SCALING LIMITS OF RANDOM GRAPH MODELS AT CRITICALITY: UNIVERSALITY AND THE BASIN OF ATTRACTION OF THE ERDŐS-RÉNYI RANDOM GRAPH

SHANKAR BHAMIDI[1], NICOLAS BROUTIN[2], SANCHAYAN SEN[3], AND XUAN WANG[4]

ABSTRACT. Over the last few years a wide array of random graph models have been postulated to understand properties of empirically observed networks. Most of these models come with a parameter t (usually related to edge density) and a (model dependent) critical time t_c which specifies when a giant component emerges. There is evidence to support that for a wide class of models, under moment conditions, the nature of this emergence is universal and looks like the classical Erdős-Rényi random graph, in the sense of the critical scaling window and (a) the sizes of the components in this window (all maximal component sizes scaling like $n^{2/3}$) and (b) the structure of components (rescaled by $n^{-1/3}$) converge to random fractals related to the continuum random tree. Till date, (a) has been proven for a number of models using different techniques while (b) has been proven for only two models, the classical Erdős-Rényi random graph and the rank-1 inhomogeneous random graph. The aim of this paper is to develop a general program for proving such results. The program requires three main ingredients: (i) in the critical scaling window, components merge approximately like the multiplicative coalescent (ii) scaling exponents of susceptibility functions are the same as the Erdős-Rényi random graph and (iii) macroscopic averaging of expected distances between random points in the same component in the barely subcritical regime. We show that these apply to a number of fundamental random graph models including the configuration model, inhomogeneous random graphs modulated via a finite kernel and bounded size rules. Thus these models all belong to the domain of attraction of the classical Erdős-Rényi random graph. As a by product we also get results for component sizes at criticality for a general class of inhomogeneous random graphs.

1. INTRODUCTION

Over the last few years, motivated both by questions in fields such as combinatorics and statistical physics as well as the explosion in the amount of data on empirically observed networks, a myriad of random graph models have been proposed. A fundamental question in this general area is understanding connectivity properties of the model, including the time and nature of emergence of the giant component. Writing $[n] = \{1, 2, \ldots, n\}$ for

[1]DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 304 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

[2]INRIA PARIS-ROCQUENCOURT, 78153 LE CHESNAY, FRANCE

[3]DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, THE NETHERLANDS

[4]DEPARTMENT OF MATHEMATICS, 307 RAWLES HALL, INDIANA UNIVERSITY, BLOOMINGTON, INDIANA

E-mail addresses: bhamidi@email.unc.edu, broutin@inria.fr, s.sen@tue.nl, xw45@indiana.edu.

2010 *Mathematics Subject Classification.* Primary: 60C05, 05C80.

Key words and phrases. Multiplicative coalescent, **p**-trees, continuum random tree, critical random graphs, branching processes, inhomogeneous random graphs, configuration model, bounded-size rules.

the vertex set, most of these models have a parameter t (related to the edge density) and a model dependent critical time t_c such that for $t < t_c$ (subcritical regime), there exists no giant component (size of the largest component $|\mathscr{C}_1(t)| = o_P(n)$) while for $t > t_c$ (supercritical regime), the size of the largest component scales like $f(t)n$ with $f(t) > 0$ and model dependent.

The classical example of such a model is the Erdős-Rényi random graph ERRG$(n, t/n)$ where the critical time $t_c(\text{ERRG}) = 1$. Similar phenomenon are observed in a wide class of random graph models. The techniques in analyzing such models in the subcritical and supercritical regime are model specific, quite often relying on branching process approximations of local neighborhoods. Understanding what happens at criticality and the nature of emergence of the giant component as one transitions from the subcritical to the supercritical regime has motivated a large body of work. In particular, it is conjectured both in combinatorics and via simulations in statistical physics [24] that the nature of this emergence is "universal" and under moment assumptions on the average degree, a wide array of models exhibit the same sort of behavior in the critical regime to that of the classical Erdős-Rényi random graph in the sense that for any fixed $i \geqslant 1$, the i-th maximal components $|\mathscr{C}_i|$ scales like $n^{2/3}$ and the components viewed as metric spaces using the graph metric, scale like $n^{1/3}$. Till date, for **component sizes** in the critical regime, this has been proven for a number of models including the rank one random graph [17], the configuration model [44, 54] and bounded size rules [11, 13]. Viewing the maximal components as metric spaces, the only model which has succumbed has been the Erdős-Rényi random graph ERRG$(n, 1/n + \lambda/n^{4/3})$ in [3] where it was shown that rescaling edge lengths in the maximal components $\mathscr{C}_i(\lambda)$ by $n^{-1/3}$, one has

$$\left(\frac{\mathscr{C}_i(\lambda)}{n^{1/3}} : i \geqslant 1\right) \xrightarrow{\text{w}} \textbf{Crit}(\lambda) := (\text{Crit}_i(\lambda) : i \geqslant 1), \tag{1.1}$$

for a sequence of limiting random fractals that are described in more detail in Section 2.

Probability theory is filled with a wide array of invariance principles, for example the central limit theorem (or Donsker's invariance principle) which study the convergence of limit processes to fundamental objects such as Brownian motion under *uniform asymptotic negligibility* conditions. In a similar spirit, we are interested in understanding assumptions required that would ensure that at criticality, maximal components in a random graph model behave similar to the Erdős-Rényi in the large network $n \to \infty$ limit and develop general mathematical principles that would enable one to prove such results.

The aim of this paper is to understand general methodology to prove results such as (1.1). We exhibit proof of concept by showing that the general techniques coupled with model specific analysis to verify the assumptions of the main results enable one to prove continuum scaling limits for three large families of random graph models. As a by-product of the metric scaling results, we also obtain and develop new techniques to study component sizes of inhomogeneous random graphs [21] in the critical regime. Of independent interest, the proof technique requires various estimates about the *barely subcritical regime* for the three models analyzed in depth in this paper. Further as shown in [4], understanding metric structure of maximal components in the critical regime is the first step in analyzing more complicated objects such as the scaling limits of the minimal spanning tree on the giant component in the supercritical regime. The techniques developed in this paper would be the first step in establishing universality for such objects as well.

Organization of the paper: An observation which might at first sight seem tangential is that many interesting random graph models can be viewed as dynamic processes on the space of graphs. We start in Section 1.1 by recalling three major families of random graph models, known results about the percolation phase transition/emergence and describe dynamic constructions of these models. The rest of Section 1 gives an informal description of our main results. In Section 2, we define mathematical constructs including metric space convergence and limiting random metric spaces required to state the results. Section 3 describes the main universality results. Section 4 contains results for the three families of random graph models in Section 1.1. Relevance of the results and connection to existing literature are explored in Section 5. Section 6 contains proofs of the universality results stated in Section 3. Section 7, Section 8 and Section 9 contain proofs for the Inhomogeneous random graph model, the configuration model and Bounded size rules respectively.

1.1. **Models.** We now describe three classes of models with vertex set $[n]$ to which the universality results in this paper enable one to prove continuum scaling limits of maximal components. Suppressing dependence on n, we write $\mathscr{C}_i$ for the i-th maximal component and $\mathscr{C}_i(t)$ if this corresponds to a dynamic random graph process at time t.

1.1.1. *Inhomogeneous random graph* [21]. We start by describing a simpler version of the model deferring the general definition to Section 4.1. Start with a Polish ground space $\mathscr{X}$, probability measure μ on $\mathscr{X}$ and a symmetric non-negative kernel $\kappa : \mathscr{X} \times \mathscr{X} \to \mathbb{R}_+$. The random graph $\mathrm{IRG}_n(\kappa, \mu)$ with vertex set $[n]$ is constructed by first generating the types x_i of $i \in [n]$ in an *iid* fashion using μ and then forming the graph by connecting $i, j \in [n]$ with probability $p_{ij} := \min\left\{1, \frac{\kappa(x_i, x_j)}{n}\right\}$ independent across edges. In our regime, this model is asymptotically equivalent to the model with connection probabilities

$$p_{ij} = 1 - \exp\left(-\frac{\kappa(x_i, x_j)}{n}\right) \tag{1.2}$$

We use this version for the rest of the paper. For technical reasons, we will mainly restrict ourselves to the finite type case where $\mathscr{X} = \{1, 2, \ldots, K\}$ for some $K \geqslant 1$. Thus here the kernel κ is just a $K \times K$ symmetric matrix and $\mu = (\mu(i) : i \in [K])$ is a probability mass function. For simplicity and to avoid pesky irreducibility issues, we assume $\kappa(i, j) > 0$ and $\mu(i) > 0$ for all $i, j \in [K]$. View κ as an operator on $L^2([K], \pi, \mu)$ via the action

$$(T_\kappa f)(i) := \sum_{j=1}^{K} \mu(j)\kappa(i, j) f(j), \qquad i \in [K].$$

Then [21] shows that if the operator norm $||\kappa|| < 1$, $\mathscr{C}_1 = o(\log n)$, while if $||\kappa|| > 1$, then $\mathscr{C}_1 \sim f(\kappa)n$, where $f(\kappa)$ is the survival probability of an associated supercritical multitype branching process. We work in the critical regime $||\kappa|| = 1$. Consider the following dynamic version of the above model: For each unordered pair of distinct vertices $u, v \in [n]$, generate *iid* rate one exponential random variables ξ_{uv}. For fixed $\lambda \in \mathbb{R}$, form the graph $\mathscr{G}_n^{\mathrm{IRG}}(\lambda)$ as follows. Connect vertices u,v with an edge if

$$\xi_{ij} \leqslant \left(1+\frac{\lambda}{n^{1/3}}\right)\frac{\kappa(x_u,x_v)}{n},$$

where $x_u, x_v \in [K]$ denote the types of the two vertices. Then by construction, for each fixed $\lambda \in \mathbb{R}$, $\mathscr{G}_n^{\mathrm{IRG}}(\lambda)$ is a random graph where we form edges between vertices independently with probability

$$p_{uv}(\lambda) := 1-\exp\left(-\left(1+\frac{\lambda}{n^{1/3}}\right)\frac{\kappa(x_u,x_v)}{n}\right). \tag{1.3}$$

More importantly, the entire process $\left\{\mathscr{G}_n^{\mathrm{IRG}}(\lambda) : -\infty < \lambda < \infty\right\}$ is a dynamic random graph process that "increases" in the sense of addition of edges as λ increases.

