Brownian Excursions, Critical Random Graphs and the Multiplicative Coalescent

David Aldous*
Department of Statistics
University of California
Berkeley CA 94720
aldous@stat.berkeley.edu

July 31, 1996

Abstract

Let $(B^t(s), 0 \le s < \infty)$ be reflecting inhomogeneous Brownian motion with drift $t - s$ at time s, started with $B^t(0) = 0$. Consider the random graph $\mathcal{G}(n, n^{-1} + tn^{-4/3})$, whose largest components have size of order $n^{2/3}$. Normalizing by $n^{-2/3}$, the asymptotic joint distribution of component sizes is the same as the joint distribution of excursion lengths of B^t (Corollary 2). The dynamics of merging of components as t increases are abstracted to define the *multiplicative coalescent process*. The states of this process are vectors $\mathbf{x}$ of nonnegative real cluster sizes (x_i), and clusters with sizes x_i and x_j merge at rate $x_i x_j$. The multiplicative coalescent is shown to be a Feller process on l_2. The random graph limit specifies the *standard* multiplicative coalescent, which starts from infinitesimally small clusters at time $-\infty$: the existence of such a process is not obvious.

AMS 1991 subject classifications. 60C05, 60J50,

Key words and phrases. Brownian motion, Brownian excursion, Markov process, random graph, critical point, stochastic coalescent, stochastic coagulation, weak convergence.

Running head: Random graph – multiplicative coalescent.

*Research supported by N.S.F. Grant DMS92-24857

1 Introduction

1.1 A stochastic process

We start by describing more carefully the stochastic process mentioned in the abstract. For readers with a background in random graphs, we should emphasize that this process is a comparatively simple instance of the kind of process that the modern martingale-based theory of stochastic calculus, treated in e.g. Rogers and Williams [27, 26] or Revuz and Yor [25], is designed to study. Conversely, for readers with a background in modern process theory, we should emphasize that our results open up challenging problems to even re-derive, via process techniques, existing random graph asymptotics formulas.

Fix $-\infty < t < \infty$. Let $(W(s), 0 \le s < \infty)$ be standard Brownian motion. Then

$$W^t(s) = W(s) + ts - \tfrac{1}{2}s^2, \quad s \ge 0 \tag{1}$$

defines the (inhomogeneous) Brownian motion with drift $t - s$ at time s. We wish to study this process restricted to the range $[0, \infty)$ by reflection at 0. As the inhomogeneous analog of the classical "Lévy presentation of reflecting Brownian motion" ([27] I.14), this reflecting process B^t may be constructed via

$$B^t(s) = W^t(s) - \min_{0 \le s' \le s} W^t(s'), \ s \ge 0. \tag{2}$$

Now append a point process of "marks" of intensity $B^t(s)$ at time s. Informally,

$$P(\text{some mark during } [s, s+ds] | B^t(u), u \le s) = B^t(s)ds. \tag{3}$$

Precisely, the number $N^t(s)$ of marks in $[0, s]$ is characterized as the counting process for which

$$N^t(s) - \int_0^s B^t(u)du \text{ is a martingale,}$$

though we shall use the more intuitive "infinitesimal" notation. An *excursion* γ of B^t is a time interval $[l(\gamma), r(\gamma)]$ such that $B^t(l(\gamma)) = B^t(r(\gamma)) = 0$ and $B^t(s) > 0$ on $l(\gamma) < s < r(\gamma)$. The excursion has length $|\gamma| = r(\gamma) - l(\gamma)$ and contains some number $\mu(\gamma) \ge 0$ of marks. Write Γ^t for the set of excursions of B^t. A stochastic calculus calculation (Lemma 25, which like other such calculations is deferred to section 5) implies that we can order excursions by length, i.e. write $\Gamma^t = \{\gamma_j, j \geq 1\}$ so that the lengths $|\gamma_j|$ are decreasing. This in turn specifies a joint distribution

$$((|\gamma_j|, \mu(\gamma_j)), j \geq 1)$$

of lengths and mark-counts of excursions.

1.2 Critical random graphs

The random graph model $\mathcal{G}(n, P(\text{edge}) = p(n))$ and its variants are perhaps the most studied model in probabilistic combinatorics, and the $n \to \infty$ asymptotics of component sizes are a classical object of study. A fundamental result going back to Erdős and Rényi [11, 12] says that when $p(n) = a/n$ for fixed $0 < a < \infty$, then
(i) for $a < 1$ the largest component size is $\Theta(\log n)$
(ii) for $a > 1$ the largest component has size $\Theta(n)$ while the second largest component size is $\Theta(\log n)$
(iii) for $a = 1$ the largest and second largest components both have size $\Theta(n^{2/3})$.
Much attention has been paid to "the emergence of the giant component" as $p(n)$ increases through $1/n$: see Bollobás [7] for results up to 1984 and Janson et al. [16] and Łuczak et al. [19] for references to subsequent work. The following "folk theorem" is implicit in this recent work, though apparently has never been proved explicitly in the form we state. Define the number of "surplus" edges in a component as

$$\text{surplus} = (\text{number of edges}) - (\text{number of vertices } -1) \geq 0 \ .$$

Fix $-\infty < t < \infty$.

Folk Theorem 1 *Let $C_n^t(1) \geq C_n^t(2) \geq \ldots$ be the ordered component sizes of $\mathcal{G}(n, n^{-1} + tn^{-4/3})$, and let $\sigma_n^t(j)$ be the surplus of the corresponding component. Then as $n \to \infty$*

$$\left(n^{-2/3} C_n^t(j), \sigma_n^t(j)), j \geq 1\right) \ \xrightarrow{d}\ \left((C^t(j), \sigma^t(j)), j \geq 1\right) = (\mathbf{C}^t, \boldsymbol{\sigma}^t) \ \textit{say}$$

for some limit $(\mathbf{C}^t, \boldsymbol{\sigma}^t)$ with $0 < C^t(j) < \infty$ and $0 \leq \sigma^t(j) < \infty$ a.s. for each $j \geq 1$.

(For the moment regard "convergence" as convergence with respect to product topology, i.e. convergence of initial segments of arbitrary fixed length). It is well known that the $n^{-4/3}$ scaling is "correct" for the emergence of the giant component, in that

$$C^t(1) \stackrel{d}{\to} 0 \text{ as } t \to -\infty$$

$$C^t(1) \stackrel{d}{\to} \infty \text{ but } C^t(2) \stackrel{d}{\to} 0 \text{ as } t \to +\infty.$$

Our main result, Theorem 3, has the following Corollary. Define $l^2_{\searrow}$ to be the set of infinite sequences $\mathbf{x} = (x_1, x_2, \ldots)$ with $x_1 \geq x_2 \geq \ldots \geq 0$ and $\sum_i x_i^2 < \infty$, and give $l^2_{\searrow}$ the natural metric $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$. We may regard the finite sequence $\mathbf{C}^t_n = (C^t_n(j), j \geq 1)$ as a random element of $l^2_{\searrow}$ by appending entries of size zero.

Corollary 2 *Folk Theorem 1 is true, and the convergence $n^{-2/3}\mathbf{C}^t_n \stackrel{d}{\to} \mathbf{C}^t$ holds with respect to the $l^2_{\searrow}$ topology. Moreover the limit $((C^t(j), \sigma^t(j)), j \geq 1)$ is distributed as the sequence $((|\gamma_j|, \mu(\gamma_j)), j \geq 1)$ of lengths and mark-counts of excursions of B^t.*

We now start to describe how this result arises.

1.3 The breadth-first walk

We first describe a deterministic construction, illustrated in figure 1. Consider a graph on vertices $\{1, 2, \ldots, n\}$. We shall specify the *breadth-first ordering* $(v(1), \ldots, v(n))$ of the vertices, and an associated integer-valued sequence $(z(i); 0 \leq i \leq n)$ we shall call *breadth-first walk*. The first of these notions is of course a well-known algorithmic procedure. In brief: order components $\mathcal{C}_1, \mathcal{C}_2, \ldots$ so that the smallest-labeled vertex $w_1, w_2, \ldots$ in each component has $w_1 < w_2 < \ldots$ (and call w_j the *root* of $\mathcal{C}_j$); within each component, order by levels (equal distance from root); within each level, use the original order of labels. Put figure 1 near here

We now elaborate this construction. Given $\{v(1), \ldots, v(i)\}$, define the *neighbor* set $\mathcal{N}_i$ to be the set of vertices outside $\{v(1), \ldots, v(i)\}$ which are neighbors of some vertex inside $\{v(1), \ldots, v(i)\}$. We can also define the set of *children* of $v(i)$ to be the set $\mathcal{N}_i \setminus \mathcal{N}_{i-1}$. Order the components $\mathcal{C}_i$ as described above, and consider the first component $\mathcal{C}_1$. Define $v(1) = w_1$ and let $v(2), \ldots, v(1 + |\mathcal{N}_1|)$ be the neighbors of $v(1)$, in increasing order. Inductively for $i = 2, \ldots, |\mathcal{C}_1|$, list the children (if any) of $v(i)$ in increasing order as $v(i+|\mathcal{N}_{i-1}|),\ldots,v(i+|\mathcal{N}_i|)$. After exhausting the first component, we set $v(|\mathcal{C}_1|+1)=w_2$, the root of the second component. List the children of w_2 as $v(|\mathcal{C}_1|+2), v(|\mathcal{C}_1|+3),\ldots$ and continue the induction through the second component. Repeat for subsequent components.

Write $c(i)=|\mathcal{N}_i\setminus\mathcal{N}_{i-1}|$ for the number of children of $v(i)$. Now define *breadth-first walk* via

$$z(0)=0;\quad z(i)-z(i-1)=c(i)-1,\ i=1,\ldots,n. \tag{4}$$

An equivalent definition is provided by (5) below. Write

$$\zeta(j)=|\mathcal{C}_1|+\ldots+|\mathcal{C}_j|$$

$$\zeta^{-1}(i)=\min\{j:\zeta(j)\geq i\}$$

so that $\zeta^{-1}(i)$ is the index of the component containing $v(i)$. We assert

$$z(i)=|\mathcal{N}_i|-\zeta^{-1}(i),\ i=1,2,\ldots,n. \tag{5}$$

To verify by induction that (5) and (4) are equivalent, we need to show

$$|\mathcal{N}_i|-|\mathcal{N}_{i-1}|=c(i)-1+\zeta^{-1}(i)-\zeta^{-1}(i-1),\ i=2,3,\ldots,n.$$

Suppose $v(i-1)$ is not the last vertex in its component. Then $\zeta^{-1}(i)=\zeta^{-1}(i-1)$ and because $v(i)\in\mathcal{N}_{i-1}$ we have $|\mathcal{N}_i|-|\mathcal{N}_{i-1}|=c(i)-1$.If on the other hand $v(i-1)$ is the last vertex in its component, then $\zeta^{-1}(i)=1+\zeta^{-1}(i-1)$ and, because $|\mathcal{N}_{i-1}|=0$, we have $|\mathcal{N}_i|-|\mathcal{N}_{i-1}|=c(i)$,as required.

Because $|\mathcal{N}_i|=0$ only if $v(i)$ is the last vertex of its component, (5) implies

$$z(\zeta(j))=-j;\quad z(i)\geq -j \text{ for all } \zeta(j)<i<\zeta(j+1). \tag{6}$$

It follows that we can reconstruct component sizes and indices from the walk via

$$\zeta(j)=\min\{i:z(i)=-j\}$$

$$|\mathcal{C}_j|=\zeta(j)-\zeta(j-1)$$

$$\zeta^{-1}(i)=1-\min_{j\leq i-1} z(j). \tag{7}$$

Our first main result says what happens when we apply this construction to the near-critical random graph.

Theorem 3 *Let* $(Z_n^t(i), 0 \le i \le n)$ *be the breadth-first walk associated with* $\mathcal{G}(n, n^{-1} + tn^{-4/3})$*. Rescale by defining*

$$\bar{Z}_n^t(s) = n^{-1/3} Z_n^t(\lfloor n^{2/3} s \rfloor).$$

Then $\bar{Z}_n^t \stackrel{d}{\to} W^t$ *as* $n \to \infty$*.*

Interpret $Z_n^t(\lfloor n^{2/3}s \rfloor)$ as $Z_n^t(n)$ for $s > n^{1/3}$. Recall that convergence of processes on the infinite interval $0 \le s < \infty$ (see e.g. [13]) is "uniform on finite intervals" rather than uniform over the infinite interval. We in fact need an extension (31) of Theorem 3 in which surplus edges are indicated as "marks" on the breadth-first walk. This (easy) extension in stated and proved in section 2.2.

The proof of Theorem 3 (section 2.1) uses standard methodology from stochastic process theory (the functional CLT for continuous-time martingales) but does not require any non-trivial facts about random graphs. Essentially, one just has to compute first-order asymptotics for the conditional mean and variance of increments of $Z_n^t(\cdot)$ – see (20,21). Details of how Corollary 2 follows are given in section 2.3, but should be intuitively clear from property (6) of breadth-first walk. The point is that component sizes are coded as lengths of path-segments above past-minima; these converge to lengths of excursions of W^t above past-minima, which are just lengths of excursions of B^t above 0.

Martin-Löf [22] and Spencer [28] have independently given results relating to random graphs and Brownian-type processes which may be viewed as aspects of Theorem 3: see section 6.

1.4 A non-uniform random graph model

It turns out that the "component size" part of Corollary 2 can be extended to a non-uniform random graph model. This extension, Proposition 4, will be proved in section 3 by modifying where needed the proof of Theorem 3 and Corollary 2.

For a positive real vector $\mathbf{x} = (x_1, \ldots, x_n)$ and $q > 0$ define a random graph $\mathcal{W}(\mathbf{x}, q)$ on vertices $\{1, 2, \ldots, n\}$ as follows. Each pair (i, j) of vertices is an edge with probability $1 - \exp(-qx_ix_j)$, independently for distinct pairs.

Interpret x_i as the *size* of vertex i, and so say a component $\mathcal{C}$ of $\mathcal{W}(\mathbf{x}, q)$ has *size* $C = \sum_{i \in \mathcal{C}} x_i$. Given $\mathbf{x}$, define

$$\sigma_r = \sum_i x_i^r, \quad r \ge 1$$

$$x_* = \max_i x_i.$$

Proposition 4 *For each n let $\mathbf{x}^{(n)}$ be a finite positive vector and let $q^{(n)} > 0$. Let $(C^{(n)}(j); j \geq 1)$ be the ordered component sizes of $\mathcal{W}(\mathbf{x}^{(n)}, q^{(n)})$. Suppose that, as $n \to \infty$,*

$$\frac{\sigma_3^{(n)}}{(\sigma_2^{(n)})^3} \to 1 \tag{8}$$

$$q^{(n)} - \frac{1}{\sigma_2^{(n)}} \to t \tag{9}$$

$$\frac{x_*^{(n)}}{\sigma_2^{(n)}} \to 0 \tag{10}$$

for some $-\infty < t < \infty$. Then

$$(C^{(n)}(j); j \geq 1) \xrightarrow{d} (C^t(j); j \geq 1)$$

with respect to the $l^2_\searrow$ topology defined in section 1.2, where $(C^t(j); j \geq 1)$ are the ordered excursion lengths of B^t.

Discussion. We have built the scaling into the hypotheses rather than the conclusion. For any constant $a > 0$, the component sizes of $\mathcal{W}(a\mathbf{x}, a^{-2}q)$ are a times the component sizes of $\mathcal{W}(\mathbf{x}, q)$, so we can always assume (8) by scaling, and then (9) is the "essential" hypothesis. Note that after scaling, the classical model $\mathcal{G}(n, n^{-1} + tn^{-4/3})$ corresponds to the case

$$x_i^{(n)} = n^{-2/3}, \quad \sigma_2^{(n)} = n^{-1/3}, \quad \sigma_3^{(n)} = n^{-1}, \quad q^{(n)} = n^{1/3} + t. \tag{11}$$

Assuming (8), hypothesis (10) is equivalent to (dropping the n's) $x_*^3/\sigma_3 \to 0$, i.e. the requirement that the contribution to σ_3 from individual terms be asymptotically negligible. Clearly some such "asymptotic negligibility" condition is necessary, to eliminate cases $(\mathbf{x}^{(n)}, q^{(n)}) = (\mathbf{x}, q)\ \forall n$, but it is not clear whether (10) itself is necessary.

