}e^{-\log_2 n}$. On the other hand, a straightforward calculation shows the probability of choosing two minimal trees to pair at every step is around $e^{-(1+o(1))2n}$, so the contribution to $\mathbf{E}\hat{Z}_{\mathrm{MC}}(n)$ from such paths is $n^{2n}e^{-(1+o(1))(2+\log 2)n}$. This is exponentially small compared to $n^{n-2}(n-1)!$, so the lion's share of the expected value comes from elsewhere.

- Suppose we condition $\hat{Z}_{\mathrm{MC}}(n)$ to be close to $n^{n-2}(n-1)! = \mathbf{E}\left[\hat{Z}_{\mathrm{MC}}(n)\right]$; we know by Corollary 4.3 that this event has exponentially small probability. Perhaps, under this conditioning, the tree $T_1^{(n)}$ built by the multiplicative coalescent might be similar to that built by the additive coalescent? At any rate, it would certainly be interesting to study, e.g., $\mathbf{E}\left[\operatorname{height}(T_1^{(n)}) \mid \hat{Z}_{\mathrm{MC}}(n) \geq n^{n-2}(n-1)!\right]$, or more generally to study observables of $T_1^{(n)}$ under unlikely conditionings of $\hat{Z}_{\mathrm{MC}}(n)$.

- Condition $T_1^{(n)}$ to have exactly k leaves. After rescaling distances appropriately, $T_1^{(n)}$ should converge in the Gromov-Hausdorff sense. What is the limit?

Write $\mathbf{E}_k$ for the coresponding conditional expectation; then we should have, for example, $\mathbf{E}_k[\mathrm{diam}(T_1^{(n)})/n] \to f(k)$ for some function $f(k)$. How does f behave as $k \to \infty$? It is known [1] that without conditioning, $\mathbf{E}\left[\mathrm{diam}(T_1^{(n)})\right] = \Theta(n^{1/3})$.

- Pitman's coalescent, Kingman's coalescent, and the multiplicative coalescent correspond to rate kernels $\kappa(x,y) = x + y$, $\kappa(x,y) = 1$, and $\kappa(x,y) = xy$, respectively. Baur [6] has shown that for any integer $k \geq 3$, the coalescent with kernel $\kappa(x_1,...,x_k) = x_1 + ... + x_k + k/(k-2)$ admits a representation in terms of $(k-1)$-ary forests. Are there further rate kernels that may be naturally enriched to form interesting forest-valued coalescent processes?

Acknowledgment

References

[1] L. Addario-Berry, N. Broutin, and B. Reed. Critical random graphs and the structure of a minimum spanning tree. *Random Structures Algorithms*, 35(3): 323–347, 2009. URL http://www.cs.mcgill.ca/~nbrout/pub/knmst_rsa.pdf. 22, 41

[2] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont. The scaling limit of the minimum spanning tree of the complete graph. arXiv:1301.1664 [math.PR], 2013. 12

[3] D.J. Aldous. A random tree model associated with random graphs. *Random Structures Algorithms*, 1(4):383–402, 1990. URL http://stat-www.berkeley.edu/~aldous/Papers/me49.pdf. 12

[4] D.J. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, 1991. URL http://www.stat.berkeley.edu/~aldous/Papers/me55.pdf. 21

[5] D.J. Aldous and J.M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, 110:1–72, 2004. URL http://www.stat.berkeley.edu/~aldous/Papers/me101.pdf. 35

[6] Erich Baur. On a ternary coalescent process. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(2):561–589, 2013. URL http://arxiv.org/pdf/1301.0409.pdf. 41

[7] M. Biskup, L. Chayes, and S. A. Smith. Large-deviations/thermodynamic approach to percolation on the complete graph. *Random Structures Algorithms*, 31(3):354–370, 2007. URL http://arxiv.org/abs/math/0506255. 40

[8] L. Devroye. A note on the height of binary search trees. *J. Assoc. Comput. Mach.*, 33(3):489–498, 1986. URL http://luc.devroye.org/devroye_1986_univ_a_note_on_the_height_of_binary_search_trees.pdf. 21

[9] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10:47–56, 1985. URL http://www.math.cmu.edu/~af1p/Texfiles/MST.pdf. 33

[10] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311 (electronic), 2005. URL http://www.i-journals.org/ps/viewarticle.php?id=53. 21

[11] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989. URL http://www.stats.ox.ac.uk/people/academic_staff/colin_mcdiarmid/?a=4113. 27

[12] N. O'Connell. Some large deviation results for sparse random graphs. *Probab. Theory Related Fields*, 110(3):277–285, 1998. URL http://homepages.warwick.ac.uk/~masgas/pubs/noc98b.pdf. 40

[13] J. Pitman. Coalescent random forests. *J. Combin. Theory Ser. A*, 85:165–193, 1999. URL http://www.stat.berkeley.edu/~pitman/457.pdf. 1, 2, 3

[14] A. Rényi. Some remarks on the theory of trees. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 4:73–85, 1959. 12

[15] A. Rényi and G. Szekeres. On the height of trees. *J. Austral. Math. Soc.*, 7:497–507, 1967. URL http://dx.doi.org/10.1017/S1446788700004432. 21

Louigi Addario-Berry

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, Québec, H3A 2K6, Canada

e-mail: louigi@math.mcgill.ca

URL: http://www.math.mcgill.ca/~louigi/