1.1.2. **Configuration model** [9, 19, 49]**:** Start with a prescribed degree sequence $\mathbf{d}_n = (d_i : i \in [n])$, where $d_i \in \mathbb{N}$ represents the degree of vertex i with $\sum_{i=1}^n d_i$ assumed even. Think of each vertex $i \in [n]$ as having d_i half edges associated with it. Form the random graph $\mathrm{CM}_n(\mathbf{d}_n)$ by performing a uniform matching of these half edges (thus two half edges form a complete edge). Special cases include:

(a) **Random r-regular graph:** Fix $r \geqslant 3$ and let $d_i = r$ for all $i \in [n]$.
(b) **Uniform random graph:** Conditioned on having no self-loops or multiple edges the resulting random graph has the same distribution as a uniform random graph amongst all simple graphs with degree sequence $\mathbf{d}_n$.

Assume regularity conditions on the degree sequence $\mathbf{d}_n$ as $n \to \infty$ in particular,

$$\nu_n := \frac{\sum_{i=1}^n d_i(d_i-1)}{\sum_{i=1}^n d_i} \to \nu > 1.$$

Then by [49], there exists a unique giant component of size $\mathscr{C}_1 \sim f(\nu)n$.

Consider the following dynamic construction of $\mathrm{CM}_n(\mathbf{d}_n)$. Assign every half-edge an exponential rate one clock. At time $t = 0$, all half edges are designated Alive. When a clock of an alive half-edge rings this half-edge selects another **alive** half-edge and forms a full edge. Both edges are then considered **dead** and removed from the collection of alive edges. We will use **free** interchangeably with alive in the above construction. This construction is related but not identical to the dynamic construction in [41]. Let $\{\mathrm{CM}_n(t) : t \geqslant 0\}$ denote this dynamic random graph process. Write $\mathrm{CM}_n(\infty) := \mathrm{CM}_n(\mathbf{d}_n)$ for the graph constructed in this fashion.

1.1.3. **Bounded size rules** [56]**:** These models were motivated in trying to understand the effect of limited choice coupled with randomness in the evolution of the network. Fix $K \geqslant 0$ and let $\Omega_K = \{1, 2, \ldots, K, \omega\}$. Now fix a subset $F \subseteq \Omega_K^4$. The process constructs a dynamically evolving graph process $\{\mathscr{G}_n(t) : t \geqslant 0\}$ as follows. For every ordered quadruple of vertices $\mathbf{v} = (v_1, v_2, v_3, v_4)$, let $\mathscr{P}_{\mathbf{v}}$ be a rate $1/2n^3$ Poisson process independent across quadruples. Start with the empty graph $\mathscr{G}_n(0) = \mathbf{0}_n$ on n vertices. For $t > 0$ and $\Delta\mathscr{P}_{\mathbf{v}}(t) = \mathscr{P}_{\mathbf{v}}(t+dt) - \mathscr{P}_{\mathbf{v}} = 1$, let $\mathbf{c}(t) = (c(v_1), c(v_2), c(v_3), c(v_4)) \in \Omega_K^4$, where $c(v_i) = |\mathscr{C}(v_i,t)|$ if the component of v_i has size $|\mathscr{C}(v_i,t)| \leqslant K$ else $c(v_i) = \omega$. Now if $\mathbf{c} \in F$ then add edge (v_1, v_2) to $\mathscr{G}_n(t)$, else add edge (v_3, v_4). Spencer and Wormald [56] showed that such rules exhibit a (rule dependent) critical time $t_c(F)$ such that for $t < t_c$, $|\mathscr{C}_1(t)| = O_P(\log n)$ while for $t > t_c$,

$|\mathscr{C}_1(t)| = \Theta_P(n)$. Note that in this construction multiple edges and self-loops are possible but have negligible effect in the connectivity structure of the maximal components in the critical regime.

To reduce notational overhead, we will state the main results for one famous class of these rules called the Bohman-Frieze process [18] $\{\mathrm{BF}_n(t) : t \geqslant 0\}$ which corresponds to $K = 1$ and the rule $F = \{(1,1,\star,\star)\}$ where $\star$ corresponds to any element in $\Omega_1 := \{1,\omega\}$. In words this means if at any time t the clock corresponding to a particular quadrupule $\mathbf{v} = (v_1, v_2, v_3, v_4)$ rings, then if v_1 and v_2 are singletons then we place an edge between them, else we place an edge between v_3 and v_4 irrespective of the size of the components of these two vertices.

1.2. **Our findings in words.** Let us now give an informal description of our results. Suppose one wanted to understand both maximal component size and metric structure in the critical scaling window of a dynamic random graph process $\{\mathscr{G}_n(t) : t \geqslant 0\}$ and in particular show that they belong to the same universality class as the Erdős-Rényi model. For $t \geqslant 0$ define the second and third susceptibility functions as

$$\bar{s}_2(t) = \frac{1}{n}\sum_i [\mathscr{C}_n^{(i)}(t)]^2, \qquad \bar{s}_3(t) = \frac{1}{n}\sum_i [\mathscr{C}_n^{(i)}(t)]^3. \tag{1.4}$$

Since for $t > t_c$ one has a giant component one expects $s_2(t), s_3(t) \to \infty$ as $t \uparrow t_c$. Note that we are interested in time scales of the form $t = t_c + \lambda/n^{1/3}$ for $\lambda \in \mathbb{R}$, namely in the *critical scaling window*. Fix $\delta < 1/3$ and let $t_n = t_c - n^{-\delta}$. We call the graph at time t_n, the *barely subcritical* regime. Now suppose for a model one can show the following:

(a) The scaling exponents of the susceptibility functions are the same as the Erdős-Rényi random graph. In particular, the second and third susceptibility functions scale like $s_2(t) \sim \alpha/(t - t_c)$ and $s_3(t) \sim \beta(s_2(t))^3$ as $t \uparrow t_c$ for model dependent constants $\alpha, \beta > 0$. Further a new susceptibility function called distanced based susceptibility (see e.g. Theorem 4.10) which measures macroscopic averaging of distances in components scales like $\mathscr{D}(t) \sim \gamma/(t_c - t)^2$.
(b) After the barely subcritical regime, in the time window $[t_n, t_c + \lambda/n^{1/3}]$ components merge approximately like the multiplicative coalescent, namely the Markov process that describes merging dynamics in the Erdős-Rényi random graph (see Section 2.4). The key conceptual point here is that whilst the dynamic random graph process does **not** behave like the multiplicative coalescent, for many models one might expect that in the time window $[t_n, t_c + \lambda/n^{1/3}]$, if δ is not "much" smaller than 1/3 then merging dynamics can be coupled with a multiplicative coalescent thorough some specific functionals of the component (which need not be the component sizes as we will see below).
(c) One has good bounds for maximal component sizes and diameter at time t_n. This is similar to uniform asymptotic negligibility conditions for the CLT.

Then in this paper, we show the following:

> The above three conditions are essentially enough to show that maximal components in the critical scaling window viewed as metric spaces with

each edge rescaled by $n^{-1/3}$ converge to random fractals that are identical to the scaling limits of the Erdős-Rényi random graph [3]. In particular one can show with reasonable amount of work that critical percolation on the configuration model, Inhomogeneous random graphs in the critical regime, as well as bounded size rules in the critical regime all satisfy the above three properties thus resulting in limits of maximal components of the metric structure all through the critical scaling window.

A by product of the analysis is the following at first slightly counter-intuitive finding:

(a) Call the components at time $t_n = t_c - n^{-\delta}$ "blobs" and view each of these as a single vertex. Note that $s_2(t_n) \sim \alpha n^{\delta}$ so the blob of a randomly selected vertex has on the average size n^{δ}. Now fix time $t = t_c + \lambda/n^{1/3}$. Metric structure for maximal components are composed of links between blobs and distances between vertices in the same blob. Our techniques imply that the number of blobs in a component scales like $n^{2/3-\delta}$ (see e.g. Lemma 8.23) which gels nicely with the fact that typical blobs are of size n^{δ} and maximal components scale like $n^{2/3}$. Viewing each blob as a single vertex call the graph composed of links *between the blobs* the "blob-level superstructure". Then one might expect that this scales like $\sqrt{n^{2/3-\delta}} = n^{1/3-\delta/2}$ while distances within a blob which are typically of size n^{δ} scale like $n^{\delta/2}$ thus resulting in the $n^{1/3}$ scaling of the metric structure. This intuitively plausible idea is **incorrect**.

(b) Due to size-biasing of connections with respect to blob size within connected components, owing to connections between the distribution of connected components and tilted versions of famous class of random trees called **p**-trees or birthday trees (Proposition 6.3), the blob-level superstructure in fact scales like $n^{1/3-\delta}$ instead of $n^{1/3-\delta/2}$.

(c) Again due to size-biasing effects, macroscopic averaging of distances over blobs gives a factor of n^{δ} instead of $n^{\delta/2}$. These two effects combined imply that distances scale like $n^{1/3}$ in the maximal components in the critical regime.

2. PRELIMINARIES

This section contains basic constructs required to state our main results.