Since $\sigma_3 \leq x_* \sigma_2$, hypotheses (8,10) imply

$$\sigma_2^{(n)} \to 0. \tag{12}$$

Then

$$\begin{aligned} \max_{i,j} q x_i x_j &\leq q x_* \sqrt{\sigma_2} \\ &= O\left(\tfrac{x_*}{\sigma_2}\sqrt{\sigma_2}\right) \text{ by } (9,12) \\ &\to 0 \text{ by } (10,12). \end{aligned}$$

So under the hypotheses of Proposition 4, the individual edge-probabilities tend to zero, and are asymptotic to qx_ix_j.

1.5 The multiplicative coalescent

There is a natural process describing $\mathcal{W}(\mathbf{x}, q)$ as q varies, which we now describe in somewhat different notation. Fix $\mathbf{x} \in l^2_\searrow$. For each pair $i < j$, create an exponential (rate 1) r.v. ξ_{ij}, independent for different pairs. Given t, consider the graph where there exists an edge (i, j) iff $\xi_{ij} \leq tx_ix_j$: this is a construction of $\mathcal{W}(\mathbf{x}, t)$, simultaneously for all $0 \leq t < \infty$. Let $X_i(\mathbf{x}, t)$ be the size of the i'th largest component of this $\mathcal{W}(\mathbf{x}, t)$, and let

$$\mathbf{X}(\mathbf{x}, t) = (X_i(\mathbf{x}, t), i \geq 1). \tag{13}$$

Picture the typical state $\mathbf{y} = (y_i)$ as a collection of "clusters" of sizes $y_1, y_2, \ldots$. For an initial vector $\mathbf{x}$ of finite length, $\mathbf{X}(\mathbf{x}, t)$ is a continuous-time finite-state (the state-space depending on the initial $\mathbf{x}$) Markov chain whose dynamics are described by

$$\text{for each pair of clusters of sizes } (x, y)\text{, they merge at rate } xy$$
$$\text{into a cluster of size } x + y. \tag{14}$$

The construction (13) makes sense for an infinite initial vector $\mathbf{x}$, if we allow individual cluster sizes to be infinite. In section 4.2 we shall prove that the natural state space for this process is $l^2_\searrow$, in the following sense.

Proposition 5 *For each $\mathbf{x} \in l^2_\searrow$ the construction (13) yields a Markov process $(\mathbf{X}(\mathbf{x}, t); t \geq 0)$ on $l^2_\searrow$. This process has the Feller property: for each t,*

$$\textit{if } \mathbf{x}^{(n)} \to \mathbf{x} \textit{ then } \mathbf{X}(\mathbf{x}^{(n)}, t) \stackrel{d}{\to} \mathbf{X}(\mathbf{x}, t).$$

In other words, there is a well-defined continuous-time Markov process on $l^2_\searrow$ whose dynamics are informally described by (14). We call this process the *multiplicative coalescent*, and write it as $\mathbf{X}(t)$. See section 6.4 for some background on general stochastic coalescence models. Note that there exists a "constant" version of the multiplicative coalescent

$$\mathbf{X}(t) = (y, 0, 0, 0, \ldots), \ -\infty < t < \infty \tag{15}$$

for constant $y \geq 0$. Corollary 2 and the Feller property imply (see Corollary 24) there exists a process $(\mathbf{X}^*(t); -\infty < t < \infty)$, the *standard multiplicative* *coalescent*, such that for each t we have $\mathbf{X}(t) \stackrel{d}{=} \mathbf{C}^t$, where $\mathbf{C}^t$ is the asymptotic joint distribution of rescaled component sizes (equivalently: the distribution of excursion lengths of B^t) appearing in Folk Theorem 1 and Corollary 2. In a companion paper [4] it is shown the constant process, the standard process and certain processes derived from the standard process are (up to scaling and mixtures) the *only* versions of the multiplicative coalescent which exist for time $-\infty < t < \infty$. The basic idea is that, for a non-constant version $(\mathbf{X}(t); -\infty < t < \infty)$, the distribution of $\mathbf{X}(0)$ is by construction just the vector of component sizes of $\mathcal{W}(\mathbf{X}(-n), n)$, so to prove it is the standard version it suffices to verify the hypotheses of Proposition 4, and this can be done via stochastic calculus.

Intuitively, a typical state for the multiplicative coalescent is an *unordered* collection of cluster sizes. The convention of defining the multiplicative coalescent as a $l^2_\searrow$-valued process by using the decreasing ordering was intended as the most elementary way to specify an explicit state space. But as will be discussed in section 3.3 there are advantages in using the more sophisticated notion of *size-biased random order*. The representation in terms of B^t shows that for each t the standard multiplicative coalescent $\mathbf{X}^*(t)$ has infinite total size, so that our $l^2_\searrow$ setup is not just "generalization for its own sake".

1.6 Summary

In case this introduction seems disjointed, we summarize the three main points.

(a) Theorem 3 and Corollary 2 link the excursion lengths of B^t to the asymptotic component sizes in the near-critical random graph process.

(b) The multiplicative coaleescent process is defined, and shown (Proposition 5) to be a Feller process. As an immediate consequence of this and (a) we deduce the existence of the *standard* multiplicative coalescent on $-\infty < t < \infty$.

(c) Proposition 4 extends (a) to certain non-uniform random graph models, which is a key ingredient in the proof in [4] that the standard process is essentially the only version of the multiplicative coalescent which starts at time $-\infty$ with infinitesimally small clusters.

One could give much more discussion of background material and known results, but it is time to start proving the new results, so we defer further discussion until section 6.

2 Weak convergence arguments

2.1 Proof of Theorem 3

We start with a technical point. Recall from section 1.3 the construction of breadth-first walk $(z(i), 0 \le i \le n)$ in the deterministic setting. We need to interpolate between integer times, and one always-available way to do this is via $z(s) = z(\lfloor s \rfloor)$. Motivated by our later extension to the non-uniform case, we make a slightly different definition which is tailored to our specific setting. Take independent uniform$(0,1)$ r.v.'s $(U_{i,j}, 1 \le i \le n, 1 \le j \le c(i))$ and then for each i set

$$z(i-1+u) = z(i-1) - u + \sum_j 1_{(U_{i,j} \le u)}, \ 0 \le u \le 1. \tag{16}$$

So $z(i) = z(i-1) - 1 + c(i)$ as required.

Here is a mental picture of the construction. After step $i-1$ we have a list $(v(1), \ldots, v(j))$ of length $j = i - 1 + |\mathcal{N}_{i-1}|$ consisting of vertices $\{v(1), \ldots, v(i-1)\}$ and their neighbors. In section 1.3 we envisaged adding the children of $v(i)$ to this list at "time" i, but now we envisage adding them at uniform random times over $[i-1, i]$.

We shall prove Theorem 3 for $Z_n^t(s)$ defined using this interpolation, with the rescaling

$$\bar{Z}_n^t(s) = n^{-1/3} Z_n^t(n^{2/3}s). \tag{17}$$

This obviously implies the stated form of the Theorem.

To ease the notation let us drop the superscript t from random variables. We may write (by general theory; we calculate explicit expressions later in Lemma 6)

$$Z_n = M_n + A_n \tag{18}$$

where $M_n(\cdot)$ is a martingale and $A_n(\cdot)$ is a continuous, bounded variation process. Then write

$$M_n^2 = Q_n + B_n \tag{19}$$

where $Q_n(\cdot)$ is a martingale and B_n is a continuous increasing process. (All these processes start with value 0 at $s = 0$.) We shall show that as $n \to \infty$ with s_0 fixed,

$$n^{-1/3} \sup_{s \le n^{2/3} s_0} \left| A_n(s) + n^{-1}s^2/2 - n^{-1/3}st \right| \xrightarrow{p} 0 \tag{20}$$

$$n^{-2/3} B_n(n^{2/3}s_0) \xrightarrow{p} s_0 \tag{21}$$

$$n^{-2/3} E \sup_{s \leq n^{2/3} s_0} |M_n(s) - M_n(s-)|^2 \to 0. \tag{22}$$

Rescaling as at (17) to define $\bar{A}_n, \bar{M}_n, \bar{B}_n$, these assertions become

$$\sup_{s \leq s_0} \left| \bar{A}_n(s) - \rho(s) \right| \xrightarrow{p} 0, \text{ where } \rho(s) = st - s^2/2$$

$$\bar{B}_n(s_0) \xrightarrow{p} s_0$$

$$E \sup_{s \leq s_0} |\bar{M}_n(s) - \bar{M}_n(s-)|^2 \to 0.$$

The latter two conditions are the hypotheses of the functional CLT for continuous-time martingales (e.g. [13] Theorem 7.1.4(b)), whose conclusion is $\bar{M}_n \xrightarrow{d} W$, standard Brownian motion. Then the former condition implies

$$\bar{Z}_n = \bar{M}_n + \bar{A}_n \xrightarrow{d} W + \rho(\cdot) = W^t$$

which is Theorem 3.

By construction the jumps of $Z_n(\cdot)$, and hence of $M_n(\cdot)$, have size exactly 1, and so (22) is obvious. So the issue is to prove (20,21). We now calculate the explicit form of the decompositions (18,19). Following (7), write

$$\zeta_n^{-1}(i) = 1 - \min_{u \leq i-1} Z_n(u). \tag{23}$$

Lemma 6

$$A_n(u) = \int_0^u (a_n(s) \ - 1) \ ds$$

$$B_n(u) = \int_0^u a_n(s) \ ds$$

where

$$a_n(s) = (n - s - \zeta_n^{-1}(\lceil s \rceil) - Z_n(s)) \frac{p(n)}{1 - (s - \lfloor s \rfloor) p(n)} \quad .$$

Proof. From the definition of Z_n as a process with drift -1 and with jumps $+1$ when a new edge appears, it is clear that the formulas for A_n and B_n hold for $a_n(s)$ defined by

$$a_n(s) \ ds = P(\text{some new edge appears during } [s, s+ds] | Z_n(u), u \leq s).$$

An elementary calculation shows: if an event occurs with probability $p(n)$ and, conditionally on occurrence, it occurs at a random time uniform on $(0, 1)$, then

$$\begin{aligned} P(\text{event occurs during } [s, s+ds] \ \mid \text{does not occur before } s) \\ = \frac{p(n)}{1 - sp(n)}\ ds. \end{aligned} \tag{24}$$

So by construction of breadth-first walk,

$$a_n(s) = (n - \nu_n(s)) \frac{p(n)}{1 - (s - \lfloor s \rfloor) p(n)} \tag{25}$$

where $\nu_n(s)$ is the number of vertices at time s which are ineligible to be children on $v(\lceil s \rceil)$. When we start looking for children of $v(i)$ at time $i-1$, the number of ineligible vertices is

$$\nu_n(i-1) = i - 1 + |\mathcal{N}_{i-1}| + (\zeta^{-1}(i) - \zeta^{-1}(i-1))$$

where the final term takes care of $v(i)$ itself. By (5) we can rewrite this as

$$\nu_n(i-1) = i - 1 + \zeta_n^{-1}(i) + Z_n(i-1). \tag{26}$$

So at time $i - 1 + u$ (for $0 < u < 1$) the number ineligible is

$$\begin{aligned} \nu_n(i-1+u) &= i - 1 + \zeta_n^{-1}(i) + Z_n(i-1) + \sum_j 1_{(U_{i,j} \le u)} \\ &= (i - 1 + u) + \zeta_n^{-1}(i) + Z_n(i-1+u) \end{aligned}$$

by our interpolation convention. In other words, $\nu_n(s) = s + \zeta_n^{-1}(\lceil s \rceil) + Z_n(s)$, establishing the Lemma. □

The expressions for A_n and B_n in Lemma 6 allow us to rewrite (21) as

$$n^{-2/3} A_n(n^{2/3} s_0) \ \xrightarrow{p}\ 0$$

which is plainly weaker than (20). So it suffices to verify (20). Consider $a_n'(s)$ defined as "a_n without the denominator", i.e.

$$a_n'(s) = (n - s - \zeta_n^{-1}(\lceil s \rceil) - Z_n(s)) p(n).$$

It is straightforward to see that $|a_n'(s) - a_n(s)| = O(1/n)$, uniformly in s. Now

$$a_n'(s) - 1 = \left(1 - \frac{s + \zeta_n^{-1}(\lceil s \rceil) + Z_n(s)}{n}\right)\left(1 + \frac{t}{n^{1/3}}\right) - 1$$

and this leads to the bound (for $n^{1/3} > |t|$)

$$\left| a_n'(s) - 1 + \frac{s}{n} - \frac{t}{n^{1/3}} + \frac{st}{n^{4/3}} \right| \leq 2 \ \frac{\zeta_n(\lceil s \rceil) + |Z_n(s)|}{n}. \qquad (27)$$

Integrating over s, and using (23),

$$\left| A_n(s) + \frac{s^2}{2n} - \frac{st}{n^{1/3}} + \frac{s^2 t}{2n^{4/3}} \right| \leq \frac{4s \max_{u \leq s} |Z_n(u)|}{n} + O(s/n).$$

So the proof of (20) reduces to proving

$$n^{-2/3} \sup_{s \leq n^{2/3} s_0} |Z_n(s)| \ \xrightarrow{p} \ 0.$$

In fact we shall prove the stronger result

$$n^{-1/3} \sup_{s \leq n^{2/3} s_0} |Z_n(s)| \text{ is stochastically bounded as } n \to \infty. \qquad (28)$$

This requires a routine argument using truncation and the martingale optional sampling theorem. Fix a large constant K and define

$$T_n^* = \min\{s : |Z_n(s)| > K n^{1/3}\}$$

$$T_n = \min(T_n^*, s_0 n^{2/3}).$$

Then

$$\begin{aligned} EM_n^2(T_n) &= EB_n(T_n) \text{ by the optional sampling theorem} \\ &= E \int_0^{T_n} a_n(s) ds \\ &\leq \int_0^{s_0 n^{2/3}} \frac{np(n)}{1 - (s - \lfloor s \rfloor) p(n)} \ ds \text{ by (25)} \\ &\leq 2 s_0 n^{2/3} \end{aligned}$$

the final inequality for n sufficiently large. Then

$$\begin{aligned} E|Z_n(T_n)| &\leq E|M_n(T_n)| + E|A_n(T_n)| \\ &\leq (2s_0)^{1/2} n^{1/3} + E \int_0^{T_n} |a_n(s) - 1| \ ds. \end{aligned}$$

Using (27) and (23),

$$E\int_0^{T_n} |a_n(s)-1|\ ds$$

$$\leq E\int_0^{s_0 n^{2/3}} |a'_n(s)-a_n(s)|\ ds+\int_0^{s_0 n^{2/3}} \left|\frac{s}{n}-\frac{t}{n^{1/3}}+\frac{st}{n^{4/3}}\right|\ ds+(s_0 n^{2/3})4\frac{Kn^{1/3}}{n}.$$

This leads to a bound, for large n,

$$E|Z_n(T_n)| \leq \alpha n^{1/3} + 4s_0 K$$

where α depends on (s_0, t) but not on (n, K). Then

$$P(\sup_{s\leq s_0 n^{2/3}} |Z_n(s)| > Kn^{1/3}) = P(|Z_n(T_n)| > Kn^{1/3}) \leq \frac{\alpha}{K}+\frac{4s_0}{n^{1/3}}$$

establishing (28).