2.1. **Gromov Hausdorff convergence of metric spaces.** We mainly follow [1, 4, 26]. All metric spaces under consideration will be measured compact metric spaces. Let us recall the Gromov-Hausdorff distance d_{GH} between metric spaces. Fix two metric spaces $X_1 = (X_1, d_1)$ and $X_2 = (X_2, d_2)$. For a subset $C \subseteq X_1 \times X_2$, the distortion of C is defined as

$$\mathrm{dis}(C) := \sup\left\{|d_1(x_1, y_1) - d_2(x_2, y_2)| : (x_1, x_2), (y_1, y_2) \in C\right\}. \tag{2.1}$$

A correspondence C between X_1 and X_2 is a measurable subset of $X_1 \times X_2$ such that for every $x_1 \in X_1$ there exists at least one $x_2 \in X_2$ such that $(x_1, x_2) \in C$ and vice-versa. The Gromov-Hausdorff distance between the two metric spaces (X_1, d_1) and (X_2, d_2) is defined as

$$d_{\mathrm{GH}}(X_1, X_2) = \frac{1}{2}\inf\left\{\mathrm{dis}(C) : C \text{ is a correspondence between } X_1 \text{ and } X_2\right\}. \tag{2.2}$$

We will use the Gromov-Hausdorff-Prokhorov distance that also keeps track of associated measures on the corresponding metric spaces which we now define. A compact measured metric space (X, d, μ) is a compact metric space (X, d) with an associated finite measure μ on the Borel sigma algebra $\mathscr{B}(X)$. Given two compact measured metric spaces (X_1, d_1, μ_1) and (X_2, d_2, μ_2) and a measure π on the product space $X_1 \times X_2$, the discrepancy of π with respect to μ_1 and μ_2 is defined as

$$D(\pi; \mu_1, \mu_2) := ||\mu_1 - \pi_1|| + ||\mu_2 - \pi_2|| \tag{2.3}$$

where π_1, π_2 are the marginals of π and $||\cdot||$ denotes the total variation of signed measures. Then define the metric d_{GHP} between X_1 and X_2 is defined

$$d_{\mathrm{GHP}}(X_1, X_2) := \inf\left\{\max\left(\frac{1}{2}\operatorname{dis}(C),\ D(\pi; \mu_1, \mu_2),\ \pi(C^c)\right)\right\}, \tag{2.4}$$

where the infimum is taken over all correspondences C and measures π on $X_1 \times X_2$.

Write $\mathscr{S}$ for the collection of all measured metric spaces (X, d, μ). The function d_{GHP} is a pseudometric on $\mathscr{S}$, and defines an equivalence relation $X \sim Y \Leftrightarrow d_{\mathrm{GHP}}(X, Y) = 0$ on $\mathscr{S}$. Let $\bar{\mathscr{S}} := \mathscr{S} / \sim$ be the space of isometry equivalent classes of measured compact metric spaces and $\bar{d}_{\mathrm{GHP}}$ be the induced metric. Then by [1], $(\bar{\mathscr{S}}, \bar{d}_{\mathrm{GHP}})$ is a complete separable metric space. To ease notation, we will continue to use $(\mathscr{S}, d_{\mathrm{GHP}})$ instead of $(\bar{\mathscr{S}}, \bar{d}_{\mathrm{GHP}})$ and $X = (X, d, \mu)$ to denote both the metric space and the corresponding equivalence class.

Since we will be interested in not just one metric space but an infinite sequence of metric spaces, the relevant space will be $\mathscr{S}^{\mathbb{N}}$ equipped with the product topology inherited from d_{GHP}.

The scaling operator: We will need to rescale both the metric structure as well as associated measures of the components in the critical regime. Let us setup some notation for this operation. For $\alpha, \beta > 0$, let $\operatorname{scl}(\alpha, \beta)$ be the scaling operator

$$\operatorname{scl}(\alpha, \beta) : \mathscr{S} \to \mathscr{S}, \qquad \operatorname{scl}(\alpha, \beta)[(X, d, \mu)] := (X, d', \mu'),$$

where $d'(x, y) := \alpha d(x, y)$ for all $x, y \in X$, and $\mu'(A) := \beta\mu(A)$ for all $A \subset X$. Write $\operatorname{scl}(\alpha, \beta)X$ for the output of the above scaling operator and $\alpha X := \operatorname{scl}(\alpha, 1)X$.

2.2. **Graph constructs and convergence.** For finite graph $\mathscr{G}$ write $V(\mathscr{G})$ for the vertex set and $E(\mathscr{G})$ for corresponding edge set. As before we write $[n] = \{1, 2, \ldots, n\}$. We will typically denote a connected component of $\mathscr{G}$ by $\mathscr{C} \subseteq \mathscr{G}$. A connected component $\mathscr{C}$, will be viewed as a compact metric space by imposing the usual graph distance d_{G}

$$d_{\mathrm{G}}(v, u) = \text{ number of edges on the shortest path between } v \text{ and } u, \qquad u, v \in \mathscr{C}.$$

Often in the applications below, the graph will come equipped with a collection of vertex weights $\{w_i : i \in [n]\}$ (e.g. the degree sequence in CM_n). There are two natural measures on $\mathscr{G}$

(i) **Counting measure:** $\mu_{\mathrm{ct}}(A) := |A|$, for $A \subset V(\mathscr{G})$.
(ii) **Weighted measure:** $\mu_{\mathbf{w}}(A) := \sum_{v \in A} w_v$, for $A \subset V(\mathscr{G})$. If no weights are specified then by default $w_v \equiv 1$ for all v resulting in $\mu_{\mathbf{w}} = \mu_{\mathrm{ct}}$.

For $\mathscr{G}$ finite and connected, the corresponding metric space is compact with finite measure. We use $\mathscr{G}$ for both the graph and the corresponding measured metric space. Finally for two graphs $\mathscr{G}_1$ and $\mathscr{G}_2$ on the same vertex set say $[n]$, we write $\mathscr{G}_1 \subseteq \mathscr{G}_2$ if $\mathscr{G}_1$ is a subgraph of $\mathscr{G}_2$.

2.3. **Real trees with shortcuts.** Let us now setup notation to describe the limiting random metric space that arose in [3] to describe maximal components at criticality for Erdős-Rényi random graph $\mathscr{G}_n(n, n^{-1} + \lambda n^{-4/3})$ where we start with n vertices and connection probability $p = n^{-1} + \lambda n^{-4/3}$. For $l > 0$, let $\mathscr{E}_l$ for the space of excursions on the interval $[0, l]$. Let $h, g \in \mathscr{E}_l$ be two excursions, fix a countable set $\mathscr{P} \subseteq \mathbb{R}_+ \times \mathbb{R}_+$ with

$$g \cap \mathscr{P} := \left\{(x, y) \in \mathscr{P} : 0 \leqslant x \leqslant l,\ 0 \leqslant y < g(x)\right\} < \infty.$$

The measured metric space $\mathscr{G}(h, g, \mathscr{P})$ is constructed as follows. First, let $\mathscr{T}(h)$ be the real tree associated with the contour function h, see e.g. [33, 47]. Here $\mathscr{T}(h)$ inherits the push forward of the Lebesgue measure from $[0, l]$. Next, $\mathscr{G}(h, g, \mathscr{P})$ is constructed by identifying the pairs of points in $\mathscr{T}(h)$ corresponding to

$$\left\{(x, r(x, y)) \in [0, l] \times [0, l] : (x, y) \in g \cap \mathscr{P},\ r(x, y) = \inf\left\{x' : x' \geqslant x,\ g(x') \leqslant y\right\}\right\}.$$

Thus $\mathscr{G}(h, g, \mathscr{P})$ is constructed by adding a finite number of shortcuts to the real tree $\mathscr{T}(h)$.
Tilted Brownian excursions: Let $\{\mathbf{e}_l(s) : s \in [0, l]\}$ be a Brownian excursion of length l. For $l > 0$ and $\theta > 0$, define the tilted Brownian excursion $\tilde{\mathbf{e}}_l^\theta$ as an $\mathscr{E}_l$-valued random variable such that for all bounded continuous function $f : \mathscr{E}_l \to \mathbb{R}$,

$$\mathbb{E}[f(\tilde{\mathbf{e}}_l^\theta)] = \frac{\mathbb{E}\left[f(\mathbf{e}_l)\exp\left(\theta\int_0^l \mathbf{e}_l(s)ds\right)\right]}{\mathbb{E}\left[\exp\left(\theta\int_0^l \mathbf{e}_l(s)ds\right)\right]}.$$

Note that $\mathbf{e}_l$ and $\tilde{\mathbf{e}}_l^\theta$ are both supported on $\mathscr{E}_l$. Writing ν_l and $\tilde{\nu}_l^\theta$ respectively for the law of $\mathbf{e}_l$ and $\tilde{\mathbf{e}}_l^\theta$ on $\mathscr{E}_l$ the Radon-Nikodym derivative is given by

$$\frac{d\tilde{\nu}_l^\theta}{d\nu_l}(h) = \frac{\exp\left(\theta\int_0^l h(s)ds\right)}{\int_{\mathscr{E}_l}\exp\left(\theta\int_0^l h'(s)ds\right)d\nu_l(dh')}, \qquad h \in \mathscr{E}_l.$$

When $l = 1$, we use $\mathbf{e}(\cdot)$ for the standard Brownian excursion. For fixed $l > 0$ and $\theta = 1$ write $\tilde{\mathbf{e}}_l(\cdot)$ for the corresponding tilted excursion. Now the limiting random metric spaces in all our results can be described as follows. For fixed $\bar{\gamma} > 0$ consider the random compact metric space $\mathscr{G}(2\tilde{\mathbf{e}}_{\bar{\gamma}}, \tilde{\mathbf{e}}_{\bar{\gamma}}, \mathscr{P})$. Here $\tilde{\mathbf{e}}_{\bar{\gamma}}$ is a tilted Brownian excursion of length $\bar{\gamma}$ independent of a rate one Poisson process $\mathscr{P}$ on $\mathbb{R}_+ \times \mathbb{R}_+$.