2.2 Surplus edges

Along with the breadth-first walk $(Z_n^t(s); 0 \leq s \leq n)$, we may associate with $\mathcal{G}(n, n^{-1}+tn^{-4/3})$ a counting process $(N_n^t(s); 0 \leq s \leq n)$ which increases by 1 at each occurrence of an excess edge. To analyze this process, recall that (26) gave an expression for the number $\nu_n(i-1)$ of ineligible vertices when we start looking for children of $v(i)$. Of these, i vertices (that is, $v(1), \ldots, v(i)$) cannot have edges to $v(i)$, and the remaining $\nu_n(i-1)-i$ vertices are candidates for having an excess edge to $v(i)$. Each of these candidates will have an excess edge with probability $p(n) = n^{-1}+tn^{-4/3}$. Representing each excess edge as a "mark" at a uniform random time on $[i-1, i]$, then the counting process N_n^t has rate (i.e. conditional intensity)

$$\frac{p(n)}{1-(s-\lfloor s\rfloor)p(n)}(\nu_n(\lfloor s\rfloor)-\lfloor s\rfloor). \tag{29}$$

(This rate is in fact a slight overcount, but we argue later that the error is asymptotically negligible). Using (23,26) this rate becomes

$$\frac{p(n)}{1-(s-\lfloor s\rfloor)p(n)}\left(Z_n^t(\lfloor s\rfloor)-\min_{u\leq\lfloor s\rfloor} Z_n^t(u)\right). \tag{30}$$

Now rescale the counting process via

$$\bar{N}_n^t(s) = N_n^t(n^{2/3}s).$$

The rate for this rescaled process, in terms of the rescaled walk $\bar{Z}^t_n$, is the rate in (30) multiplied by $n^{2/3} \times n^{1/3} = n$, and since $np(n) \to 1$ the rate is asymptotic to

$$\bar{Z}^t_n(s) - \min_{u \leq s} \bar{Z}^t_n(u).$$

But by Theorem 3 this process converges to $W^t(s) - \min_{u \leq s} W^t(u) = B^t(s)$. By routine weak convergence theory, such convergence of rates is enough to extend Theorem 3 to give joint convergence of processes:

$$(\bar{Z}^t_n(s), \bar{N}^t_n(s); s \geq 0) \ \xrightarrow{d}\ (W^t(s), N^t(s); s \geq 0) \tag{31}$$

for N^t defined at (3).

Formula (29) slightly overestimates the chance that a vertex $v(i)$ has two or more surplus edges, but even this overestimated chance that some one of the first $O(n^{2/3})$ vertices has two or more excess edges must tend to zero, otherwise the limit N^t in (31) would have multiple coincident points.

2.3 Proof of Corollary 2

We shall prove the part of Corollary 2 dealing with component sizes: the full result incorporating component surpluses is just the same argument, invoking the joint convergence (31).

Recall that the reflecting process B^t is derived from W^t via (2): excursions of B^t above 0 are excursions of W^t above its past-minimum. There are two issues in deducing Corollary 2 from Theorem 3. The first is to check that excursions of the limit process are matched by excursions of the breadth-first walks (representing components of the random graph); the second is to check that no components of size $\Omega(n^{2/3})$ are overlooked by virtue of their positions in the walk going off to $+\infty$.

The first issue is mostly handled by the following deterministic lemma, whose straightforward proof we omit.

Lemma 7 *Suppose $f : [0, \infty) \to R$ is continuous. Let $\mathcal{E}$ be the set of non-empty intervals $e = (l, r)$ such that*

$$f(r) = f(l) = \min_{s \leq l} f(s); \quad f(s) > f(l) \text{ for } l < s < r.$$

Suppose that, for intervals $e_1, e_2 \in \mathcal{E}$ with $l_1 < l_2$ we have

$$f(l_1) > f(l_2). \tag{32}$$

Suppose also that the complement of $\cup_{e\in\mathcal{E}}(l,r)$ *has Lebesgue measure zero. Let* $\Xi = \{(l, r-l) : (l,r) \in \mathcal{E}\}$. *Now let* $f_n \to f$ *uniformly on bounded intervals. Suppose* $(t_{n,i}, i \geq 1)$ *satisfy*
(i) $0 = t_{n,1} < t_{n,2} < t_{n,3} \ldots$ *and* $\lim_{i\to\infty} t_{n,i} = \infty$
(ii) $f_n(t_{n,i}) = \min_{u \leq t_{n,i}} f_n(u)$
(iii) $\max_{i:t_{n,i}\leq s_0}(f_n(t_{n,i}) - f_n(t_{n,i+1})) \to 0$ *as* $n \to \infty$, *for each* $s_0 < \infty$.
Write $\Xi^{(n)} = \{(t_{n,i}, t_{n,i+1} - t_{n,i}); i \geq 1\}$. *Then* $\Xi^{(n)} \to \Xi$ *as* $n \to \infty$.

Here and below, we regard Ξ and $\Xi^{(n)}$ as point processes on $[0,\infty)\times(0,\infty)$, and convergence is the natural notion of vague convergence of counting measures on $[0,\infty)\times(0,\infty)$: see e.g. [17].

Define $\gamma(n,i)$ by: $v(\gamma(n,i))$ is the last vertex in the $i-1$'st component of the random graph encountered by breadth-first walk. Let $C_{n,i}$ be the size of this i'th component.

Lemma 8 *Let* $\Xi^{(\infty)}$ *be the point process with points*

$$\{(l(\gamma), |\gamma|), \gamma \text{ an excursion of } B^t\ \}.$$

Let $\Xi^{(n)}$ *be the point process with points* $\{(n^{-2/3}\gamma(n,i), n^{-2/3}C_{n,i}) : i \geq 1\}$. *Then* $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$ *as* $n \to \infty$.

Proof. Using (2), $\Xi^{(\infty)}$ is just the Ξ of Lemma 7 applied to W^t. And writing $t_{n,i} = n^{-2/3}\gamma(n,i)$, the $\Xi^{(n)}$ in Lemma 8 is just the $\Xi^{(n)}$ in Lemma 7 applied to Z_n. Theorem 3 gave $\bar{Z}_n \xrightarrow{d} W^t$, and by the Skorohod representation theorem ([27] II.86.1) it is enough to verify the hypotheses of Lemma 7. It is standard that the hypotheses on f hold a.s. for Brownian motion W, and hence they hold a.s. for W^t by the absolute continuity given by Cameron-Martin-Girsanov theory. And conditions (i)-(iii) follow from construction of breadth-first walk (recall (6)). $\square$

The subject of Corollary 2 is the decreasing ordering of $\{n^{-2/3}C_{n,i}, i \geq 1\}$, i.e. of the second coordinates of the points in $\Xi^{(n)}$. To deduce Corollary 2 from Lemma 8 requires some extra work. Consider

$$T(y) = \min\{s : W^t(s) = -y\}$$

$$T_n(y) = \min\{i : Z_n(i) = -\lfloor yn^{1/3}\rfloor\}.$$

Note that by step $T_n(y)$ the breadth-first walk has encountered all vertices labeled $\{1, 2, \ldots, \lfloor yn^{1/3}\rfloor\}$ in the original labeling. Theorem 3 implies $n^{-2/3}T_n(y) \xrightarrow{d} T(y)$. Since $T(y) \to \infty$ as $y \to \infty$, we have established a restricted version of Corollary 2 in which we consider only excursions of B^t starting before T_{y_0} and components with minimal vertex-label (in the original labeling) is $\leq y_0 n^{1/3}$, for some fixed y_0. Now Corollary 2 itself will follow from the next lemma.

Lemma 9 *Let $p(n, y, \delta)$ be the chance that $\mathcal{G}(n, n^{-1} + tn^{-4/3})$ contains a component of size $\geq \delta n^{2/3}$ which does not contain any vertex i with $1 \leq i \leq yn^{1/3}$. Then*

$$\lim_{y \to \infty} \limsup_{n} p(n, y, \delta) = 0 \text{ for all } \delta > 0.$$

Proof. Fix $\delta > 0$. For an interval I define $q(n, I)$ to be the mean number of components of size $\geq \delta n^{2/3}$ whose minimal vertex is in $n^{1/3} I$. Conditional on component sizes, the labels $\{1, 2, \ldots, n\}$ of the vertices of the random graph are in random order. For a component having size $vn^{2/3}$, write $\chi_n = n^{-1/3}$(label of minimal vertex). Then $\chi_n \xrightarrow{d}$ exponential (rate v), implying $P(\chi_n > y) \sim \frac{e^{-vy}}{1-e^{-v}} P(\chi_n \leq 1)$. By summing over components,

$$\limsup_{n} \frac{q(n, [y, \infty))}{q(n, [0, 1])} \leq \sup_{v \geq \delta} \frac{e^{-vy}}{1 - e^{-v}} = \frac{e^{-\delta y}}{1 - e^{-\delta}}.$$

Because $p(n, y, \delta) \leq q(n, [y, \infty))$, it suffices to prove

$$\sup_{n} q(n, [0, 1]) < \infty. \tag{33}$$

But results in the random graphs literature imply $\sup_n q(n, [0, \infty)) < \infty$. (Boris Pittel (personal communication) observes this follows from bounds on the numbers of tree components, unicyclic components and complex components given in [19] Theorem 2 and Lemma 2.1). So by quoting that result, we finish the proof of Corollary 2. Note that for the analogous part of the proof of Proposition 4, the non-uniform case, we will need a novel argument (see section 3.4) and that argument could be used here to make our proof of Corollary 2 be independent of existing random graphs results.

3 The non-uniform case

In this section we give the proof of Proposition 4. The proof follows the general lines of the proof of Theorem 3 and Corollary 2, with modified definitions and extra technical lemmas where needed.

3.1 Breadth-first walk

In section 1.3 we defined the breadth-first walk for an arbitrary deterministic unweighted graph. In the current "weighted" setting, it is more convenient to give a simultaneous construction of the random graph $\mathcal{W}(\mathbf{x} = \{x_1, \ldots, x_n\}, q)$ and its interpolated walk $z(u)$ analogous to (16). Figure 2 illustrates part of the construction. For each ordered pair (i,j), $i \neq j$ let $U_{i,j}$ have exponential(qx_j) distribution, independent over pairs. Choose $v(1)$ by size-biased sampling, i.e. vertex v is chosen with probability proportional to x_v. Let $\{v : U_{v(1),v} \leq x_{v(1)}\}$ be the set of children of $v(1)$, and order these children as $v(2), v(3), \ldots$ so that $U_{v(1),v(i)}$ is increasing. Start the walk $z(\cdot)$ with $z(0) = 0$ and let

Put figure 2 near here

$$z(u) = -u + \sum_v x_v 1_{(U_{v(1),v} \leq u)}, \ 0 \leq u \leq x_{v(1)}.$$

So

$$z(x_{v(1)}) = -x_{v(1)} + \sum_{v \text{ child of } v(1)} x_v.$$

Inductively, write $\tau_{i-1} = \sum_{j \leq i-1} x_{v(j)}$. If $v(i)$ is in the same component as $v(1)$, then the set

$$\{v \notin \{v(1), \ldots, v(i-1)\} : v \text{ is a child of one of } \{v(1), \ldots, v(i-1)\}\}$$

consists of $v(i), \ldots, v(l(i))$ for some $l(i) \geq i$. Let the children of $v(i)$ be $\{v \notin \{v(1), \ldots, v(l(i))\} : U_{v(i),v} \leq x_{v(i)}\}$, and order them as $v(l(i)+1), v(l(i)+2), \ldots$ such that $U_{v(i),v}$ is increasing. Set

$$z(\tau_{i-1} + u) = z(\tau_{i-1}) - u + \sum_{v \text{ child of } v(i)} x_v \ 1_{(U_{v(i),v} \leq u)}, \ 0 \leq u \leq x_{v(i)}. \quad (34)$$

After exhausting the component containing $v(1)$, choose the next vertex by size-biased sampling, i.e. each available vertex v is chosen with probability proportional to x_v. Continue.

This construction yields a forest on the vertices $\{1, \ldots, n\}$, an ordering $v(1), \ldots, v(n)$ of the vertices and a walk $(z(u); 0 \leq u \leq \sum_v x_v)$. Add extra edges (i,j) for each pair such that $i < j \leq l(i)$ and $U_{v(i),v(j)} \leq x_{v(i)}$. It is easy to check that the resulting random graph is $\mathcal{W}(\mathbf{x}, q)$. Briefly, for any pair of vertices (i,j) one (say i) appears first in the ordering, and then (i,j) is an edge iff $U_{i,j} \leq x_i$, which happens with probability $1 - \exp(-qx_jx_i)$.

Note also that by construction the ordering $(v(i))$ is the *size-biased* random ordering (see section 3.3) of the vertices.

In figure 2 the weight of vertex $v(i)$ is given below the label $v(i)$. A helpful way to think about the construction, illustrated in figure 2, is to picture the successive vertices $v(i)$ occupying successive intervals of the "time" axis, the length of the interval for v being the weight x_v. During this time interval we "search for" children of $v(i)$, and any such child $v(j)$ causes a jump in $z(\cdot)$ of size $x_{v(j)}$. The time of this jump is the *birth time* $\beta(j)$ of $v(j)$, which in this case (i.e. provided $v(j)$ is not the first vertex of its component) is $\beta(j) = \tau_{i-1} + U_{v(i),v(j)}$. These jumps are superimposed on a constant drift of rate -1. If $v(j)$ is the first vertex of its component, its birth time is the start of its time interval: $\beta(j) = \tau_{j-1}$.

The relationship between the walk and the graph is less simple than in the uniform case. In the uniform case we could reconstruct the graph (up to vertex labels) from the walk, but this is not true in the non-uniform case, because it is not clear from the walk where one vertex's time interval ends and the next's begins. In particular, the relationship between the walk and the components is less simple than (7). But we do have an analog of (6): a component consists of vertices $\{v(i), v(i+1), \ldots, v(j)\}$, and the walk $z(\cdot)$ satisfies

$$z(\tau_j) = z(\tau_{i-1}) - x_{v(i)}, \quad z(u) \geq z(\tau_j) \text{ on } \tau_{i-1} < u < \tau_j.$$

Note that by construction, the order in which the *components* appear in the breadth-first walk is also size-biased order.

3.2 Asymptotics

We now apply the construction above to $\mathcal{W}(\mathbf{x}^{(n)}, q^{(n)})$ satisfying the hypotheses of Proposition 4. Write

$$Z_n(s), \ 0 \leq s \leq \sum_v x_v^{(n)}$$

for the breadth-first walk. Rescale to define

$$\bar{Z}_n(s) = \sqrt{\frac{\sigma_2^{(n)}}{\sigma_3^{(n)}}} \, Z_n(s).$$

We use the same decompositions $Z_n = M_n + A_n$; $M_n^2 = Q_n + B_n$ as at (18,19), and analogous to (20 - 22) we seek to show that for fixed s_0

$$\sqrt{\frac{\sigma_2^{(n)}}{\sigma_3^{(n)}}} \sup_{s \leq s_0} \left| A_n(s) + \frac{q^{(n)}\sigma_3^{(n)}}{2\sigma_2^{(n)}} s^2 - (q^{(n)}\sigma_2^{(n)} - 1)s \right| \xrightarrow{p} 0 \tag{35}$$

$$\frac{\sigma_2^{(n)}}{\sigma_3^{(n)}} B_n(s_0) \xrightarrow{p} s_0 \tag{36}$$

$$\frac{\sigma_2^{(n)}}{\sigma_3^{(n)}} E \sup_{s \leq s_0} |M_n(s) - M_n(s-)|^2 \to 0. \tag{37}$$

Then exactly as before we deduce the analog of Theorem 3,

Proposition 10 $\bar{Z}_n \xrightarrow{d} W^t$.

Since the maximum jump is x_*, property (37) is immediate from hypotheses (8,10). Using hypotheses (8,9) and (12), we see that (35,36) reduce to

$$\sup_{s \leq s_0} \left| \frac{A_n(s)}{\sigma_2^{(n)}} + \tfrac{1}{2}s^2 - ts \right| \xrightarrow{p} 0 \tag{38}$$

$$\frac{B_n(s_0)}{(\sigma_2^{(n)})^2} \xrightarrow{p} s_0. \tag{39}$$

The rest of section 3.2 is devoted to proving (38,39) and hence Proposition 10. Here are the explicit forms of A_n, B_n, analogous to Lemma 6. To ease notation we shall mostly omit the superscripts n.

Lemma 11

$$dA_n(s) = -\ ds\ + q(\sigma_2 - Q_2(s) - \tilde{Q}_2(s))\ ds$$

$$dB_n(s) = q(\sigma_3 - Q_3(s) - \tilde{Q}_3(s))\ ds$$

where, for $\tau_{i-1} \leq s < \tau_i$,

$$Q_2(s) = \sum_{j \leq i} x_{v(j)}^2, \quad Q_3(s) = \sum_{j \leq i} x_{v(j)}^3$$

$$\tilde{Q}_2(s) = \sum_{j > i, \beta(j) < s} x_{v(j)}^2, \quad \tilde{Q}_3(s) = \sum_{j > i, \beta(j) < s} x_{v(j)}^3.$$

Proof. The proof follows the proof of Lemma 6 but is simpler: the set of ineligible vertices at s is exactly $\{j : \beta(j) < s\}$. $\square$

Because $\frac{1}{2}s^2 - ts = \int_0^s (u-t)\, du$, showing (38) reduces to showing

$$\sup_{u \le s_0} |d(u)| \xrightarrow{p} 0$$

where

$$d(u) = \frac{-1 + q(\sigma_2 - Q_2(u) - \tilde{Q}_2(u))}{\sigma_2} + (u - t).$$

Using hypotheses (8,9) and (12), this in turn reduces to proving Lemmas 12 and 13 below. Similarly, (39) reduces to showing

$$\frac{Q_3(s_0) + \tilde{Q}_3(s_0)}{\sigma_2^3} \xrightarrow{p} 0.$$

Since $Q_3(s_0) \le x_* Q_2(s_0)$ and $\tilde{Q}_3(s_0) \le x_* \tilde{Q}_2(s_0)$, this also follows from Lemma 12 and 13, using hypothesis (10).

Lemma 12 $\sup_{u \le s_0} \tilde{Q}_2(u)/\sigma_2^2 \xrightarrow{p} 0.$

Lemma 13

$$\sup_{u \le s_0} \left| \tfrac{\sigma_2}{\sigma_3} Q_2(u) \ - u \right| \xrightarrow{p} 0.$$

We will see that Lemma 12 reduces to the analog of (28). Lemma 13 is trivial in the uniform setting, and so will require a new argument.

Proof of Lemma 12. $\tilde{Q}_2(s) \le x_* \tilde{Q}_1(s)$, where for $\tau_{i-1} \le s < \tau_i$

$$\tilde{Q}_1(s) = \sum_{j > i, \beta(j) < s} x_{v(j)}.$$

So by hypothesis (10) it is enough to prove

$$\frac{1}{\sigma_2} \sup_{s \le s_0} \tilde{Q}_1(s) \text{ is stochastically bounded as } n \to \infty.$$

We assert

$$\tilde{Q}_1(s) = Z(s) - (\tau_i - s) - (Z(\tau_{\mu-1}) - x_{v(\mu)}), \ \tau_{i-1} \le s \le \tau_i \qquad (40)$$

where $v(\mu)$ is the first vertex of the component containing $v(i)$. Indeed, (40) is true instantaneously after $\tau_{\mu-1}$, when both sides are zero. Traversing the component, when vertex $v(j)$ occurs as a child of some $v(i)$, both sides increase by $x_{v(j)}$, while at times $s = \tau_{i-1}$, both sides decrease by $x_{v(i)}$, and at other times, both sides stay unchanged. This verifies (40). Using hypothesis (10) again, proving the lemma reduces to proving

$$\frac{1}{\sigma_2} \sup_{s \leq s_0} |Z(s)| \text{ is stochastically bounded as } n \to \infty.$$

But this can be established by following the proof of (28) – we omit details.

Proof of Lemma 13. We exploit the fact that the $(v(i))$ are in size-biased order. Introduce an artificial time parameter θ, let (T_i) be independent with exponential(x_i) distribution, and consider

$$\begin{aligned} D_1(\theta) &= \sum_j x_j 1_{(T_j \leq \theta)} \quad - \sigma_2 \theta \\ D_2(\theta) &= \sum_j x_j^2 1_{(T_j \leq \theta)} \quad - \sigma_3 \theta \\ D_0(\theta) &= \tfrac{\sigma_2}{\sigma_3} D_2(\theta) \ - D_1(\theta). \end{aligned}$$

Ordering vertices i according to the (increasing) values of T_i gives the size-biased ordering. So the process

$$\left(\frac{\sigma_2}{\sigma_3} Q_2(\tau_i) - \tau_i, i \geq 0 \right)$$

is distributed as the process $(D_0(\theta_i), i \geq 0)$, where

$$\theta_i = \min\{\theta : T_j \leq \theta \text{ for exactly } i \text{ different } j\text{'s } \}.$$

So the quantity featuring in Lemma 13 can be rewritten as

$$D(s_0) = \sup \{ |D_0(\theta)| \ : D_1(\theta) + \sigma_2 \theta \leq s_0 \} . \tag{41}$$

For $u = 1, 2$ the process $D_u(\theta)$ is a supermartingale, and so by a maximal inequality ([27] Lemma 2.54.5), for $\varepsilon > 0$

$$\varepsilon P(\sup_{\theta' \leq \theta} |D_u(\theta')| > 3\varepsilon) \leq 3E|D_u(\theta)| \leq 3 \left(|ED_u(\theta)| + \sqrt{\text{var } D_u(\theta)} \right).$$

Now

$$\begin{aligned} |ED_2(\theta)| &= -ED_2(\theta) \\ &= \sum_j x_j^2 (x_j \theta - 1 + \exp(-x_j \theta)) \\ &\leq \sum_j x_j^2 \ (x_j \theta)^2 / 2 \\ &= \theta^2 \sigma_4 / 2 \end{aligned}$$

$$\begin{aligned} \text{var } D_2(\theta) &= \sum_j x_j^4 P(T_j \le \theta) P(T_j > \theta) \\ &\le \sum_j x_j^4 \ (x_j \theta) \\ &= \theta \sigma_5. \end{aligned}$$

Similarly

$$|ED_1(\theta)| \le \theta^2 \sigma_3/2; \text{ var } D_1(\theta) \le \theta\sigma_3. \tag{42}$$

Combining these bounds,

$$\tfrac{\varepsilon}{3} P(\sup_{\theta' \le \theta} |D_0(\theta')| > 6\varepsilon) \le \frac{\sigma_2}{\sigma_3} \left(\frac{\theta^2 \sigma_4}{2} + \theta^{1/2} \sigma_5^{1/2} \right) + \frac{\theta^2 \sigma_3}{2} + \theta^{1/2} \sigma_3^{1/2}.$$

Setting $\theta = 2s_0/\sigma_2$ and using the bounds $\sigma_4 \le x_* \sigma_3$, $\sigma_5 \le x_*^2 \sigma_3$, the bound becomes

$$O\left(\frac{x_*}{\sigma_2} + \frac{\sigma_2^{1/2} x_*}{\sigma_3^{1/2}} + \frac{\sigma_3}{\sigma_2^2} + \frac{\sigma_3^{1/2}}{\sigma_2^{1/2}} \right)$$

and this $\to 0$ using (8,10,12). So in view of (41) it is enough to show that, for $\theta = 2s_0/\sigma_2$,

$$P(D_1(\theta) + \sigma_2 \theta \le s_0) \to 0.$$

But this follows from Chebyshev's inequality and (42).

3.3 Size-biased ordering for random sequences in l^2

Central to this paper is the notion of convergence of "random unordered sets of positive numbers", where the index sets are not fixed. In this section we discuss representations of unordered sets which permit discussion of convergence. In particular, a general result on convergence of size-biased orderings (Proposition 15) will enable us to complete the proof of Proposition 4 in section 3.4.

For a countable index set Γ write $l_+^2(\Gamma)$ for the set of sequences $\mathbf{x} = (x_\gamma; \gamma \in \Gamma)$ such that each $x_\gamma \ge 0$ and $\sum_\gamma x_\gamma^2 < \infty$. Recall that $l_\searrow^2$ denotes the set of sequences $\mathbf{x} = (x_i; i = 1, 2, \ldots)$ such that $x_1 \ge x_2 \ge \ldots \ge 0$ and $\sum_\gamma x_\gamma^2 < \infty$. Give $l_\searrow^2$ the natural metric $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$. Writing ord $: l_+^2(\Gamma) \to l_\searrow^2$ for the "decreasing ordering" map, it is elementary that

$$d^2(\text{ord } \mathbf{x}, \text{ord } \mathbf{y}) \le \sum_\gamma (x_\gamma - y_{b(\gamma)})^2 \tag{43}$$

for any bijection b between the index sets of $\mathbf{x}$ and $\mathbf{y}$, a fact we will use without explicit mention in section 4.

Given a random collection $\mathbf{Y} = \{Y_\gamma : \gamma \in \Gamma\}$ in $l^2_+(\Gamma)$, where Γ may depend on the realization, the most elementary way to represent $\mathbf{Y}$ without mentioning Γ is to use ord $\mathbf{Y}$ to create a random element of $l^2_\searrow$. But an alternative, in some ways more elegant and mathematically natural, is to use the notion of *size-biased order*. As well as being classical in statistical sampling theory, size-biasing in the l^1 setting has been prominent in recent mathematical work in probabilistic combinatorics: see [24] for an extensive list of references. Given $\mathbf{Y} = \{Y_\gamma : \gamma \in \Gamma\}$ with each $Y_\gamma > 0$, construct r.v.'s (ξ_γ) such that, conditional on $\mathbf{Y}$, the (ξ_γ) are independent and ξ_γ has exponential(Y_γ) distribution. These define a random linear ordering on Γ, i.e. $\gamma_1 \le \gamma_2$ iff $\xi_{\gamma_1} \le \xi_{\gamma_2}$. In the "$l^1$" case, i.e. when $\sum_\gamma Y_\gamma < \infty$ a.s., this coincides with the elementary notion of size-biased order: there is a first element $\gamma_{(1)}$ such that

$$P(\gamma_{(1)} = \gamma | \mathbf{Y}) = \frac{Y_\gamma}{\sum_{\gamma'} Y_{\gamma'}}.$$

But the ordering makes sense without any l^1 assumption, although (as with the ordering of the positive rationals) there will be no first element in the ordering. Consider the following construction. For $0 \le a < \infty$ define

$$S(a) = \sum_{\gamma : \xi_\gamma < a} Y_\gamma. \tag{44}$$

Note that

$$E(S(a)|\mathbf{Y}) = \sum_\gamma Y_\gamma (1 - \exp(-aY_\gamma)) \le a \sum_\gamma Y_\gamma^2.$$

So if $\mathbf{Y} \in l^2_+(\Gamma)$ then we have $S(a) < \infty$ a.s.. So we can define $S_\gamma = S(\xi_\gamma) < \infty$ and finally define the *size-biased point process* (SBPP) associated with $\mathbf{Y}$ to be the set $\Xi = \{(S_\gamma, Y_\gamma) : \gamma \in \Gamma\}$. So Ξ is a random element of $\mathcal{M}$, the space of configurations of points on $[0,\infty) \times (0,\infty)$ with only finitely many points in each compact rectangle $[0, s_0] \times [\delta, 1/\delta]$. Note that Ξ depends only on the ordering, rather than the actual values, of the ξ's. Clearly Ξ has the properties

$$\text{if } (s,y) \in \Xi \text{ then } \sum_{\{y' : (s',y') \in \Xi, s' < s\}} y' = s \tag{45}$$

$$\max\{y : (s,y) \in \Xi \text{ for some } s > s_0\} \xrightarrow{p} 0 \text{ as } s_0 \to \infty. \tag{46}$$

Writing π for the "project onto the y-axis" map

$$\pi(\{(s_\gamma, y_\gamma)\}) = \{y_\gamma\} \tag{47}$$

we can recover ord $\mathbf{Y}$ from Ξ via ord $\mathbf{Y}$ = ord $\pi(\Xi)$.

We now turn to notions of convergence. On $l^2_\searrow$, convergence $\mathbf{x}^{(n)} \to \mathbf{x}$ shall mean convergence with respect to the metric d, except when we explicitly write $\mathbf{x}^{(n)} \to_{\text{prod}} \mathbf{x}$ to indicate convergence in the product topology: $\lim_n x_i^{(n)} = x_i \ \forall i$. The set $\mathcal{M}$ has its own natural topology – pointwise convergence, uniform over compact subsets. These deterministic notions extend in the usual way to notions of convergence in distribution for random elements of the spaces (see e.g. [17] for discussion of convergence in distribution for point processes).

The following straightforward lemma provides a connection between these modes of convergence in distribution.

Lemma 14 *Let $\mathbf{Y}^{(n)} \in l^2_+(\Gamma^n)$ for each $1 \le n \le \infty$, and let $\Xi^{(n)}$ be the associated SBPP. The following are equivalent.*

(a) ord $\mathbf{Y}^{(n)} \xrightarrow{d}$ ord $\mathbf{Y}^{(\infty)}$.

(b) ord $\mathbf{Y}^{(n)} \xrightarrow{d}_{\text{prod}}$ ord $\mathbf{Y}^{(\infty)}$ *and*

$$\lim_{\delta \downarrow 0} \limsup_n P\left(\sum_{Y_\gamma^{(n)} \le \delta} (Y_\gamma^{(n)})^2 > \varepsilon \right) = 0 \textit{ for each } \varepsilon > 0. \tag{48}$$

(c) $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, and (48) holds.

We need a subtle variation of these ideas. Suppose we know $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, but we do not know that $\Xi^{(\infty)}$ is the SBPP of some $\mathbf{Y}^{(\infty)}$, and we do not know (48). Can we still deduce that the assertions of Lemma 14 hold, by imposing only conditions on $\Xi^{(\infty)}$?

Proposition 15 *Let $\mathbf{Y}^{(n)} \in l^2_+(\Gamma^n)$ for each $1 < n \le \infty$, and let $\Xi^{(n)}$ be the associated SBPP. Suppose $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, where $\Xi^{(\infty)}$ is a point process satisfying (45,46) and*

$$\sup\{s : (s, y) \in \Xi^{(\infty)} \textit{ for some } y\} = \infty \textit{ a.s.} \tag{49}$$

Then $\mathbf{Y}^{(\infty)}$ = ord $\pi(\Xi^{(\infty)})$ is in $l^2_\searrow$, and ord $\mathbf{Y}^{(n)} \xrightarrow{d}$ ord $\mathbf{Y}^{(\infty)}$.

The following three examples show that none of the three conditions (45,46,49) can be removed.

(a) Let $\mathbf{Y}^{(n)}$ consist of a fixed $\mathbf{y} \in l^2$ with $\sum_i y_i = \infty$, together with n^2 terms of size $1/n$. Here only (45) fails.

(b) Let $\mathbf{Y}^{(n)}$ consist of n terms of size 1. Here only (46) fails.

(c) Let $\mathbf{Y}^{(n)}$ consist of n^2 terms of size $1/n$. Here $\Xi^{(\infty)}$ is empty, so (45,46) are vacuously satisfied, and only (49) fails.

Proof of Proposition 15. We shall make several uses of the following technical device. If Q_n is some positive real-valued function of $\mathbf{Y}^{(n)}$, then we may assume that one of the two cases

$$Q_n \xrightarrow{p} \infty; \qquad (Q_n) \text{ is tight}$$

holds, because by considering subsequences the general case can be viewed as a mixture of these cases.