2.4. **Standard Multiplicative coalescent and the random graph $\mathscr{G}(\mathbf{x}, q)$.** In [5], Aldous constructed the multiplicative coalescent on the space $l^2_\downarrow$

$$l^2_\downarrow = \{(x_1, x_2, \ldots) : x_1 \geqslant x_2 \geqslant \cdots \geqslant 0, \sum_i x_i^2 < \infty\}, \tag{2.5}$$

endowed with the natural metric inherited from l^2. This Markov process described both finite time distributions as well as component merger dynamics of the critical Erdős-Rényi random graph and can be informally described as follows: Fix $\mathbf{x} \in l^2_\downarrow$. Given that at some time t the process is in configuration $\mathbf{X}(t) = \mathbf{x}$, each pair of clusters i and j merge at rate $x_i x_j$ to form a new cluster of size $x_i + x_j$. While this description makes sense for a finite collection of clusters (namely $x_i = 0$ for $i > K$ for some $K < \infty$), Aldous showed that this makes sense in general for $\mathbf{x} \in l^2_\downarrow$ and in fact defines a Feller process on $l^2_\downarrow$. See [5] for an in depth analysis of the construction and properties of the multiplicative coalescent.

A key ingredient in the proof of the existence of such a process is the following random graph $\mathscr{G}(\mathbf{x},q)$. Fix vertex set $[n]$, a collection of positive vertex weights $\{x_i : i \in [n]\}$ and parameter $q > 0$. Construct the random graph $\mathscr{G}(\mathbf{x},q)$ by placing an edge between $i, j \in [n]$ with probability $1-\exp\{-qx_i x_j\}$, independent across edges. Here $q = q^{(n)}$ and $\mathbf{x} = \mathbf{x}^{(n)}$ both depend on n, but we suppress this dependence for notational convenience. For a connected component $\mathscr{C}$ of $\mathscr{G}(\mathbf{x},q)$, define $\mathrm{mass}(\mathscr{C}) = \sum_{i\in\mathscr{C}} x_i$ for the weight of all vertices in the component. Rank components in terms of their mass and let $\mathscr{C}_i$ be the i-th largest component of $\mathscr{G}(\mathbf{x},q)$.

Assumption 2.1. *Let $\sigma_k := \sum_{i\in[n]} x_i^k$ for $k = 2,3$ and $x_{\max} := \max_{i\in[n]} x_i$. Assume there exists a constant $\lambda \in \mathbb{R}$ such that*

$$\frac{\sigma_3}{(\sigma_2)^3} \to 1, \;\; q - \frac{1}{\sigma_2} \to \lambda, \;\; \frac{x_{\max}}{\sigma_2} \to 0,$$

as $n \to \infty$.

Note that the Erdős-Rényi random graph in the critical regime $\mathrm{ERRG}(n, 1/n + \lambda/n^{4/3})$ falls in this class of models with the special choice

$$x_i = n^{-2/3}, \qquad q = n^{1/3} + \lambda.$$

Now fix $\lambda \in \mathbb{R}$. Define the Brownian motion with parabolic drift W_λ and the corresponding process $\tilde{W}_\lambda$,

$$\tilde{W}_\lambda(t) := W_\lambda(t) - \inf_{s\in[0,t]} W_\lambda(s), \;\; W_\lambda(t) := B(t) + \lambda t - \frac{1}{2}t^2, \qquad t \geqslant 0, \tag{2.6}$$

where $\{B(t) : t \geqslant 0\}$ is standard Brownian motion. Aldous [5] proves the following result.

Theorem 2.2 ([5])**.** *Under Assumption* 2.1*,*

$$(\mathrm{mass}(\mathscr{C}_i) : i \geqslant 1) \xrightarrow{\mathrm{w}} \boldsymbol{\xi}(\lambda) := (\gamma_i(\lambda) : i \geqslant 1), \text{ as } n \to \infty,$$

where weak convergence is with respect to the space $l^2_\downarrow$ and $(\gamma_i(\lambda) : i \geqslant 1)$ are the decreasing order of excursion lengths from zero of the process $\{\tilde{W}_\lambda(t) : t \geqslant 0\}$. Further there exists a version of the multiplicative coalescent $\{X(\lambda) : -\infty < \lambda < \infty\}$ called the standard multiplicative coalescent such that for each fixed $\lambda \in \mathbb{R}$, $\mathbf{X}(\lambda) \stackrel{d}{=} \boldsymbol{\xi}(\lambda)$.

2.5. **Scaling limits of components in critical Erdős-Rényi random graph.** Fix $\lambda \in \mathbb{R}$. Now we can define the sequence of limit metric spaces that describe maximal components in the critical regime $\mathrm{ERRG}(n, 1/n + \lambda/n^{4/3})$ as constructed in [3]. Let $\boldsymbol{\xi}(\lambda)$ be as in Theorem 2.2. Conditional on the excursion lengths $\boldsymbol{\xi}(\lambda)$, let $\{\tilde{\mathbf{e}}_{\gamma_i(\lambda)} : i \geqslant 1\}$ be a sequence of independent tilted Brownian excursions with $\tilde{\mathbf{e}}_{\gamma_i(\lambda)}$ having length $\gamma_i(\lambda)$. Let $\{\mathscr{P}_i : i \geqslant 1\}$ be a sequence of independent rate one Poisson processes on $\mathbb{R}^2_+$. Recall the metric space $\mathscr{G}(h, g, \mathscr{P})$ constructed in Section 2.3. Now consider the sequence of random metric spaces

$$\mathrm{Crit}_i(\lambda) := \mathscr{G}(2\tilde{\mathbf{e}}_{\gamma_i(\lambda)}, \tilde{\mathbf{e}}_{\gamma_i(\lambda)}, \mathscr{P}_i), \qquad i \geqslant 1. \tag{2.7}$$

For the rest of the paper we let

$$\mathbf{Crit}(\lambda) := (\mathrm{Crit}_i(\lambda) : i \geqslant 1). \tag{2.8}$$

3. RESULTS: UNIVERSALITY

We start by extending Theorem 2.2 in two stages. Since both are abstract results, let us first give a brief idea of how these will be used in the sequel. In the first stage, we prove under additional assumptions on the weight sequence $\mathbf{x}$, that for each fixed $i \geqslant 1$, the component $\mathscr{C}_i$, properly rescaled, converges to $\mathbf{Crit}(\lambda)$ as in (2.8). For a dynamic random graph model $\{\mathscr{G}_n(t) : t \geqslant 0\}$ we will fix some appropriate $\delta < 1/3$ and study the connectivity patters between components at time t_n that is formed in the interval $[t_n, t_c + \lambda/n^{1/3}]$. With a good deal of effort, we will show that for the models of interest in this paper, in this interval connectivity patterns can be *approximately* described through the graph $\mathscr{G}(\mathbf{x}, q)$. Here each vertex i is in fact a blob (a connected component at time t_n) viewed as a single vertex and x_i is an some functional of the blob (as we will see later this need not be the size of the component, it could be some other functional of the blob, for example the number of still alive edges at time t_n in the configuration model). Thus $\mathscr{G}(\mathbf{x}, q)$ should be thought of as describing the blob-level superstructure, namely the connectivity pattern between these “blobs” owing to edges created in the interval $[t_n, t_c + \lambda/n^{1/3}]$ where each blob is viewed as a single vertex, ignoring all internal structure. Note that here the vertex set $[n]$ for blob-level is a misnomer as in fact we should be using the number of components at time t_n but in order to prevent notation from exploding we will use n to describe the abstract result below.

In the second stage, we replace each vertex $i \in [n]$ in the graph $\mathscr{G}(\mathbf{x}, q)$ with a connected compact measured metric space (M_i, d_i, μ_i) (referred to as “blobs”) and describe how one incorporates blob-blob junction points within the metric spaces, overlayed with the superstructure analyzed in stage one. We show that the metric space now associated with $\mathscr{C}_i$, under natural regularity assumptions on the blobs (Assumption 3.3) converge to the same limit after proper rescaling, owing to macroscopic averaging of within blob distances. Here convergence of metric spaces is under the metric d_{GHP} as in Section 2.1. This second result now gives the convergence of the full metric on the maximal components as we have now taken into account the metric within blobs as well.

3.1. Stage One: The blob-level structure.

Recall the random graph $\mathscr{G}(\mathbf{x}, q)$ in Section 2.4. In addition to Assumption 2.1 we need the following.

Assumption 3.1. *Assume there exist $\eta_0 \in (0, 1/2)$ and $r_0 \in (0, \infty)$ as $n \to \infty$, we have*

$$\frac{x_{\max}}{\sigma_2^{3/2+\eta_0}} \to 0, \quad \frac{\sigma_2^{r_0}}{x_{\min}} \to 0.$$

Theorem 3.2. *Treat $(\mathscr{C}_i : i \geqslant 1)$ as measured metric spaces using the graph distance for the distance between vertices and the weighted measure where each vertex $i \in [n]$ is given weight x_i. Under Assumptions 2.1 and 3.1, we have*

$$\left(\mathrm{scl}(\sigma_2, 1)\mathscr{C}_i^{(n)} : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}(\lambda), \ \text{as } n \to \infty.$$

Here the convergence of each component is with respect to the GHP topology and the joint convergence of the sequences of components is with respect to the product topology.

Remark 1. As described before, the result for critical Erdős-Rényi random graph is recovered as a special case by taking $x_i \equiv n^{-2/3}$ and $q = \lambda + n^{1/3}$. Here $\sigma_2 = n^{-1/3}$. These choices recover the continuum scaling limits of the Erdős-Rényi model in the critical regime proved in [3]. All scaling constants are absorbed into the condition $\sigma_3/\sigma_2^3 \to 1$.

3.2. **Stage Two: Inter-blob distances.** In this section, we will study the case where each vertex in the random graph $\mathscr{G}(\mathbf{x}, q)$ actually corresponds to a metric space. To describe the general setup we will need the following ingredients.

(a) **Blob level superstructure:** This consists of a simple finite graph $\mathscr{G}$ with vertex set $\mathcal{V} := [n]$ and a weight sequence $\mathbf{x} := (x_i : i \in \mathcal{V})$.
(b) **Blobs:** A family of compact connected measured metric spaces $\mathbf{M} := \{(M_i, d_i, \mu_i) : i \in \mathcal{V}\}$, one for each vertex in $\mathscr{G}$. Further assume that for all $i \in \mathcal{V}$, μ_i is a probability measure namely $\mu_i(M_i) = 1$.
(c) **Blob to blob junction points:** This is a collection of points $\mathbf{X} := (X_{i,j} : i \in \mathcal{V}, j \in \mathcal{V})$ such that $X_{i,j} \in M_i$ for all i, j.

Using these three ingredients, we can define a metric space $(\bar{M}, \bar{d}, \bar{\mu}) = \Gamma(\mathscr{G}, \mathbf{x}, \mathbf{M}, \mathbf{X})$ as follows: Let $\bar{M} := \bigsqcup_{i\in[n]} M_i$. Define the measure $\bar{\mu}$ as

$$\bar{\mu}(A) = \sum_{i\in[n]} x_i \mu_i(A \cap M_i), \text{ for } A \subset \bar{M}. \tag{3.1}$$

The metric $\bar{d}$ is the natural metric obtained by using the blob-level distance functions $\{d_i : i \in \mathcal{V}\}$ incorporated with the graph metric $\mathscr{G}$ by putting an edge of length one between the pairs of vertices:

$$\left\{(X_{i,j}, X_{j,i}) : (i,j) \text{ is an edge in } \mathscr{G}\right\}.$$

More precisely, for $x, y \in \bar{M}$ with $x \in M_{j_1}$ and $y \in M_{j_2}$, define

$$\bar{d}(x,y) = \inf_{k\geqslant 1; i_1,\ldots,i_{k-1}} \left\{ k + d_{j_1}(x, X_{j_1,i_1}) + \sum_{\ell=1}^{k-1} d_{i_\ell}(X_{i_\ell,i_{\ell-1}}, X_{i_\ell,i_{\ell+1}}) + d_{j_2}(X_{j_2,i_{k-1}}, y) \right\},$$

where the infimum is taken over all paths $(i_1, \ldots, i_{k-1})$ in $\mathscr{G}$ and we interpret i_0 and i_k as j_1 and j_2 respectively. Note that there is a one-to-one correspondence between components in $\mathscr{G}$ and components in $\Gamma(\mathscr{G}, \mathbf{x}, \mathbf{M}, \mathbf{X})$.

The above gives a deterministic procedure, using the three ingredients above to create a new metric space. We now describe, how these three ingredients are selected in our applications. Assume that we are provided with the weight sequence and the family of metric spaces

$$q^{(n)}, \mathbf{x}^{(n)} := \left\{x_i^{(n)} : i \in [n]\right\}, \mathbf{M}^{(n)} := \left\{(M_i^{(n)}, d_i^{(n)}, \mu_i^{(n)}) : i \in [n]\right\}$$

where as before we will suppress n in the notation. Let $\mathscr{G}(\mathbf{x}, q)$ be the random graph defined in Section 2.4 constructed using the weight sequence $\mathbf{x}$ and the parameter q. Let $\{\mathscr{C}_i : i \geqslant 1\}$ denote the connected components of $\mathscr{G}(\mathbf{x}, q)$ ranked in terms of their masses. For fixed i, let $(X_{i,j} : i \in [m], j \in [m])$ be *iid* random variables (and independent of the graph $\mathscr{G}(\mathbf{x}, q)$) taking values in M_i with distribution μ_i.

Let $\bar{\mathscr{G}}(\mathbf{x}, q, \mathbf{M}) = \Gamma(\mathscr{G}, \mathbf{x}, \mathbf{M}, \mathbf{X})$ be the (random) compact measured metric space constructed as above with the blob-level superstructure taken to be $\mathscr{G} = \mathscr{G}(\mathbf{x}, q)$ and let $\bar{\mathscr{C}}_i$ be the component in $\bar{\mathscr{G}}(\mathbf{x}, q, \mathbf{M})$ that corresponds to the i-th largest component $\mathscr{C}_i$ in $\mathscr{G}(\mathbf{x}, q)$. Thus the law of $(\bar{\mathscr{C}}_i : i \geqslant 1)$ consists of two levels of randomness: the random edges in $\mathscr{G}(\mathbf{x}, q)$ and the random choice of junction points in the blobs $\left\{(M_i, d_i, \mu_i) : i \in [n]\right\}$. Let

$\mathbf{X} = \{X_{i,j} : i, j \in [n]\}$ be the random blob-blob junction points constructed above. For blob M_i let

$$u_i := \mathbb{E}[d_i(X_{i,1}, X_{i,2})], \tag{3.2}$$

denote the first moment of the distance between two randomly chosen points in M_i using the measure μ_i. Let $d_{\max} = \max_{i\in[n]} \operatorname{diam}(M_i)$. We will make the following assumption on the entire collection of weights $\mathbf{x}$ and blob-level distances. Note that the last two assertions are the new assumptions added to Assumption 3.1.

Assumption 3.3. *Suppose there exist $\eta_0 \in (0, 1/2)$ and $r_0 \in (0,\infty)$ such that as $n \to \infty$, we have*

$$\frac{x_{\max}}{\sigma_2^{3/2+\eta_0}} \to 0, \quad \frac{\sigma_2^{r_0}}{x_{\min}} \to 0, \quad \frac{d_{\max}\sigma_2^{3/2-\eta_0}}{\sum_{i=1}^{\infty} x_i^2 u_i + \sigma_2} \to 0, \quad \frac{\sigma_2 x_{\max} d_{\max}}{\sum_{i\in[n]} x_i^2 u_i} \to 0.$$

Then we have the following result.

Theorem 3.4. *Under Assumptions 2.1 and 3.3, we have*

$$\left(\operatorname{scl}\left(\frac{(\sigma_2)^2}{\sigma_2 + \sum_{i\in[n]} x_i^2 u_i}, 1\right)\bar{\mathscr{C}}_i : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}(\lambda), \ \textit{as } n \to \infty,$$

where the convergence of each component is with respect to the GHP topology using the measure $\bar{\mu}$ as in (3.1) on $\bar{\mathscr{G}}(\mathbf{x}, q, \mathbf{M})$.

Remark 2. Depending on whether $\lim_{n\to\infty} \sum_{i\in[n]} x_i^2 u_i / \sigma_2 = 0$, $\in (0,\infty)$ or $= \infty$, the above theorem deals with three different scales. Critical Erdős-Rényi random graph corresponds to the case where each blob is just a single vertex thus with inner blob distance zero. This is an example of the "$= 0$" case where the graph distance inherited from $\mathscr{G}(\mathbf{x}, q)$ dominates the inner vertex distances. In the applications below for general random graphs evolving after the barely subcritical window, we will find $\sigma_2 \sim n^{\delta-1/3}$ while $\sum_{i\in[n]} x_i^2 u_i \sim n^{2\delta-1/3}$ which corresponds to the "$= \infty$" case.

4. Results for associated random graph models

Let us now state our results for the three families of models described in Section 1.1. For each model we describe the continuum scaling limit of maximal components at criticality and then describe the results in the barely subcritical regime which play a key role in the proof and are of independent interest.

4.1. Inhomogeneous random graphs.

Recall the IRG model from Section 1.1.1. We will work in a slightly more general setup where the kernel could depend on n and the types need not be chosen in an iid fashion, rather the empirical distribution of types needs to satisfy regularity conditions in the large network limit. We now describe the precise model. Let $\mathscr{X} = [K] = \{1, 2, \dots, K\}$ be the type space and $\kappa_n(\cdot,\cdot) : [K]\times[K] \to \mathbb{R}^+$ be symmetric kernels for $n \geqslant 1$. Let $\mathcal{V}^{(n)} := [n]$ be the vertex space where each vertex i has a "type", $x_i \in [K]$. The inhomogeneous random graph (IRG) $\mathscr{G}_{\mathrm{IRG}}^{(n)} := \mathscr{G}^{(n)}(\kappa_n, \mathcal{V}^{(n)})$ is a random graph on the vertex set $\mathcal{V}^{(n)}$ constructed as follows. For each $i, j \in \mathcal{V}^{(n)}$, $i \neq j$, with types x_i and x_j respectively, place an edge between i and j with probability

$$p_{ij} = p_{ij}^{(n)} = 1 - \exp\left\{-\frac{\kappa_n(x_i, x_j)}{n}\right\}, \tag{4.1}$$

independent across distinct pairs. Denote the empirical distribution of types by μ_n namely

$$\mu_n(x) := \frac{|\{i \in \mathcal{V}^{(n)} : i \text{ has the type } x\}|}{n}, \text{ for } x \in [K].$$

We will write $\boldsymbol{\mu}_n$ for the vector $(\mu_n(1), \ldots, \mu_n(K))^t$. Consider the associated operator T_{κ_n},

$$(T_{\kappa_n} f)(x) := \sum_{y \in [K]} \kappa_n(x,y) f(y) \mu_n(y), \;\; x \in [K], f \in \mathbb{R}^{[K]}.$$

We will make the following assumptions about $\{(\kappa_n, \mu_n) : n \geqslant 1\}$.

Assumption 4.1. (a) **Convergence of the kernels:** *There exists a kernel* $\kappa(\cdot,\cdot) : [K] \times [K] \to \mathbb{R}$ *and a matrix* $A = \big((a_{xy})\big)_{x,y \in [K]}$ *with real valued (not necessarily positive) entries such that*

$$\min_{x,y \in [K]} \kappa(x,y) > 0 \textit{ and } \lim_n n^{1/3} \big(\kappa_n(x,y) - \kappa(x,y)\big) = a_{xy} \textit{ for } x,y \in [K].$$

(b) **Convergence of the empirical measures:** *There exists a probability measure* μ *on* $[K]$ *and a vector* $\mathbf{b} = (b_1, \ldots, b_K)^t$ *such that*

$$\min_{x \in [K]} \mu(x) > 0 \textit{ and } \lim_n n^{1/3} \big(\mu_n(x) - \mu(x)\big) = b_x \textit{ for } x \in [K].$$

(c) **Criticality of the model***: The operator norm of* T_κ *in* $L^2([K], \mu)$ *equals one.*

Remark 3. The conditions in Assumption 4.1 can be viewed as the critical window for IRG model.