First assume a boundedness condition:

$$K \equiv \sup_n \max_\gamma Y_\gamma^{(n)} < \infty. \tag{50}$$

Suppose

$$\sigma^{(n)} \equiv \sum_\gamma (Y_\gamma^{(n)})^2 \xrightarrow{p} \infty. \tag{51}$$

From the definition of $S^{(n)}(a)$ we have, for fixed $\lambda > 0$,

$$E(S^{(n)}(\lambda/\sigma^{(n)})|\mathbf{Y}^{(n)}) = \sum_\gamma Y_\gamma^{(n)} \ (1 - \exp(-Y_\gamma^{(n)}\lambda/\sigma^{(n)})).$$

Since $\sum_\gamma Y_\gamma^{(n)} \left(Y_\gamma^{(n)}/\sigma^{(n)}\right)^2 \leq K/\sigma^{(n)} \xrightarrow{p} 0$ by (50,51), we have

$$E(S^{(n)}(\lambda/\sigma^{(n)})|\mathbf{Y}^{(n)}) \ - \ \lambda \xrightarrow{p} 0.$$

And

$$\begin{aligned} \text{var}\,(S^{(n)}(\lambda/\sigma^{(n)})|\mathbf{Y}^{(n)}) &= \sum_\gamma (Y_\gamma^{(n)})^2 (1 - \exp(-Y_\gamma^{(n)}\lambda/\sigma^{(n)}))\exp(-Y_\gamma^{(n)}\lambda/\sigma^{(n)}) \\ &\leq KE(S^{(n)}(\lambda/\sigma^{(n)})|\mathbf{Y}^{(n)}) \\ &\xrightarrow{p} K\lambda. \end{aligned}$$

Using Chebyshev's inequality, it is easy to deduce that, for arbitrary random τ_n,

$$S^{(n)}(\tau_n) \xrightarrow{p} \infty \text{ iff } \tau_n\sigma^{(n)} \xrightarrow{p} \infty. \tag{52}$$

Now fix $1 > \delta > 0$, let B_n be the set $\{Y_\gamma^{(n)} : \delta \le Y_\gamma^{(n)} \le \delta^{-1}\}$ and write $|B_n|$ for its cardinality. Let $\tau_{n,k}$ be the k'th smallest element of $\{\xi_\gamma : \delta \le Y_\gamma^{(n)} \le \delta^{-1}\}$. If $(\sigma^{(n)}/|B_n|)$ is tight then $|B_n| \xrightarrow{p} \infty$, and $\tau_{n,k} = O(1/|B_n|)$ for fixed k, so $(\tau_{n,k}\sigma^{(n)})$ is tight and then by (52) $(S^{(n)}(\tau_{n,k}), n \ge 1)$ is tight. But then the convergence $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$ implies that $\Xi^{(\infty)}$ has an infinite number of points in $[0,\infty) \times [\delta, \delta^{-1}]$, contradicting hypothesis (46). So we may suppose the other case, $\sigma^{(n)}/|B_n| \xrightarrow{p} \infty$. But in this case $\tau_{n,1}\sigma^{(n)} \xrightarrow{p} \infty$, so $S^{(n)}(\tau_{n,1}) \xrightarrow{p} \infty$, and so $\Xi^{(\infty)}$ has no points in $[0,\infty) \times [\delta, \delta^{-1}]$. This must hold for each $\delta > 0$, so $\Xi^{(\infty)}$ is empty, contradicting (49). This means (51) must be false, and so we may assume the other case, that $(\sigma^{(n)})$ is tight. Since

$$E(S^{(n)}(a)|\mathbf{Y}^{(n)}) \le a \sum_\gamma (Y_\gamma^{(n)})^2 = a\sigma^{(n)}$$

we see that, for fixed a, the sequence $(S^{(n)}(a), n \ge 1)$ is tight. Together with the hypothesis $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, this implies ord $\mathbf{Y}^{(n)} \xrightarrow{d}{}_{\text{prod}}\mathbf{Y}^{(\infty)} \equiv$ ord $\pi(\Xi^{(\infty)})$, and that $\mathbf{Y}^{(\infty)}$ is in $l^2_\searrow$. Passing to a subsequence we may assume

$$\lim_{k\to\infty} \lim_n \sum\{(Y_\gamma^{(n)})^2 : \ Y_\gamma^{(n)} \le 1/k\} = \bar{\sigma}$$

exists, where limits are "in distribution". The convergence ord $\mathbf{Y}^{(n)} \xrightarrow{d}{}_{\text{prod}}\mathbf{Y}^{(\infty)}$ then implies $S^{(n)}(a) \xrightarrow{d} S^{(\infty)}(a) + a\bar{\sigma}$ where $S^{(\infty)}$ is defined in terms of $Y^{(\infty)}$ as at (44). But the convergence $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, and hypothesis (45) on $\Xi^{(\infty)}$, show that $\bar{\sigma} = 0$. In other words, (48) holds, and the conclusion follows from Lemma 14.

This establishes the Proposition under the boundedness assumption (50), and the general case follows using a truncation argument – we omit the details.

3.4 Proof of Proposition 4

We shall apply Proposition 15 to the component sizes $\mathbf{Y}^{(n)}$ of $\mathcal{W}(\mathbf{x}^{(n)}, q^{(n)})$. Let $(Y_u^{(n)}, u = 1, 2, \ldots)$ be the component sizes of $\mathcal{W}(\mathbf{x}^{(n)}, q^{(n)})$, in the order of appearance in breadth-first walk. Write $S_{u-1}^{(n)} = \sum_{j=1}^{u-1} Y_j^{(n)}$. Let $\Xi^{(n)}$ be the point process on $[0,\infty) \times (0,\infty)$ with points at $(S_{u-1}^{(n)}, Y_u^{(n)}), u = 1, 2, \ldots$. Proposition 10 showed $\bar{Z}_n \xrightarrow{d} W^t$. Repeating the argument for Lemma 8 shows $\Xi^{(n)} \xrightarrow{d} \Xi^{(\infty)}$, where $\Xi^{(\infty)}$ is the point process with points

$\{(l(\gamma), |\gamma|), \gamma$ an excursion of $B^t\}$. Standard qualitative properties of Brownian motion establish properties (45,49). We observed in section 3.1 that components appeared in size-biased order in the breadth-first walk. To apply Proposition 15 the only further hypothesis which needs checking is (46), which is a consequence of Lemma 25. The conclusion of the Proposition now establishes Proposition 4.

Proposition 15 and Lemma 14 imply that excursions of B^t appear in size-biased order, a fact we record as follows.

Corollary 16 *The point process $\{(l(\gamma), |\gamma|)\}$ consisting of left endpoints and lengths of excursions γ of B^t is distributed as the size-biased point process $\{(S_\gamma, |\gamma|)\}$ associated with $\{|\gamma|\}$.*

4 Analysis of the multiplicative coalescent

4.1 Preliminaries

Say $\mathbf{x} \in l^2_\searrow$ is *finite-length* if $x_i = 0$ ultimately. For finite-length $\mathbf{x}$, the process $\mathbf{X}(t)$ constructed in (13) as the decreasing-ordered component sizes of $\mathcal{W}(\mathbf{x}, t)$ can clearly be regarded as a $l^2_\searrow$-valued Markov process, which we now call the *multiplicative coalescent*. (The Markov property is a simple consequence of the "memoryless" property for the exponential r.v.'s (ξ_{ij})). In the next section we shall prove Proposition 5, which asserts that for any $\mathbf{x} \in l^2_\searrow$ this construction of $\mathbf{X}(t)$ yields a $l^2_\searrow$-valued process possessing the Feller property.

When $\mathbf{X}(0) = \mathbf{x}$ is finite-length, the dynamics (14) of the multiplicative coalescent can be expressed in martingale form as follows. Let $\mathbf{x}^{(i+j)}$ be the configuration obtained from $\mathbf{x}$ by merging the i'th and j'th clusters, that is $\mathbf{x}^{(i+j)} = (x_1, \ldots, x_{u-1}, x_i + x_j, x_u, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots)$ for some u. Write $\mathcal{F}(t) = \sigma\{\mathbf{X}(u); u \le t\}$. Then

$$E(\Delta g(\mathbf{X}(t))|\mathcal{F}(t)) = \sum_i \sum_{j>i} X_i(t) X_j(t) \left(g(\mathbf{X}^{(i+j)}(t)) - g(\mathbf{X}(t)) \right) \, dt \quad (53)$$

for all $g : l^2_\searrow \to R$ (for *all* g because there are only finitely many possible states). Of course, our "infinitesimal" notation $E(\Delta Y(t)|\mathcal{F}(t)) = A(t)dt$ is just an intuitive way of expressing the rigorous assertion that $M(t) = Y(t) - \int_0^t A(s)ds$ is a local martingale; similarly the notation var $(\Delta Y(t)|\mathcal{F}(t)) = B(t)dt$ means that $M^2(t) - \int_0^t B(s)ds$ is a local martingale. Following a paradigm in modern stochastic process theory, one could seek to *define* the multiplicative coalescent via such a martingale characterization, but that requires technical discussion of the class of g's where the formula should hold. Our "constructive" definition of the multiplicative coalescent finesses that issue, but requires us to give ad hoc justifications of uses of (53) in the infinite-length setting.

4.2 Proof of Proposition 5

$\mathbf{X}(t)$ denotes the multiplicative coalescent started from some initial state $\mathbf{x}$, and $S(t) = \sum_i X_i^2(t)$. When we wish to indicate explicitly the initial state we write $\mathbf{X}(\mathbf{x}, t)$ and $S(\mathbf{x}, t)$.

Our proof involves coupling arguments to bound the effect on $\mathbf{X}(\mathbf{x}, t)$ of changing $\mathbf{x}$, and martingale arguments to bound the effect of changing t. We start with a deterministic coupling lemma. Recall d denotes distance in $l^2_\searrow$.

Lemma 17 *Let $\bar{G}$ be a graph with vertex-weights $(\bar{x}_i)$. Let G be a subgraph of $\bar{G}$ (that is, each edge of G is an edge of $\bar{G}$) with vertex weights $x_i \leq \bar{x}_i$. Let $\bar{\mathbf{a}}$ and $\mathbf{a}$ be the decreasing orderings of the component sizes of $\bar{G}$ and G. Then*

$$d^2(\bar{\mathbf{a}}, \mathbf{a}) \leq \sum_i \bar{a}_i^2 - \sum_i a_i^2$$

provided $\sum_i a_i^2 < \infty$.

Proof. By considering different components of $\bar{G}$ separately and using (43) it is enough to treat the case where $\bar{G}$ is a single component. Then, writing $\bar{a} = \sum_i \bar{x}_i$,

$$d^2(\bar{\mathbf{a}}, \mathbf{a}) = (\bar{a} - a_1)^2 + \sum_{i\geq 2} a_i^2.$$

We need to prove this is $\leq \bar{a}^2 - \sum_{i\geq 1} a_i^2$, and after rearranging we need to prove

$$\bar{a}^2 - (\bar{a} - a_1)^2 \geq a_1^2 + 2\sum_{i\geq 2} a_i^2.$$

The left side increases with $\bar{a}$, and since $\bar{a} \geq a \equiv \sum_i a_i$ it is enough to prove

$$a^2 - (a - a_1)^2 \geq a_1^2 + 2\sum_{i\geq 2} a_i^2.$$

But this holds because the left side equals $a_1^2 + 2\sum_{i\geq 2} a_1 a_i$, and $a_1 \geq a_i$ for $i \geq 2$. □

The construction (13) of $\mathcal{W}(\mathbf{x},t)$ (and hence $\mathbf{X}(\mathbf{x},t)$) in terms of $(\xi_{ij}; 1 \le i < j < \infty)$ works simultaneously for all $\mathbf{x}$. When we want to exploit a joint distribution $(\mathbf{X}(\bar{\mathbf{x}},t), \mathbf{X}(\mathbf{x},t))$ arising in this way, we call it the *ξ-coupling.* Lemma 17 often enables us to bound $l^2_\searrow$-distances for the multiplicative coalescent in terms of the real-valued r.v.'s $S(t)$. In particular,

Corollary 18 .

(a) If $t_1 < t_2$ *then* $d^2(\mathbf{X}(t_1), \mathbf{X}(t_2)) \le S(t_2) - S(t_1)$*on* $\{S(t_1) < \infty\}$.

(b) If $x_i \le \bar{x}_i \ \forall i$ *then the ξ-coupling satisfies* $d^2(\mathbf{X}(\mathbf{x},t), \mathbf{X}(\bar{\mathbf{x}},t)) \le S(\bar{\mathbf{x}},t) - S(\mathbf{x},t)$ *on* $\{S(\mathbf{x},t) < \infty\}$.

(c) For $\mathbf{x} \in l^2_\searrow$ *write* $\mathbf{x}^{(k)} = (x_1, \ldots, x_k)$. *For the ξ-coupling,* $S(\mathbf{x}^{(k)},t) \uparrow S(\mathbf{x},t) \le \infty$ *and* $\mathbf{X}(\mathbf{x}^{(k)},t) \to \mathbf{X}(\mathbf{x},t)$ *on* $\{S(\mathbf{x},t) < \infty\}$.

A formally different use of the same idea is where we have a collection $\bar{\mathbf{y}} = (y_\alpha; \alpha \in \bar{A})$ and a subcollection $\mathbf{y} = (y_\alpha; \alpha \in A \subseteq \bar{A})$. In this setting we can construct $\mathcal{W}(\text{ord } \bar{\mathbf{y}}, t)$ jointly with $\mathcal{W}(\text{ord } \mathbf{y}, t)$ by using the same family $(\xi_{\{\alpha,\beta\}}; \alpha, \beta \in \bar{A})$. Call this the *subgraph-coupling.*

We now turn to martingale estimates. The full form of Lemma 19 will be used in later sections, but for now we need only the "submartingale" assertion for use in Lemma 20. Because a merge of clusters of sizes x_i and x_j causes an increase in S of size $(x_i + x_j)^2 - x_i^2 - x_j^2 = 2x_i x_j$, (53) specializes to

$$E(\Delta f(S(t))|\mathcal{F}(t)) = \sum_i \sum_{j>i} X_i(t)X_j(t) \left(f(S(t) + 2X_i(t)X_j(t)) - f(S(t))\right) dt \,. \tag{54}$$

Lemma 19 *If* $\mathbf{x} = \mathbf{X}(0)$ *is finite-length then the process* $Y(t) = t + \frac{1}{S(t)}$ *is a submartingale. In fact*

$$E(\Delta Y(t)|\mathcal{F}(t)) = \left(\frac{\sum X_i^4(t)}{S^2(t)} + A(t) \right) \, dt \tag{55}$$

where

$$0 \le A(t) \le \frac{2(\sum X_i^3(t))^2}{S^3(t)}.$$

Moreover

$$\text{var } (\Delta Y(t)|\mathcal{F}(t)) \le \frac{2(\sum X_i^3(t))^2}{S^4(t)} \, dt. \tag{56}$$

Proof. Because $\frac{1}{s+2xy} - \frac{1}{s} = \frac{-2xy}{s(s+2xy)}$, applying (54) gives $E(\Delta Y(t)|\mathcal{F}(t)) = (1-Q)dt$ where

$$Q = \sum\sum_{i<j} \frac{2X_i(t)X_j(t)}{S(t)(S(t)+2X_i(t)X_j(t))} \times X_i(t)X_j(t)$$

$$= \sum\sum_{i<j} \frac{2X_i^2(t)X_j^2(t)}{S^2(t)} - \sum\sum_{i<j} 2X_i^2(t)X_j^2(t)\left(\frac{1}{S^2(t)} - \frac{1}{S(t)(S(t)+2X_i(t)X_j(t))}\right)$$

$$= 1 - \frac{\sum X_i^4(t)}{S^2(t)} - \sum\sum_{i<j} 2X_i^2(t)X_j^2(t)\left(\frac{2X_i(t)X_j(t)}{S^2(t)(S(t)+2X_i(t)X_j(t))}\right).$$

This establishes (55). Similarly, var $(\Delta Y(t)|\mathcal{F}(t))$ equals dt times

$$\sum_i\sum_{j>i} X_i(t)X_j(t)\left(\frac{2X_i(t)X_j(t)}{S(t)(S(t)+2X_i(t)X_j(t))}\right)^2$$

and (56) follows. $\square$

Lemma 20 *For* $\mathbf{x} \in l^2_{\searrow}$,

$$P(S(\mathbf{x},t) > s) \le \frac{tsS(\mathbf{x},0)}{s - S(\mathbf{x},0)}, \ s > S(\mathbf{x},0).$$

Proof. First assume $\mathbf{x}$ is finite-length. Write $b = S(\mathbf{x},0)$ and $S(t) = S(\mathbf{x},t)$. From the submartingale property of $Y(t) = t + \frac{1}{S(t)}$ (Lemma 19)

$$\begin{aligned} \frac{1}{b} &\le t + E(1/S(t)) \\ &\le t + \tfrac{1}{s} + E\left(\frac{1}{S(t)} - \frac{1}{s}\right)^+ \\ &\le t + \tfrac{1}{s} + \left(\frac{1}{b} - \frac{1}{s}\right) P(S(t) \le s) \end{aligned}$$

because $S(t)$ is increasing and $S(0) = b$. Rearranging gives the stated inequality. If $\mathbf{x}$ is not finite-length, consider (as in Corollary 18(c)) $\mathbf{x}^{(k)} = (x_1, \ldots, x_k)$. Since $S(\mathbf{x}^{(k)}, t) \uparrow S(\mathbf{x}, t) \le \infty$, the inequality extends from $\mathbf{x}^{(k)}$ to $\mathbf{x}$. $\square$

Remark. The final "extension by truncation" argument finds similar uses later.