Fix $\delta \in (1/6, 1/5)$ and define the kernel κ_n^- by

$$\kappa_n^-(x,y) := \kappa_n(x,y) - n^{-\delta}, \text{ for } x,y \in [K]. \tag{4.2}$$

We will write κ (resp. κ_n, κ_n^-) to denote the $K \times K$ matrix with entries $\kappa(i,j)$ (resp. $\kappa_n(i,j), \kappa_n(i,j) - n^{-\delta}$) and it will be clear from the context whether the reference is to the kernel or the matrix. Define $\boldsymbol{\mu}$ to be the vector $(\mu(1), \ldots, \mu(K))^t$ and write

$$D = \mathrm{Diag}(\boldsymbol{\mu}), \; D_n = \mathrm{Diag}(\boldsymbol{\mu}_n), \text{ and } B = \mathrm{Diag}(\mathbf{b})$$

where $\mathrm{Diag}(\boldsymbol{\mu})$ denotes the $K \times K$ diagonal matrix whose diagonal entries are $\mu(1), \ldots, \mu(K)$.

For a square matrix $\overline{M}$ with positive entries, we will denote by $\rho(\overline{M})$, its Perron root. Define $m_{ij} = \mu(j)\kappa(i,j)$ for $i,j \in [K]$ and let $M = \big((m_{ij})\big)$. Note that, Assumption 4.1 (c) is equivalent to the condition $\rho(M) = 1$. Let $\mathbf{u}$ and $\mathbf{v}$ be right and left eigenvectors respectively of M corresponding to the eigenvalue $\rho(M) = 1$ subject to $\mathbf{v}^t\mathbf{u} = 1$ and $\mathbf{u}^t\mathbf{1} = 1$, *i.e.,*

$$M\mathbf{u} = \mathbf{u}, \; \mathbf{v}^t M = \mathbf{v}^t, \; \mathbf{u}^t \mathbf{1} = 1 \text{ and } \mathbf{v}^t\mathbf{u} = 1. \tag{4.3}$$

Writing $\mathbf{u} = (u_1, \ldots, u_K)^t$ and $\mathbf{v} = (v_1, \ldots, v_K)^t$, we define

$$\alpha = \frac{1}{(\mathbf{v}^t\mathbf{1}) \cdot (\boldsymbol{\mu}^t\mathbf{u})}, \; \beta = \frac{\sum_{x \in [K]} v_x u_x^2}{(\mathbf{v}^t\mathbf{1}) \cdot (\boldsymbol{\mu}^t\mathbf{u})^2} \text{ and } \zeta = \alpha \cdot \big[\mathbf{v}^t (AD + \kappa B)\mathbf{u}\big]. \tag{4.4}$$

Theorem 4.2. *Let $\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)})$, $i \geq 1$, be the i-th largest component in $\mathscr{G}_{\mathrm{IRG}}^{(n)}$. View each $\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)})$ as a measured metric space by assigning measure* 1 *to each vertex. Under Assumption* 4.1*, we have,*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\alpha n^{1/3}}, \frac{\beta^{1/3}}{n^{2/3}}\right)\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)}) : i \geq 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}\left(\frac{\zeta}{\beta^{2/3}}\right) \text{ as } n \to \infty.$$

As a by product of proving Theorem 4.2, we obtain the following result about the sizes of the components of $\mathscr{G}_{\mathrm{IRG}}^{(n)}$.

Theorem 4.3. *With notation as in Theorem* 4.2*, we have, under Assumption* 4.1*,*

$$\left(\frac{\beta^{1/3}}{n^{2/3}}|\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)})| : i \geq 1\right) \xrightarrow{\mathrm{w}} \boldsymbol{\xi}\left(\frac{\zeta}{\beta^{2/3}}\right) \text{ as } n \to \infty$$

with respect to $l^2_{\downarrow}$ topology where $|\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)})|$ denotes the size of $\mathscr{C}_i(\mathscr{G}_{\mathrm{IRG}}^{(n)})$.

We now define a closely related model $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}(\kappa_n^-, \mathscr{V}^{(n)})$ as follows. For each $i, j \in [n]$ with $i \neq j$, place an edge between them independently with probability

$$p_{ij}^\star = 1 \wedge (\kappa_n^-(x_i, x_j)/n).$$

As before, $x_i, x_j \in [K]$ denote the types of i and j respectively. In order to apply Theorem 3.3, we have to study the barely subcritical random graph $\mathscr{G}_{\mathrm{IRG}}^{(n),-} := \mathscr{G}^{(n)}(\kappa_n^-, \mathscr{V}^{(n)})$ (i.e. the connection probabilities will be given by (4.1) with κ_n^- replacing κ_n). It is more convenient to work with $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}$ when checking the necessary conditions in the barely subcritical regime. These results can then be translated to $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ by asymptotic equivalence [37].

For any graph G with vertex set $\mathscr{V}(G)$, define

$$\mathscr{S}_k(G) := \sum_{\mathscr{C} \subset G} |\mathscr{C}|^k, \ \bar{s}_k(G) := \mathscr{S}_k(G)/|\mathscr{V}(G)| \text{ for } k = 1, 2, \ldots \text{ and}$$
$$\mathscr{D}(G) := \sum_{i,j \in \mathscr{V}(G)} d(i,j) \mathbb{1}_{\{d(i,j) < \infty\}}, \ \bar{\mathscr{D}}(G) = \mathscr{D}(G)/|\mathscr{V}(G)|.$$

We prove the following theorem for the graph $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}$, which plays an important role in proving Theorem 4.2.

Theorem 4.4. *Let $\mathscr{C}_i^\star$ be the i-th largest component in $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}$. Define $\mathscr{D}_{\max}^\star = \max_{i \geq 1} \mathrm{diam}(\mathscr{C}_i^\star)$. Write $\bar{\mathscr{D}}^\star := \bar{\mathscr{D}}(\mathscr{G}_{\mathrm{IRG}}^{(n),\star})$ and $\bar{s}_k^\star := \bar{s}_k(\mathscr{G}_{\mathrm{IRG}}^{(n),\star})$ for $k = 1, 2, \ldots$. Then, there exists a positive constant $A_1 = A_1(\kappa, \mu)$ such that*

$$\frac{\bar{s}_3^\star}{(\bar{s}_2^\star)^3} \xrightarrow{\mathrm{P}} \beta, \ n^{1/3}\left(\frac{1}{n^\delta} - \frac{1}{\bar{s}_2^\star}\right) \xrightarrow{\mathrm{P}} \zeta \text{ and } \mathbb{P}(|\mathscr{C}_1^\star| \geqslant A_1 n^{2\delta} \log n) \to 0 \tag{4.5}$$

Further, there exists a positive constant $A_2 = A_2(\kappa, \mu)$ such that

$$\frac{\bar{\mathscr{D}}^\star}{n^{2\delta}} \xrightarrow{\mathrm{P}} \alpha \text{ and } \mathbb{P}(\mathscr{D}_{\max}^\star \geqslant A_2 n^\delta \log n) \to 0. \tag{4.6}$$

A simple consequence of this theorem is the analogous result for $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$.

Corollary 4.5. *Let $\mathscr{C}_i^-$ be the i-th largest component in $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$. Define $\mathscr{D}_{\max}^- = \max_{i\geq 1}\operatorname{diam}(\mathscr{C}_i^-)$, $\bar{\mathscr{D}} := \bar{\mathscr{D}}(\mathscr{G}_{\mathrm{IRG}}^{(n),-})$ and $\bar{s}_k := \bar{s}_k(\mathscr{G}_{\mathrm{IRG}}^{(n),-})$ for $k = 1,2,\ldots$. Then, the conclusions of Theorem* 4.4 *hold if we replace $\mathscr{C}_i^\star, \mathscr{D}_{\max}^\star, \bar{\mathscr{D}}^\star, \bar{s}_2^\star$ and $\bar{s}_3^\star$ by $\mathscr{C}_i^-, \mathscr{D}_{\max}^-, \bar{\mathscr{D}}, \bar{s}_2$ and $\bar{s}_3$ respectively.*

4.2. **Configuration model and random graphs with prescribed degrees.** Recall the configuration model $\mathrm{CM}_n(\infty)$ constructed via a degree distribution $\mathbf{d}_n$ in Section 1.1.2. We will assume that the degree sequence $\mathbf{d}_n$ is generated in an iid fashion using a probability mass function $\mathbf{p}$. This is not essential and one can make similar assumptions as the IRG model on the rate of convergence of the empirical distribution of degrees; for simplicity for stating the results we assume this iid generation of the degree sequence. Let

$$\mu = \sum_k k p_k, \qquad \nu = \frac{\sum_k k(k-1)p_k}{\mu}, \qquad \beta = \sum_k k(k-1)(k-2)p_k. \tag{4.7}$$

We make the following assumptions on the degree distribution.

Assumption 4.6. *Assume $\nu > 1$, $0 < \beta < \infty$. Further assume that the degree distribution has exponential moments. Namely, there exists $\lambda_0 > 0$ such that $\sum_k \exp(\lambda_0 k) p_k < \infty$.*

Now consider percolation on $\mathrm{CM}_n(\infty)$ where we retain each edge with probability p. Write $\mathrm{Perc}_n(p)$ for the corresponding random graph. It is known [34,36,49] that the critical value for percolation is $p_c = 1/\nu$. Fix $\lambda \in \mathbb{R}$ and let $p = p(\lambda)$ where

$$p(\lambda) = \frac{1}{\nu} + \frac{\lambda}{n^{1/3}}. \tag{4.8}$$

Then it is known that the number of vertices in the maximal components all scale like $n^{2/3}$, see [51] for the random regular graph and [44,54] for the general case. Let $\mathscr{C}_{i,\mathrm{Perc}}$ for the i-th maximal component in $\mathrm{Perc}_n(p(\lambda))$.

Theorem 4.7 (Percolation on the configuration model)**.** *Fix $\lambda \in \mathbb{R}$ and consider percolation on the configuration model with $p(\lambda)$ as in* (8.5)*. View $\mathscr{C}_{i,\mathrm{Perc}}(\lambda)$ as a measured metric space via the graph metric equipped with the counting measure. Then*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu}\frac{1}{n^{1/3}}, \frac{\beta^{1/3}}{\mu n^{2/3}}\right)\mathscr{C}_{i,\mathrm{Perc}}(\lambda) : i \geq 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty\left(\frac{\nu^2}{\beta^{2/3}}\lambda\right), \qquad \text{as } n \to \infty.$$

As a corollary we get the following result for random regular graphs by noting that in this case $\beta = r(r-1)(r-2)$ and $\mu = r, \nu = r-1$.

Corollary 4.8. *Fix $r \geq 3$ and $\lambda \in \mathbb{R}$. Consider percolation on the random r-regular graph with edge probability*

$$p(\lambda) = \frac{1}{r-1} + \frac{\lambda}{n^{1/3}}.$$

Then the maximal components viewed as measured metric spaces where each vertex is assigned mass $(r(r-1)(r-2))^{1/3}/rn^{2/3}$ satisfy

$$\left(\frac{(r(r-1)(r-2))^{2/3}}{r(r-1)}\frac{1}{n^{1/3}}\mathscr{C}_i(\lambda) : i \geq 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty\left(\frac{(r-1)^2}{(r(r-1)(r-2))^{2/3}}\lambda\right), \qquad \text{as } n \to \infty.$$

The above two results follow easily from asymptotics for the dynamic construction $\{\mathrm{CM}_n(t) : t \geqslant 0\}$ of the configuration model in Section 1.1.2 which we now state and the equivalence between percolation on the full configuration model and the dynamic construction [34, 36]. The critical time in the dynamic construction where the system transitions from maximal component $|\mathscr{C}_1(t)| = O_P(\log n)$ to the emergence of a giant component turns out to be

$$t_c = \frac{1}{2}\log\frac{\nu}{\nu-1}. \tag{4.9}$$

Theorem 4.9. *Fix $\lambda \in \mathbb{R}$ and consider the dynamic construction of the configuration model at time $\mathrm{CM}_n(t_c+\lambda/n^{1/3})$. Then the maximal components viewed as measured metric spaces equipped with the graph distance and the counting measure at this time satisfy*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu}\frac{1}{n^{1/3}}, \frac{\beta^{1/3}}{\mu n^{2/3}}\right)\mathscr{C}_i\left(t_c+\frac{\lambda}{n^{1/3}}\right) : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty\left(\frac{2\nu(\nu-1)\mu}{\beta^{2/3}}\lambda\right), \qquad \textit{as } n\to\infty.$$

A key ingredient in the proof of the above Theorem is the analysis of the barely subcritical regime which we now describe. Fix $\delta < 1/3$. The entrance boundary corresponding to the barely subcritical regime turns out to be

$$t_n = t_c - \frac{1}{2}\frac{\nu}{\nu-1}\frac{1}{n^\delta}. \tag{4.10}$$

Here the constant in front of $n^{-\delta}$ is not important and is only useful for simplify constants during the analysis. Recall that $\mathscr{C}_i(t)$ denotes the i-th largest component in $\mathrm{CM}_n(t)$ and note that some of the half-edges of vertices in $\mathscr{C}_i(t)$ might have been used by time t to form full edges and thus are no longer considered free (in fact we call these half-edges dead half-edges). Let $f_i(t)$ denote the total number of still free (sometimes referred to as still alive to emphasize that these have not yet been used at time t) half edges in $\mathscr{C}_i(t)$. We will need to define analogs of the susceptibility functions but in this case the quantities to keep track off turn out to be these still free half-edges. To see why, note that components $\mathscr{C}_i(t)$ and $\mathscr{C}_j(t)$ merge in $[t, t+dt)$ if either one of the still free edges in $\mathscr{C}_i(t)$ rings and it chooses one of the free edges in $\mathscr{C}_j(t)$ or vice-versa, forming a full edge and both of these half-edges then removed from the system. Thus the rate of merger is given by

$$f_i(t)\frac{f_j(t)}{n\bar{s}_1(t)-1} + f_j(t)\frac{f_i(t)}{n\bar{s}_1(t)-1} = 2\frac{f_i(t)f_j(t)}{n\bar{s}_1(t)-1},$$

where $n\bar{s}_1(t) = \sum_i f_i(t)$ is the total number of free edges. Since we will eventually compare this process to a multiplicative coalescent and use Theorem 3.4, we see that the objects to keep track of are the number of *free* edges in connected components.

Now for each free half-edge u in some connected component $\mathscr{C}(t)$ write $\boldsymbol{\pi}(u) \in [n]$ for the vertex that corresponds to this half-edge. Define the distance between two free half-edges u, v in the same component as $d(u,v) = d(\boldsymbol{\pi}(u), \boldsymbol{\pi}(v))$. For a connected component $\mathscr{C}(t)$ write

$$\mathscr{D}_1(\mathscr{C}(t)) = \sum_{u,v\in\mathscr{C}(t), u,v \text{ free}} d(u,v).$$

For fixed free half-edge $u \in \mathscr{C}(t)$, we will use $\mathscr{D}(u) = \sum_{e \in \mathscr{C}, e \,\mathrm{free}} d(e,u)$, with the convention that $d(u,u) = 0$. Define the functions

$$\bar{s}_l(t) := \frac{1}{n}\sum_i [f_i(t)]^l, \qquad \bar{g}(t) := \frac{1}{n}\sum_i f_i(t)|\mathscr{C}_i(t)|, \qquad \bar{\mathscr{D}}(t) := \frac{1}{n}\sum_i \mathscr{D}_1(\mathscr{C}_i(t)). \tag{4.11}$$

Here note that the summation is over all connected components $\mathscr{C}_i(t)$ at time t and **not** over vertices.

Theorem 4.10. *Fix $\delta \in (1/6, 1/5)$ and let t_n be as in* (4.10)*. Then, as $n \to \infty$, the susceptibility functions satisfy*

$$\left|\frac{n^{1/3}}{\bar{s}_2(t_n)} - \frac{\nu^2 n^{1/3-\delta}}{\mu(\nu-1)^2}\right| \xrightarrow{\mathrm{P}} 0, \tag{4.12}$$

$$\frac{\bar{s}_3(t_n)}{[\bar{s}_2(t_n)]^3} \xrightarrow{\mathrm{P}} \frac{\beta}{\mu^3(\nu-1)^3}. \tag{4.13}$$

Further

$$\frac{\bar{\mathscr{D}}(t_n)}{n^{2\delta}} \xrightarrow{\mathrm{P}} \frac{\mu(\nu-1)^2}{\nu^3}, \qquad \frac{\bar{g}(t_n)}{n^{\delta}} \xrightarrow{\mathrm{P}} \frac{(\nu-1)\mu}{\nu^2}. \tag{4.14}$$

To prove this result, we need good upper bounds on the size of the largest component as well as diameter all through the subcritical window. More precisely we show:

Theorem 4.11 (Bounds on diameter and the maximal component)**.** *Given any $\delta < 1/4$ and $\alpha > 0$, there exists $C = C(\delta,\alpha) > 0$ such that*

$$\mathbb{P}\left(|\mathscr{C}_1(t_c - t)| \le \frac{C(\log n)^2}{(t_c-t)^2}, \mathrm{diam}_{\max}(t_c - t) \le \frac{C(\log n)^2}{(t_c - t)} \textit{ for all } 0 \le t < t_c - \frac{\alpha}{n^\delta}\right) \to 1,$$

as $n \to \infty$.

4.3. **Bounded size rules.** Recall the (continuous time) construction of the Bohman-Frieze process $\{\mathrm{BF}_n(t) : t \ge 0\}$ in Section 1.1.3, started with n isolated vertices at $t = 0$. Note that singletons (isolated vertices) play a special role in the evolution of the model. Write $X_n(t)$ and $\bar{x}(t) = X_n(t)/t$ for the number and density of singletons respectively at time t. For $k \ge 1$ let $\bar{s}_k(\cdot)$ denote the k-th susceptibility corresponding to component sizes namely $\bar{s}_k(t) = \sum_i [|\mathscr{C}_i(t)|]^k/n$. The general analysis of bounded size rules in [56] applied to the special case of the Bohman-Frieze process shows that there exist deterministic functions $x(\cdot), s_2(\cdot), s_3(\cdot)$ such that for each fixed $t \ge 0$,

$$\bar{x}(t) \xrightarrow{\mathrm{P}} x(t), \qquad \bar{s}_k(t) \xrightarrow{\mathrm{P}} s_k(t), \text{ for } k = 2, 3.$$

The limiting function $x(t)$ is continuous and differentiable for all $t \in \mathbb{R}_+$. For $k \ge 2$, $s_k(t)$ is finite, continuous and differentiable for $0 \le t < t_c$, and $s_k(t) = \infty$ for $t \ge t_c$. Furthermore, x, s_2, s_3 solve the following differential equations.

$$x'(t) = -x^2(t) - (1 - x^2(t))x(t) \qquad \text{for } t \in [0,\infty) \qquad x(0) = 1 \tag{4.15}$$

$$s_2'(t) = x^2(t) + (1 - x^2(t))s_2^2(t) \qquad \text{for } t \in [0, t_c), \qquad s_2(0) = 1 \tag{4.16}$$

$$s_3'(t) = 3x^2(t) + 3(1 - x^2(t))s_2(t)s_3(t) \qquad \text{for } t \in [0, t_c), \qquad s_3(0) = 1. \tag{4.17}$$

Both s_2, s_3 have singularities at t_c and by [35, Theorem 3.2], there exist constants $\alpha = (1-x^2(t_c))^{-1} \approx 1.063$ and $\beta \approx .764$ such that

$$s_2(t) \sim \frac{\alpha}{t_c - t}, \qquad s_3(t) \sim \beta (s_2(t))^3 \sim \beta \frac{\alpha^3}{(t_c-t)^3} \qquad \text{as } t \uparrow t_c. \tag{4.18}$$