For the remainder of this section 4.2 we fix $t > 0$ and study $\mathbf{X}(\mathbf{x}, t)$ as the initial state $\mathbf{x}$ varies.

Lemma 21 *For* $\mathbf{z} \in l^2_\searrow$ *and* $u > 0$ *let* (V_j) *be the decreasing ordering of the component sizes of the random graph on vertices* $\{0; 1, 2, \ldots\}$ *with vertex-weights* $\{u; z_1, z_2, \ldots\}$ *for which*

$$P(\ (0,i) \text{ is an edge }) = 1 - \exp(-tuz_i),\ i \geq 1$$

independently as i varies. Then $E \sum_j V_j^2 < \infty$.

Proof. Write A_i for the event that $(0, i)$ is an edge. Then

$$\sum_j V_j^2 = u^2 + 2u \sum_{i \geq 1} z_i 1_{A_i} + \sum_{i \geq 1} \sum_{j \geq 1} z_i z_j 1_{A_i \cap A_j} 1_{(j \neq i)} + \sum_{i \geq 1} z_i^2$$

and so

$$E \sum_j V_j^2 \leq u^2 + 2tu^2 \sum_{i \geq 1} z_i^2 + t^2 u^2 \sum_{i \geq 1} \sum_{j \geq 1} z_i^2 z_j^2 + \sum_{i \geq 1} z_i^2 < \infty.$$

□

We can now show that for any $\mathbf{x} \in l^2_\searrow$ we have $S(\mathbf{x}, t) < \infty$ a.s. and hence $\mathbf{X}(\mathbf{x}, t)$ is $l^2_\searrow$-valued. Write $\mathbf{x}^{[k]} = (x_k, x_{k+1}, \ldots)$ and consider the subgraph coupling of the random graphs $\mathcal{W}(\mathbf{x}^{[k]}, t)$. Lemma 20 implies $P(S(\mathbf{x}^{[k]}, t) < \infty) > 1 - t \sum_{i \geq k} x_i^2$ and hence $P(S(\mathbf{x}^{[k]}, t) < \infty \text{ for some } k) = 1$. Lemma 21 shows

$$\text{if } S(\mathbf{x}^{[j]}, t) < \infty \text{ then } S(\mathbf{x}^{[j-1]}, t) < \infty \text{ a.s.}$$

and so by backwards induction for $j = k, k-1, \ldots, 1$ we have $S(\mathbf{x}, t) < \infty$ a.s..

We now start to prove the Feller property by recording a "Fatou-like" lemma.

Lemma 22 *Suppose* $\mathbf{x}^{(n)} \to \mathbf{x}$ *in* $l^2_\searrow$*. Then, in the* ξ*-coupling,*

$$\liminf_{n \to \infty} S(\mathbf{x}^{(n)}, t) \geq S(\mathbf{x}, t) \text{ a.s. } . \tag{57}$$

To prove $\mathbf{X}(\mathbf{x}^{(n)}, t) \xrightarrow{d} \mathbf{X}(\mathbf{x}, t)$ *it suffices to prove there is some coupling for which*

$$\lim_{n \to \infty} P(S(\mathbf{x}^{(n)}, t) - S(\mathbf{x}, t) > \varepsilon) = 0,\ \text{all } \varepsilon > 0. \tag{58}$$

Proof. Let $A_{ij}^{(n)}$ (resp. A_{ij}) be the indicator of the event "vertices i and j are in the same component of $\mathcal{W}(\mathbf{x}^{(n)}, t)$" (resp. $\mathcal{W}(\mathbf{x}, t)$). Then, in the ξ-coupling,

$$\liminf_n A_{ij}^{(n)} \geq A_{ij} \text{ a.s. .} \tag{59}$$

Because if i and j are in the same component of $\mathcal{W}(\mathbf{x}, t)$ then they are linked by a finite path, each of whose edges (k, l) has $\xi_{kl} \leq t x_k x_l$. The only way (59) can fail is if $\xi_{kl} = t x_k x_l$ for some edge, which has probability zero. Now let $\mathcal{C}^{(n)}$ be the class of *modified* components C of $\mathcal{W}(\mathbf{x}^{(n)}, t)$, where i and j are in the same modified component if they are in the same component of $\mathcal{W}(\mathbf{x}^{(n)}, t)$ and are also in the same component of $\mathcal{W}(\mathbf{x}, t)$. Write $B_{ij}^{(n)}$ for the indicator of the event "i and j are in the same modified component". Using (59),

$$B_{ij}^{(n)} \leq A_{ij}; \quad \lim_{n\to\infty} B_{ij}^{(n)} = A_{ij} \text{ a.s. .}$$

So for fixed k,

$$\lim_{n\to\infty} \sum_{C\in\mathcal{C}^{(n)}} \left(\sum_{i\in C, i\leq k} x_i \right)^2 = \sum_{C\in\mathcal{C}} \left(\sum_{i\in C, i\leq k} x_i \right)^2 \text{ a.s.}$$

where $\mathcal{C}$ is the set of components of $\mathcal{W}(\mathbf{x}, t)$. Since $\mathbf{x}^{(n)} \to \mathbf{x}$ we get

$$\lim_{n\to\infty} \sum_{C\in\mathcal{C}^{(n)}} \left(\sum_{i\in C, i\leq k} x_i^{(n)} \right)^2 = \sum_{C\in\mathcal{C}} \left(\sum_{i\in C, i\leq k} x_i \right)^2 \text{ a.s.}$$

Letting $k \to \infty$,

$$\liminf_{n\to\infty} \sum_{C\in\mathcal{C}^{(n)}} \left(\sum_{i\in C} x_i^{(n)} \right)^2 \geq S(\mathbf{x}, t).$$

Since the modified components are a refinement of the original components of $\mathcal{W}(\mathbf{x}^{(n)}, t)$,

$$\sum_{C\in\mathcal{C}^{(n)}} \left(\sum_{i\in C} x_i^{(n)} \right)^2 \leq S(\mathbf{x}^{(n)}, t)$$

and we have established (57). Now if (58) holds for some coupling then, in view of (57), it must hold for the ξ-coupling and then (by a standard subsequence argument) we may suppose $S(\mathbf{x}^{(n)}, t) \to S(\mathbf{x}, t)$ a.s.. Now it is routine to see that, for $\mathbf{Y}^{(n)}, \mathbf{Y}$ in l^2_+, to prove ord $\mathbf{Y}^{(n)} \to$ ord $\mathbf{Y}$ a.s. in $l^2_\searrow$ it suffices to prove

$$\sum_j \left(y_j^{(n)}\right)^2 \to \sum_j y_j^2 \text{ a.s.}$$

$$\liminf_{n\to\infty} y_i^{(n)} \geq y_i \text{ a.s. } \forall i. \tag{60}$$

Thus the desired convergence $\mathbf{X}(\mathbf{x}^{(n)}, t) \to \mathbf{X}(\mathbf{x}, t)$ a.s. will hold provided we verify (60) for

$$y_j^{(n)} = \text{ size of component containing } j \text{ in } \mathcal{W}(\mathbf{x}^{(n)}, t),$$

if j is the smallest-labeled vertex in that component, and $y_j^{(n)} = 0$ if not. (Define y_j similarly in terms of $\mathcal{W}(\mathbf{x}, t)$). But this is clear from (59), applied to each i in the component of $\mathcal{W}(\mathbf{x}, t)$ containing j. $\Box$

The next lemma gives the key estimate we shall use in verifying (58).

Lemma 23 *Let $(z_i, 1 \leq i \leq n)$ be strictly positive vertex-weights, and let $1 \leq m < n$. Consider the bipartite random graph $\mathcal{B}$ on vertices $\{1, 2, \ldots, m\} \cup \{m+1, \ldots, n\}$ defined by: for each pair (i, j) with $1 \leq i \leq m < j \leq n$, the edge (i, j) is present with probability $1 - \exp(-tz_iz_j)$, independently for different pairs. Write $\alpha_1 = \sum_{i=1}^m z_i^2$, $\alpha_2 = \sum_{i=m+1}^n z_i^2$. Let (Z_i) be the sizes of components of $\mathcal{B}$. Then*

$$\varepsilon P(\sum_i Z_i^2 > \alpha_1 + \varepsilon) \leq (2t(\alpha_1 + \varepsilon) + (t(\alpha_1 + \varepsilon))^2)\alpha_2, \ \varepsilon > 0.$$

Proof. For $m \leq k \leq n$ let $\mathcal{B}_k$ be the subgraph of $\mathcal{B}$ on vertices $\{1, \ldots, k\}$ and let Q_k be the sum of squares of component sizes of $\mathcal{B}_k$. So $Q_m = \alpha_1$. Let $A_i, 1 \leq i \leq m$ be the events that $(i, m+1)$ is an edge of $\mathcal{B}$. Then

$$Q_{m+1} - Q_m = 2\sum_{i=1}^m z_i z_{m+1} 1_{A_i} + \sum_{i=1}^m \sum_{j=1}^m z_i z_j 1_{A_i \cap A_j} 1_{(j \neq i)}$$

and so

$$\begin{aligned} E(Q_{m+1} - Q_m) &\leq 2t\sum_{i=1}^m z_i^2 z_{m+1}^2 + t^2 z_{m+1}^2 \sum_{i=1}^m \sum_{j=1}^m z_i^2 z_j^2 1_{(j\neq i)} \\ &\leq (2tQ_m + t^2 Q_m^2) z_{m+1}^2. \end{aligned}$$

Similarly,

$$E(Q_{k+1} - Q_k | \mathcal{B}_k) \leq (2tQ_k + t^2 Q_k^2) z_{k+1}^2, \ m \leq k \leq n.$$

In other words,

$$M_k \equiv Q_k - \alpha_1 - \sum_{j=1}^{k-1} (2tQ_j + t^2 Q_j^2) z_{j+1}^2, \ m \leq k \leq n$$

is a supermartingale with $M_m = 0$. Give $\varepsilon > 0$ set $T = \min\{k : Q_k > \alpha_1 + \varepsilon\}$. Then $EM_{\min(T,n)} \leq 0$ by the optional sampling theorem. So

$$E(Q_{\min(T,n)} - \alpha_1) \leq (2t(\alpha_1 + \varepsilon) + (t(\alpha_1 + \varepsilon))^2)\alpha_2.$$

But $Q_{\min(T,n)} - \alpha_1 > \varepsilon$ on $\{Q_n > \alpha_1 + \varepsilon\}$, establishing the lemma. □

To complete the proof of the Feller property, we need to prove (58). Let $\mathbf{x}^{(n)} \to \mathbf{x}$ in $l^2_\searrow$. Write $\mathbf{x}^{(n,k)}$ for the decreasing ordering of $\{x_i, i \geq 1\} \cup \{x_i^{(n)}, i \geq k\}$. Consider the subgraph-coupling of $\mathcal{W}(\mathbf{x}^{(n,k)}, t)$ and $\mathcal{W}(\mathbf{x}, t)$. We assert

$$\lim_{k \to \infty} \limsup_{n \to \infty} P(S(\mathbf{x}^{(n,k)}, t) - S(\mathbf{x}, t) > \varepsilon) = 0, \ \text{all } \varepsilon > 0. \tag{61}$$

The point is that we can construct $\mathcal{W}(\mathbf{x}^{(n,k)}, t)$ from $\mathcal{W}(\mathbf{x}, t)$ and $\mathcal{W}(\mathbf{y}^{(n,k)}, t)$, where $\mathbf{y}^{(n,k)} = (x_i^{(n)}, i \geq k)$, via the procedure of Lemma 23. Now Lemma 23 extends by truncation to infinite graphs, and implies

$$\begin{aligned} &\varepsilon P(S(\mathbf{x}^{(n,k)}, t) - S(\mathbf{x}, t) > \varepsilon | S(\mathbf{x}, t), S(\mathbf{y}^{(n,k)}, t)) \\ &\leq \left(2t(S(\mathbf{x}, t) + \varepsilon) + t^2(S(\mathbf{x}, t) + \varepsilon)^2\right) S(\mathbf{y}^{(n,k)}, t). \end{aligned}$$

Now we know $S(\mathbf{x}, t) < \infty$ a.s.. And because $x^{(n)}$ is convergent in $l^2_\searrow$, $S(\mathbf{y}^{(n,k)}, 0) \equiv \sum_{i \geq k} (x_i^{(n)})^2$ satisfies $\lim_k \limsup_n S(\mathbf{y}^{(n,k)}, 0) = 0$ and so Lemma 20 implies

$$\lim_{k \to \infty} \limsup_{n \to \infty} P(S(\mathbf{y}^{(n,k)}, t) > s) = 0, \text{all } s > 0.$$

These estimates imply (61). Now write $\mathbf{z}^{(n,k)}$ for the decreasing ordering of $\{x_i, 1 \leq i < k\} \cup \{x_i^{(n)}, i \geq k\}$. In the subgraph-coupling we have $S(\mathbf{z}^{(n,k)}, t) \leq S(\mathbf{x}^{(n,k)}, t)$ and so by (61)

$$\lim_{k \to \infty} \limsup_{n \to \infty} P(S(\mathbf{z}^{(n,k)}, t) - S(\mathbf{x}, t) > \varepsilon) = 0, \ \text{all } \varepsilon > 0. \tag{62}$$

Now fix $\delta > 0$. For given k, for all sufficiently large n we have $x_i^{(n)} x_j^{(n)} t \leq z_i^{(n,k)} z_j^{(n,k)} (t+\delta)\ \forall i,j$ and therefore in the ξ-coupling we have

$$S(\mathbf{x}^{(n)}, t) \leq S(\mathbf{z}^{(n,k)}, t+\delta) \text{ for all sufficiently large } n.$$

Combined with (62) with $t+\delta$, we have constructed a coupling for which

$$\lim_{n\to\infty} P(S(\mathbf{x}^{(n)}, t) - S(\mathbf{x}, t+\delta) > \varepsilon) = 0, \text{ all } \varepsilon > 0.$$

But Lemma 20 implies that as $\delta \downarrow 0$ we have $S(\mathbf{x}, \delta) \downarrow S(\mathbf{x}, 0)$ a.s.. Applying this to $\mathbf{x} = \mathbf{X}(t)$ shows

$$S(t+\delta) \to S(t) \text{ a.s. as } \delta \downarrow 0$$

and (58) follows.

4.3 The standard multiplicative coalescent

Corollary 24 *There exists a version of the multiplicative coalescent* $(\mathbf{X}^*(t); -\infty < t < \infty)$, *the* standard multiplicative coalescent, *such that for each t we have* $\mathbf{X}^*(t) \stackrel{d}{=} \mathbf{C}^t$, *where* $\mathbf{C}^t$ *is the joint distribution of rescaled component sizes or of excursion lengths appearing in Folk Theorem 1 and Corollary 2.*

Proof. Fix $t_1 < t_2$. Consider the "classical" setting (11) of Proposition 4, and let $\mathbf{C}_{t_1}^{(n)}$ and $\mathbf{C}_{t_2}^{(n)}$ be the component sizes obtained with $q^{(n)} = n^{1/3} + t_1$ and with $q^{(n)} = n^{1/3} + t_2$. Then $\mathbf{C}_{t_2}^{(n)}$ is the distribution at time t_2 of the multiplicative coalescent started at time t_1 with distribution $\mathbf{C}_{t_1}^{(n)}$. Proposition 4 asserts that $\mathbf{C}_{t_i}^{(n)} \stackrel{d}{\to} \mathbf{C}^{t_i}$ with respect to the $l^2_\searrow$ topology. The Feller property then verifies that, if we start a multiplicative coalescent at time t_1 with distribution $\mathbf{C}^{t_1}$, then the distribution at time t_2 is $\mathbf{C}^{t_2}$. The Kolmogorov extension theorem now yields the existence of $\mathbf{X}^*$.

5 Stochastic calculus computations with B^t

We shall study what routine stochastic calculus reveals about certain questions concerning B^t and the process N^t of marks. The work in this section is independent of results in earlier sections, except

(i) Lemma 25, or rather the weaker fact that only finitely many excursions are longer than $\delta > 0$, was used in section 3.4 in the proof of Proposition 4.

(ii) The proof of Proposition 27 uses Corollary 16.

5.1 Excursion lengths

We set up our theory of the multiplicative coalescent as an l^2-valued process, so it is nice to have a simple direct argument that the limit process in Corollary 2 is indeed in l^2.

Lemma 25 *Let Γ^t be the set of excursions of B^t, and let $|\gamma|$ be the length of excursion γ. Then $E\sum_{\gamma\in\Gamma^t}|\gamma|^2 < \infty$.*

Proof. For an excursion $\gamma = (l, r)$,

$$|\gamma|^2 = 2\int_l^r (r-u)\,du.$$

So, writing $H_u = \min\{s > 0 : B^t(u+s) = 0\}$,

$$\sum_\gamma |\gamma|^2 = 2\int_0^\infty H_u\,du.$$

So we need to prove $\int_0^\infty EH_u\,du < \infty$. We first claim

$$E(H_u|B^t(u)) \le \frac{B^t(u)}{u-t},\ u > \max(0,t).$$

Because for fixed u, the process $(B^t(u+s), s \ge 0)$ can be coupled with reflecting Brownian motion $(\tilde{B}(s), s \ge 0)$ with constant drift $-(u-t)$, started at the same position $B^t(u)$, in such a way that $B^t(u+s) \le \tilde{B}(s)$ for all s. And for $\tilde{B}$ started at x, the mean hitting time to 0 equals $x/(u-t)$.

Using (2) it is easy to show that $EB^t(\tau) < \infty$ for each τ. Taking $\tau > \max(0, t+1)$ we have, for all $0 \le u \le \tau$,

$$EH_u \le \tau - u + EH_\tau \le EB^t(\tau) + \tau < \infty.$$

So it suffices to prove that for some τ

$$\int_\tau^\infty \frac{EB^t(u)}{u-t}\,du < \infty. \tag{63}$$

Using (2) and invariance of Brownian motion W under time-reversal,

$$\begin{aligned} B^t(s) &\stackrel{d}{=} \sup_{0\le u\le s} (W(u) + (t-s)u + \tfrac{1}{2}u^2) \\ &\le \sup_{0\le u\le s} (W(u) + (t - \tfrac{1}{2}s)u) \text{ by convexity} \\ &\le \sup_{0\le u\le \infty} (W(u) + (t - \tfrac{1}{2}s)u). \end{aligned}$$

Assuming $s > 2t$, the final quantity has exponential$(s-2t)$ distribution, and so $EB^t(s) \le \frac{1}{s-2t}$, establishing (63).

Remark. A more elaborate argument (Vlada Limic, personal communication) shows $EB^t(u) \sim \frac{1}{2u}$ and $2uB^t(u) \stackrel{d}{\to}$ exponential(1) as $u \to \infty$.

5.2 The excursion length measure

Associated with B^t is an (inhomogeneous) excursion law, analogous to the Itô excursion law for Brownian motion. In particular there is a sigma-finite "excursion length measure" $\rho^t_v(\cdot)$, whose most intuitive interpretation is as follows. Write $H^t_v = \min\{u > 0 : B^t(v+u) = 0\}$. Then

$$\lim_{b\downarrow 0} b^{-1} P(H^t_v > s \mid B^t(v) = b) = \rho^t_v(s,\infty). \tag{64}$$

Clearly $\rho^t_v = \rho^{t-v}_0$, so it suffices to consider ρ^t_0, i.e. excursions starting at $v = 0$. Write $\rho^t_0(\cdot; l)$ Recall the "marks" process defined at (3). for the measure ρ^t_0 restricted to excursions with exactly l marks. Lemma 26 below gives formulas for the densities of these measures in terms of Brownian excursion W^* of length 1. Write

$$\begin{aligned} I &= \int_0^1 W^*(u)du \\ a_l &= EI^l,\ l \ge 0 \\ \Phi(\theta) &= E\exp(\theta I),\ \theta \ge 0 \\ F_t(s) &= \frac{1}{6}((s-t)^3 + t^3). \end{aligned} \tag{65}$$

Lemma 26

$$\frac{d\rho^t_0}{ds}(s) = (2\pi)^{-1/2} s^{-3/2} e^{-F_t(s)} \Phi(s^{3/2}) \tag{66}$$

$$\frac{d\rho^t_0}{ds}(s;l) = (2\pi)^{-1/2} e^{-F_t(s)} s^{3(l-1)/2} a_l/l! \tag{67}$$

Proof. The Cameron-Martin-Girsanov formula ([26] IV.38.5) says that the density of W^t with respect to W, on the set of paths $(W(u), 0 \leq u \leq s)$, is

$$\exp\left(\int_0^s \gamma(u)dW(u) \ - \tfrac{1}{2}\int_0^s \gamma^2(u)du\right)$$

where $\gamma(u) = t - u$ is the drift. On the set of excursions of length s, we have $\frac{1}{2}\int_0^s \gamma^2(u)du = F_t(s)$ and $\int_0^s \gamma(u)dW(u) = \int_0^s W(u)du = I_s$, say, and so the density becomes

$$e^{-F_t(s)}\exp(I_s).$$

Moreover, conditional on the excursion $(W(u), 0 \leq u \leq s)$, the number of marks during the excursion has Poisson(I_s) distribution, and so the corresponding density on paths with exactly l marks is

$$e^{-F_t(s)}(I_s)^l/l!.$$

For Brownian motion itself, the excursion length has density $(2\pi)^{-1/2}s^{-3/2}$, and so

$$\frac{d\rho_0^t}{ds}(s) = (2\pi)^{-1/2}s^{-3/2}e^{-F_t(s)}E\exp\left(\int_0^s \tilde{W}(u)\ du\right)$$

$$\frac{d\rho_0^t}{ds}(s;l) = (2\pi)^{-1/2}s^{-3/2}e^{-F_t(s)}E\left(\int_0^s \tilde{W}(u)\ du\right)^l/l!$$

where $\tilde{W}$ is Brownian excursion of length s. Now (66,67) follow from the Brownian scaling property $\int_0^s \tilde{W}(u)du \ \stackrel{d}{=}\ s^{3/2}\int_0^1 W^*(u)du$.

5.3 The size-biased property

In the random graph $\mathcal{G}(n, n^{-1} + tn^{-4/3})$ we have an obvious "size-biasing" relationship between the size $C_n[1]$ of the component containing vertex 1 and the mean number ($m_n(c)$, say) of components of size c:

$$P(C_n[1] = c) = \tfrac{c}{n}m_n(c). \tag{68}$$

There must be some analogous identity in the Brownian world, and here it is. Write ψ^t for the mean occupation measure for excursion-lengths of B^t:

$$\psi^t(\cdot) = E|\{\gamma \in \Gamma^t : |\gamma| \in \cdot\}|.$$

Proposition 27 $\frac{d\psi^t}{d\rho_0^t}(s) = s^{-1},\ 0 < s < \infty.$

Though it is natural to seek to prove this from (68) and a weak convergence argument, we do not see any simple such argument. Rather than start novel weak convergence arguments, we shall combine stochastic calculus arguments with the size-biased-order result for excursions of B^t, Corollary 16.

Proof of Proposition 27. Write $L(u)$ for local time at 0 for B^t. Constructing B^t from W^t as in (2) we may define

$$L(u) = -\min_{0 \le s \le u} W^t(s)$$

(here and below we omit explicit dependence on t). Write $Q(\delta, s) =$ number of excursions of B^t with length $> \delta$ which begin before s. Then

$$\frac{Q(\delta, s)}{n(\delta)} \to L(s) \text{ a.s. as } \delta \downarrow 0 \tag{69}$$

where $n(\delta) = \int_\delta^\infty (2\pi)^{-1/2} x^{-3/2}\, dx = (2/\pi)^{1/2}\, \delta^{-1/2}$. This is standard for Brownian motion ([27] equation II.37.9) and extends to B^t by absolute continuity. Now consider the total number $M(\delta)$ of excursions of B^t with length $> \delta$. For reflecting Brownian motion with constant drift $-u$, the mean intensity per unit time of excursions of length l is $u(2\pi)^{-1/2} l^{-3/2} \exp(-u^2 l/2)$. Routine but tedious arguments, based on the analog of (69) in the constant drift case and comparison arguments, verify that

$$\frac{M(\delta)}{m(\delta)} \to 1 \text{ a.s. as } \delta \downarrow 0 \tag{70}$$

where

$$m(\delta) = \int_0^\infty \int_\delta^\infty u(2\pi)^{-1/2} l^{-3/2} \exp(-u^2 l/2)\, dl\, du = \int_\delta^\infty (2\pi)^{-1/2} l^{-5/2}\, dl.$$

Now consider the set $\{|\gamma| : \gamma \in \Gamma^t\}$ of all excursion-lengths of B^t. Apply the size-biasing construction of section 3.3, that is introduce (ξ_γ) such that the conditional distribution of ξ_γ given $|\gamma|$ is exponential($|\gamma|$), and set

$$S(x) = \sum_{\gamma : \xi_\gamma < x} |\gamma|.$$

From (70) and the definition of the ξ's, it is not hard to show that as $\delta \downarrow 0$

$$|\{\gamma :\ |\gamma| > \delta, \xi_\gamma < x\}| \sim \int_\delta^\infty (1 - \exp(-xl))(2\pi)^{-1/2} l^{-5/2}\, dl \sim x n(\delta) \text{ a.s.} \tag{71}$$

Now $(L^{-1}(x), 0 \le x < \infty)$ is a pure-jump process whose jump-sizes are the excursion-lengths of B^t, in the order that excursions occur. The size-biasing property in Corollary 16 is that $(S(x); 0 \le x < \infty)$ is also a pure-jump process whose jump-sizes are the excursion-lengths of B^t, in the order that excursions occur. So these two processes are random time-changes of each other, i.e. $(L^{-1}(x), 0 \le x < \infty) \stackrel{d}{=} (S(\Theta(x)); 0 \le x < \infty)$ for some random increasing continuous function Θ. But now (69) and (71) identify $\Theta(x) = x$, and so

$$(L^{-1}(x), 0 \le x < \infty) \quad \stackrel{d}{=} \quad (S(x); 0 \le x < \infty). \tag{72}$$

By definition of the ξ's, as $x \downarrow 0$

$$x^{-1} E|\{\gamma : \xi_\gamma < x, a < |\gamma| < b\}| \to \int_a^b y \psi^t(dy)$$

while the interpretation of the excursion-length measure ρ_0^t as the rate of excursions with respect to local time implies that, as $x \downarrow 0$

$$x^{-1} E|\{\gamma : l(\gamma) < L^{-1}(x), a < |\gamma| < b\}| \ \to \rho_0^t(a, b).$$

But the left sides are equal by (72), and the equality of the right sides is the assertion of the Proposition. $\square$

In an attempt to downplay the abstract aspects of excursion theory, we haven't said what's really going on in terms of the space U (in the notation of [26] section VI.47) of excursion functions f. Our measure ρ_0^t on excursion lengths is induced from a certain measure $\tilde{\rho}_0^t$ on U, whose density with respect to the Itô measure on U can be obtained from the argument for Lemma 26. Writing $\tilde{\psi}^t$ for the mean occupation measure on U for excursions of B^t, the proof of Proposition 27 shows

$$\frac{d\tilde{\psi}^t}{d\tilde{\rho}_0^t}(f) = \frac{1}{|f|}$$

where $|f|$ is the length of excursion f. Since the distribution of the number of marks during an excursion f depends only on f (i.e. not on the starting-time of the excursion) we may deduce the analogous size-biasing relationship for the joint distribution of lengths and number-of-marks in excursions, which we state as follows. Let $\rho_0^t(\cdot, m)$ be as in Lemma 26, and let $\psi^t(\cdot, m)$ be the mean number of excursions of B^t with $|\gamma| \in \cdot$ and with exactly m marks.

Corollary 28

$$\frac{d\psi^t(\cdot, m)}{d\rho_0^t(\cdot, m)}(s) = s^{-1}.$$

As an application, write

$$M^t(l) = \text{ number of excursions of } B^t \text{ with } l \text{ marks}$$

$$Q^t(l) = \text{ total length of excursions of } B^t \text{ with } l \text{ marks .}$$

Then

$$EM^t(l) = \int_0^\infty \psi^t(ds,l) = \int_0^\infty s^{-1}\rho_0^t(ds,l) < \infty,\ l \geq 2 \qquad (73)$$

$$EQ^t(l) = \int_0^\infty s\psi^t(ds,l) = \int_0^\infty \rho_0^t(ds,l) < \infty,\ l \geq 1 \qquad (74)$$

where in each case the first equality is by definition of ψ^t, and the second by Corollary 28. See section 6.2 for discussion of the "random graph asymptotics" interpretation.

5.4 $t \to -\infty$ asymptotics

Fix $t < 0$. For Brownian motion with constant drift t, the excursion length measure ν^t analogous to (64) is

$$\nu^t(ds) = (2\pi)^{-1/2} s^{-3/2} \exp(-t^2 s/2)\ ds.$$

By coupling with B^t, it is easy to see that ρ_0^t is stochastically smaller than ν^t, that is

$$\rho_0^t(x,\infty) \leq \nu^t(x,\infty),\ x > 0. \qquad (75)$$

On the other hand, in formula (66) we have $\Phi(s^{3/2}) \geq 1$ and so

$$\rho_0^t(ds) \geq (2\pi)^{-1/2} s^{-3/2} e^{-F_t(s)}\ ds = \nu^t(ds)\ \exp\left(\tfrac{s^2 t}{2} - \tfrac{s^3}{6}\right). \qquad (76)$$

Consider the sum of r'th powers of excursion lengths of B^t:

$$S_r(t) = \sum_{\gamma \in \Gamma^t} |\gamma|^r,\ r \geq 2.$$

Write $\psi^t(\cdot)$ for the mean number of excursions γ of B^t with $|\gamma| \in \cdot$. Then

$$\begin{aligned} ES_r(t) &= \int_0^\infty s^r \psi^t(ds) \\ &= \int_0^\infty s^{r-1}\rho_0^t(ds) \text{ by Corollary 28} \\ &\sim \int_0^\infty s^{r-1}\nu^t(ds) \text{ as } t \to -\infty \text{ using (75,76)} \\ &= |t|^{3-2r}(2r-5)!! \text{ after a brief calculation} \qquad (77) \end{aligned}$$

where for m odd, $m!! = m(m-2)(m-4)\dots 1$ and $(-1)!! = 1$. The lower bound in the $\sim$ is justified by changing variables in (76) to $du = t^2 ds$ and applying dominated convergence.

Finally, we can bound the maximal excursion length $X_*(t) \equiv \max_{\gamma \in \Gamma^t} |\gamma|$ by a similar argument, as follows.

$$\begin{aligned}
P(X_*(t) > s) &\leq E|\{\gamma : |\gamma| > s\}| \\
&= \int_s^\infty \psi^t(du) \\
&= \int_s^\infty u^{-1} \rho_0^t(du) \text{ by Corollary 28} \\
&\leq s^{-1} \rho_0^t(s, \infty) \\
&\leq s^{-1} \nu^t(s, \infty) \text{ by (75)} \\
&\leq (2\pi)^{-1/2} s^{-5/2} \int_s^\infty \exp(-t^2 u/2)\, du \\
&= (2\pi)^{-1/2} s^{-5/2} \tfrac{2}{t^2} \exp(-t^2 s/2).
\end{aligned}$$

A crude consequence is

$$P(X_*(t) > |t|^{-2+\varepsilon}) = o(\exp(-|t|^{\varepsilon/2})) \text{ as } t \to -\infty. \tag{78}$$

Remark. From (77) with $r = 2$ we have $|t| E S_2(t) \to 1$ as $t \to -\infty$, and it is not hard to improve this to $|t| S_2(t) \xrightarrow{p} 1$. Analysis of the standard multiplicative coalescent in [4] gives the stronger result

$$t + \frac{1}{S_2(t)} \xrightarrow{p} 0 \text{ as } t \to -\infty$$

but this seems hard to deduce from the definition of $S_2(t)$ in terms of B^t.

6 Further discussion

6.1 Methodological discussion

Of course the point of this paper is to exhibit the connection between critical random graphs and Brownian-type processes, a connection not visible in the voluminous literature on either subject. Whether this connection will reveal anything essentially new about distributional asymptotics of random graphs is uncertain. The underlying methodology – what the author terms the *weak convergence paradigm* – is to separate the issue of convergence to some well-defined limit process from the issue of doing explicit calculations, so that one can seek to do the calculations in the continuous world. In section 6.2 we outline how the calculations we have done with excursions of B^t relate to random graph asymptotics. Re-deriving certain other known results, e.g. (80), provides an interesting challenge for stochastic calculus. Weak convergence arguments typically have some robustness under changes in model, as our proof of Proposition 4 shows: it is less clear whether the generating function arguments employed in the random graphs literature can so naturally be extended to the setting of Proposition 4. Of course the weak convergence approach has countervailing disadvantages: Theorem 3 is tied to $p(n) = n^{-1} + tn^{-4/3}$ rather than a wider range of $p(n)$; one loses error bounds in n; and imposing an artificial "time" structure obscures symmetry properties (e.g. size-biasing, section 5.3) of the underlying random graph.

This paper continues a line of work which identifies limits in probabilistic combinatorics with Brownian-type processes (random trees with Brownian excursion [1, 2]; random mappings with Brownian bridge [5]). But these previous results (based on "depth-first search") are not used here, and indeed the weak convergence arguments in this paper are technically rather easier. Pitman [23] discusses a similar kind of problem – numbers of excursions containing j marks – for a different process (Brownian and Bessel processes and bridges, with a rate-1 process of marks).

6.2 Formulas for random graph asymptotics

Carrying through the weak convergence program of obtaining formulas for random graph asymptotics from Theorem 3 requires verifications of technical side conditions, and this is not the place to start such technicalities. But let us give one example of how our stochastic calculus formulas match those in the random graphs literature. Write $M^t_n(l), Q^t_n(l)$ for the number and the total size of components of $\mathcal{G}(n, n^{-1} + tn^{-4/3})$ with surplus l. Then Theorem 3, and verification of further tightness and integrability conditions would imply

$$M^t_n(l) \ \stackrel{d}{\to}\ M^t(l),\ EM^t_n(l) \to EM^t(l)$$

$$n^{-2/3}Q^t_n(l) \ \stackrel{d}{\to}\ Q^t(l),\ n^{-2/3}EQ^t_n(l) \to EQ^t(l)$$

where $M^t(l)$ and $Q_t(l)$ were defined above (73). Lemmas 2.2 and 2.3 of Luczak et al. [19] contain formulas for the asymptotics of $EM_n(l)$ and $n^{-2/3}EQ_n(l)$. These agree with our stochastic calculus formulas (73,74), except that our constant a_l defined by (65) is replaced by $l!\gamma_l$, where γ_l is defined by

$$\gamma_l = \lim_{k\to\infty} C(k, k+l-1)/k^{k-2+3l/2} \qquad (79)$$

where $C(k, k+l-1)$ is the number of connected graphs with k labeled vertices and $k+l-1$ edges. The implicit identity $a_l = l!\gamma_l$ touches upon a large set of ideas: see section 6.5(f).

6.3 Epidemic models

The number of vertices at successive heights in a random component evolve essentially as a classical *epidemic* model. Martin-Löf [22] proves a result in the epidemic setting which roughly translates to the fact that rescaled breadth-first walk in the first component, conditioned on reaching height $\delta > 0$, subsequently evolves as B^t until returning to zero. In principle one could seek to prove Theorem 3 by "stringing together" this fact for successive components, but this is technically complicated due to the σ-finite measure on excursion lengths. Carrying through the weak convergence argument by first showing $\bar{Z}_n^t \xrightarrow{d} W^t$ and then defining the reflecting process B^t via (2) avoids these technical complications.

6.4 General stochastic coalescent processes

The multiplicative coalescent may be viewed as the special case $K(x, y) = xy$ of the "general rate" coalescent scheme with dynamics

for each pair of clusters of sizes (x, y), they merge at rate $K(x, y)$

into a cluster of size $x + y$.

It turns out there is an extensive scientific literature on coalescence, mostly for integer-valued cluster sizes. A survey will be given elsewhere [3], but here are some highlights. Work through the 1960s emphasized the deterministic first-order approximation (*Smoluchowski coagulation equation*) in which the concentrations $c_j(t)$ of size-j clusters are assumed to satisfy the differential equations

$$\tfrac{d}{dt}c_j(t) = \tfrac{1}{2}\sum_{i=1}^{j-1} r(i, j-i)c_i(t)c_{j-i}(t) \ - c_j(t)\sum_{i=1}^{\infty} r(j, i)c_i(t),\ j = 1, 2, \ldots.$$

The survey by Drake [10] has 250 references. The general stochastic model was introduced by Marcus [21] and studied by Lushnikov [20] as a model for gelation. Ernst and van Dongen [31, 30] and references therein indicate subsequent work from a statistical physics viewpoint. The case $K(x, y)$ is essentially the classic random graphs process, and the case $K(x, y) = 1$ is essentially Kingman's coalescent [18, 9], but other cases have not been considered from a rigorous viewpoint until very recently. Evans and Pitman [14] discuss foundational issues and the Feller property (in l_1, rather than the l_2 setting of this paper) for general rate kernels K, and study the additive case $K(x, y) = x + y$.

6.5 Open problems

We collect some problems explicitly or implicitly mentioned, plus some further problems.

(a) The growth of percolation clusters in the usual bond-percolation model on the d-dimensional lattice, near the critical point, is loosely analogous to (but much harder to analyze than) the near-critical behavior of the random graph process. See Borg et al [8] for the latest results. It is unclear whether any continuous-space limit process, analogous to the multiplicative coalescent, might be anticipated in that context.

(b) Find exact necessary-and-sufficient conditions in Proposition 4. It is clear that by truncating away a few large cluster sizes one can weaken (10), but we have not tried to discover how far this idea can be pushed.

(c) Is there an explicit construction of the entire standard multiplicative coalescent in terms of familiar stochastic processes, avoiding any weak convergence argument? In other words, our results imply that there exists a two-parameter process $(B^t(s), 0 \leq s < \infty, -\infty < t < \infty)$ which for fixed t is distributed as B^t and whose excursion lengths evolve as the multiplicative coalescent. But we do not know how to directly define such a process. Janson [15] describes the two-parameter point process giving the times t and component sizes where multicyclic components arise; we would like Janson's process to be included in a two-parameter process description. Using (1,2) with the same W for each t definitely does not work: we need the excursions to merge in a much more complicated way. It seems intuitive that for a fixed pair $t_1 < t_2$ one *cannot* define a bivariate Markov process $((B^{t_1}(s), B^{t_2}(s)), 0 \leq s < \infty)$ such that the joint distribution of the excursion-lengths is the distribution of $(\mathbf{X}(t_1), \mathbf{X}(t_2))$: specifically, if the construction of Z_n^t is made simultaneously for $t = t_1$ and t_2, then Theorem 3 has a bivariate version which leads to a joint distribution $((B^{t_1}(s), B^{t_2}(s)), 0 \leq s < \infty)$, but the joint process is not Markov because coalescence of clusters between t_1 and t_2 leads to excursions of B^{t_1} becoming embedded within earlier (in terms of s) excursions of B^{t_2}, which is incompatible with the Markov property.

(d) Our proofs of the size-biasing properties of excursions of B^t (Corollary 16 and Proposition 27) rely ultimately on weak convergence from the random graphs setting. Can these be proved directly via stochastic calculus on B^t?

(e) Our results enable many random graphs results to be reformulated as results about B^t: can one find stochastic calculus proofs? For example ([16] Theorem 4): let $M(l)$ be the number of excursions of B^t containing exactly l marks. Then, for $t = 0$,

$$P(M(2) = m, M(l) = 0 \ \forall l > 2) = \left(\frac{5}{18}\right)^m \sqrt{\frac{2}{3}} \, \frac{1}{(2m)!}. \tag{80}$$

(f) Elucidate the implicit identity $a_l = l!\gamma_l$ (recall (79,65) for definitions). Briefly, asymptotics of $C(k, k+l-1)$ have been studied in detail in the combinatorics literature (for references see [7] p. 114, [16] p. 262, [19] p. 735). Distributional properties of I have been studied by probabilists: see Takàcs [29] for references and rederivation of formulas for the moments (a_l). In [29] it is noted that the classical depth-first-search $1-1$ correspondence between walk-excursions and planar trees identifies I as the rescaled limit of the sum-of-heights of all vertices in a random planar tree. But directly identifying (a_l) with $(l!\gamma_l)$ seems more subtle, and we do not see any explanation more simple than the following, which is (roughly speaking) implicit in the argument for Theorem 3. Apply the breadth-first walk construction of this paper to the random connected graph on k vertices and $k+l-1$ edges; the rescaled walk converges to Brownian excursion; the vertices with extra edges asymptotically appear as the process of "marks" analogous to N^t. The convergence of walks holds because the spanning tree induced by breadth-first walk is asymptotically like the uniform random labeled unordered tree, and (analogous to the general results of [2]) one expects a Brownian excursion limit for this and other simply-generated families of random trees (the particular case of planar trees being obvious by another $1-1$ correspondence). Spencer [28] elaborates this argument.

(g) Implicit in section 5.4 is the following idea. For large negative t, the sizes of the longer excursions of B^t, i.e. the asymptotic rescaled sizes of the larger components of $\mathcal{G}(n, n^{-1} + tn^{-4/3})$, are approximately like the positions s of the rightmost points in an inhomogeneous Poisson process of rate $(2\pi)^{-1/2}s^{-5/2}\exp(-t^2s/2)$. This idea is also folklore in random graphs, and indeed can be viewed as a consequence of general facts about the tail of the distribution of total population in a just-subcritical branching process. It would be interesting to use the modern Stein-Chen machinery [6] to obtain explicit bounds on the error in this Poisson process approximation.

(h) From formulas in [22] one may derive complicated expressions for the marginal distribution $B^t(s)$ (Vlada Limic, personal communication). Can one obtain expressions for e.g. the distribution of $\sup_{s\geq 0} B^t(s)$?

Acknowledgements. I thank Joel Spencer for the suggestion to abstract the process of coalescence in the random graph process, Jim Pitman for ongoing discussions about coalescent processes and references to the statistical physics literature, and Vlada Limic and a referee for careful reading of draft versions.

References

[1] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[2] D.J. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993.

[3] D.J. Aldous. Deterministic and stochastic models for coalescence: a review of the mean-field theory for probabilists. Survey, in preparation, 1996.

[4] D.J. Aldous and V. Limic. The entrance boundary of the multiplicative coalescent. In Preparation, 1996.

[5] D.J. Aldous and J. Pitman. Brownian bridge asymptotics for random mappings. *Random Struct. Alg.*, 5:487–512, 1994.

[6] A.D. Barbour, L. Holst, and S. Janson. *Poisson Approximation.* Oxford University Press, 1992.

[7] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[8] C. Borgs, J. Chayes, H. Kesten, and J. Spencer. The birth of the infinite cluster .... In preparation, 1996.

[9] P. Donnelly and P. Joyce. Weak convergence of population genealogical processes to the coalescent with ages. *Ann. Probab.*, 20:322–341, 1992.

[10] R.L. Drake. A general mathematical survey of the coagulation equation. In G.M. Hidy and J.R. Brock, editors, *Topics in Current Aerosol Research (Part 2)*, volume 3 of *International Reviews in Aerosol Physics and Chemistry*, pages 201–376. Pergammon, 1972.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.

[12] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[13] S.N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[14] S.N. Evans and J. Pitman. Construction of Markovian coalescents. In preparation, 1996.

[15] S. Janson. Multicyclic components in a random graph process. *Random Structures and Algorithms*, 4:71–84, 1993.

[16] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures and Algorithms*, 4:233–358, 1993.

[17] O. Kallenberg. *Random Measures*. Akademie-Verlag, Berlin, 1983.

[18] J.F.C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13:235–248, 1982.

[19] T. Łuczak, B. Pittel, and J.C. Wierman. The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.*, 341:721–748, 1994.

[20] A.A. Lushnikov. Coagulation in finite systems. *J. Colloid and Interface Science*, 65:276–285, 1978.

[21] A.H. Marcus. Stochastic coalescence. *Technometrics*, 10:133–143, 1968.

[22] A. Martin-Löf. The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.*, xxx:xxx, xxx.

[23] J. Pitman. Partition structures derived from Brownian motion and stable subordinators. Technical Report 346, Statistics Dept, U.C. Berkeley, 1995.

[24] J.W. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Technical Report 433, Statistics Dept, U.C. Berkeley, 1995.

[25] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer, Berlin-Heidelberg, 1991.

[26] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Itô Calculus*, volume 2. Wiley, 1987.

[27] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Foundations*, volume 1. Wiley, second edition, 1994.

[28] J. Spencer. Enumerating graphs and Brownian motion. Courant Institute, New York, 1996.

[29] L. Takàcs. A Bernouilli excursion and its various applications. *Adv. in Appl. Probab.*, 23:557–585, 1991.

[30] P.G.J. van Dongen. Fluctuations in coagulating systems II. *J. Statist. Phys.*, 49:927–975, 1987.

[31] P.G.J. van Dongen and M.H. Ernst. Fluctuations in coagulating systems. *J. Statist. Phys.*, 49:879–926, 1987.

7 Dump

ρ_0^t related to ρ_0^0: say explicitly?

Start section 5: main tools are coupling, CMG, excursions?

formula for $ES_2(t)$?