Now define $y(t) = 1/s_2(t)$ and note that y is a monotonically decreasing continuously differentiable function on $[0, t_c)$ with $y(t) = 0$ for all $t \geqslant t_c$. Define the function $v(\cdot)$ on $[0, t_c)$ as the unique solution of the differential equation

$$v'(t) := -2x^2(t)^2 y(t) v(t) + \frac{x^2(t)y^2(t)}{2} + 1 - x^2(t), \qquad v(0) = 0. \tag{4.19}$$

It is easy to check that $v(\cdot)$ is monotonically increasing on $[0, t_c)$ with

$$\lim_{t\uparrow t_c} v(t) := \varrho \approx .811. \tag{4.20}$$

Theorem 4.12. *Fix $\lambda \in \mathbb{R}$ and consider the Bohman-Frieze process in the critical scaling window at time* $\mathrm{BF}_n(t_c + \frac{\beta^{2/3}\alpha}{n^{1/3}}\lambda)$ *where α, β are as in* (4.18). *Then the distance within the maximal components at this time scales like $n^{1/3}$ and further*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\varrho n^{1/3}}, \frac{\beta^{1/3}}{n^{2/3}}\right)\mathscr{C}_i^{(n)}\left(t_c + \frac{\beta^{2/3}\alpha}{n^{1/3}}\lambda\right) : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty(\lambda), \qquad \text{as } n \to \infty, \tag{4.21}$$

where ϱ is as in (4.20).

This model turns out to be the easier amongst the three general families of random graph models since most of the heavy technical estimates on the scaling exponents for the susceptibility functions and maximal component size bounds in the barely subcritical regime have already been proven in [11] which then used these results to show $n^{2/3}$ scaling of maximal component sizes in the critical regime. These were later extended to all bounded size rules in [13, 14]. The same proof as in this paper (with more notation) using the general results in [13] allows us to extend the above result to the following. We omit the proof.

Theorem 4.13. *Fix $\lambda \in \mathbb{R}$. For all $K \geqslant 1$ and bounded size rule $F \subseteq \Omega_K^4$, there exist $\alpha_F, \beta_F, \varrho_F > 0$ such that the maximal components at time $t_c(F) + \frac{\beta_F^{2/3}\alpha_F}{n^{1/3}}\lambda$ satisfy the asymptotics in* (4.21).

5. Discussion

We now discuss the main results. In Section 5.1 we describe qualitative features of the pre-limits that the limits are able to describe. In Section 5.2 and 5.3 we place these results in the context of known results on scaling limits of maximal components at criticality. In Section 5.4 we discuss the differential equations method and in particular their application in this paper in understanding average distances in connected components in the barely subcritical regime. Since many of the results in this paper deal with the barely subcritical regime, we connect the results in this paper to existing results in Section 5.5. We conclude in Section 5.6 with open problems and possible extensions.

5.1. **Universality.** The main of this paper was to develop techniques to prove that for a wide array of models, maximal components in the critical regime scale like $n^{1/3}$ and properly rescaled, converge to the same family of limiting random metric spaces. The key conceptual ideas for carrying out the program was viewing these models not as static but dynamic models evolving over time and showing that

(a) The evolution of the system till the barely subcritical regime ($t_n = t_c - n^{-\delta}$) could be arbitrary and in most cases will be far from that of the multiplicative coalescent (namely the evolution of Erdős-Rényi random graph process) but the configuration of components at time t_n satisfy good properties in terms of moments of component sizes as well the average behavior of distances and maximal distances (Assumption 3.3). For all the random graph models considered in this paper, this boils down to
$$\frac{1}{n}\sum_i |\mathscr{C}_i(t_n)|^2 \sim \alpha n^{\delta}, \frac{1}{n}\sum_i |\mathscr{C}_i(t_n)|^3 \sim \beta n^{3\delta}, \frac{1}{n}\sum_i \sum_{u,v\in\mathscr{C}_i(t_n)} d(u,v) \sim \gamma n^{2\delta},$$
for model dependent constants, coupled with bounds on the maximal diameter and maximal component size in the barely subcritical regime.

(b) After the barely subcritical regime through the critical scaling window, the dynamic version of the model evolves approximately like the multiplicative coalescent.

Coupled with showing that distances in maximal components scale like $n^{1/3}$, this technique also gives results on the sizes of the components, see e.g. Theorem 4.3 for the IRG model where a by product of the analysis is the $n^{2/3}$ scaling of sizes of maximal components. For this model, we were unable to use typical component exploration techniques via breadth-first walk as used in other models [5, 44, 54]. Till date the only other critical IRG model whose component sizes have been analyzed is the rank-one model [17, 57] where the special form of the connection probabilities allows one to explore this graph in a size-biased manner.

Secondly, note that the limit objects are obtained by considering random real trees with a *finite* collection of "shortcuts". In the context of random graph models, this does **not** immediately imply that the complexity or surplus of maximal components at criticality is finite. It only implies that the number of surplus edges created in the time interval $[t_n, t_c + \lambda/n^{1/3}]$ is finite and converges in distribution. To see this distinction, consider the dynamic version of the Erdős-Rényi random graph and let it evolve till time t_n. Now for every component with at least one cherry (two neighboring leaves), choose such a pair and add an edge between them. Now let the process continue to evolve as before. Since we do not change the component sizes, distributional limits of component sizes remain unchanged **and** it is easy to check using Theorem 3.4 that maximal components at time $t = 1 + \lambda/n^{1/3}$ rescaled by $n^{1/3}$ converge to appropriate limits. However the **total** surplus of each maximal component will now be infinite owing to the creation of surplus edges in the modification of the process.

5.2. **Critical random graphs: Known results regarding the Metric structure.** In terms of the actual metric structure of components in the critical regime, only two particular models have been so far analyzed. The first results in this direction were for the Erdős-Rényi random graph were proven in [2, 3]. The first "inhomogenous" random graph model for which similar results were shown was for the so-called rank one model was carried out in

[15]. Some of the technical estimates required for this work, in particular the representation of connected components in terms of tilted versions of $\mathbf{p}$-trees [6,27] were studied in [15]. In [50] critical percolation on random r-regular graphs was studied and it was shown that the diameter of the maximal components in the critical scaling window scaled like $\Theta_P(n^{1/3})$.

5.3. **Critical random graph models: Known results for sizes of components.** For general mathematical treatment of various models of random graphs, see [20,29,30,42,58]. Specific to the connectivity phase transition and component sizes through the scaling window, [22] is recent paper giving a nice overview including the intense activity analyzing the emergence of the giant over the past two decades. With references to sizes of maximal components, there are an enormous number of results in the critical regime. For the Erdős-Rényi random graph there are now a vast number of beautiful results starting with the original paper [31], and explored and expanded in great detail in [39,48]. Particularly significant for this work is Aldous's description [5] of the limiting component sizes in terms of excursions of inhomogeneous reflected Brownian motion, Theorem 2.2. Specific to the other models considered in this paper:

Inhomogenous random graph model: This model was first introduced in its general form in [21] where a wide array of results, including the location of phase transition and properties of the graph in the sub and supercritical regime including typical distances were established. For the critical regime, the only results we are aware of are for the *rank one* model which in the regimes considered is equivalent to the Norros-Reittu [52], Britton-Deijfen [25] and Chung-Lu [28] model. Order of magnitude results for the largest component were derived in [59]. This was sharpened to distributional convergence results in [16,17] and independently in [57].

Configuration model: In [49] the size of the largest component and in particular necessary and sufficient conditions for the existence of a giant component were derived. The continuous time construction is similar to the dynamic construction used in [41] to give a different proof for existence (or lack thereof) of the giant component. Component sizes in the critical regime have been studied starting [51] for the random r-regular graphs in, in [44] under general second moment conditions of the degree, whilst more detailed results applicable to the barely subcritical and supercritical regimes under the assumption of bounded degrees were derived in [54].

Bounded size rules: The first bounded size rule to be rigorously studied was the Bohman-Frieze process [18] which showed that this rule delayed the emergence of the giant component as compared to Erdős-Rényi random graph process. Spencer and Wormald in [56] proved the existence of rule dependent critical times for all bounded size rules. The barely subcritical and critical regime of this class of models with regards to maximal component sizes was studied in [11,13,14].

5.4. **The differential equations method and average distances.** One major tool in dealing with dynamic random graph processes is the differential equations technique where one considers functionals of the process, for example the susceptibility functions $\bar{s}_2, \bar{s}_3$ and show that these converge to limiting deterministic functions s_2, s_3 obtained as solutions to differential equations. See e.g. [45] and for an exhaustive survey of of applications of this technique to random graph processes see [60]. In the context of this paper, the technique proved to be one of the main building blocks in showing the existence of a rule dependent critical point $t_c(F)$ for bounded size rules in [56], obtained as the time at which the limiting susceptibility functions exploded. The limiting differential equations for the susceptibility functions were analyzed in more detail in [35] for the Bohman-Frieze process and were then extended to all bounded size rules in [13]. Since the susceptibility functions of interest explode at t_c while one would still like to read of scaling properties of $\bar{s}$ using the behavior of the limit function s, approximation results with sharp error bounds using semi-martingale techniques were developed in [13] which will play a key role in this paper (Lemma 8.13). The paper [13] used the scaling exponents of susceptibility functions to derive limiting component sizes of maximal components for bounded size rules in the critical scaling window.

In this paper, this technique will be used to understand average distance scaling within components in the barely subcritical regime and in particular show that at time $t_n = t_c - n^{-\delta}$

$$\bar{\mathscr{D}}(t_n) = \frac{1}{n}\sum_i \sum_{u,v\in\mathscr{C}_i(t_n)} d(u,v) \sim \gamma n^{2\delta},$$

for a model dependent constant γ. See Theorem 4.10, Lemma 8.11 and Proposition 8.12 for the configuration model and Proposition 9.6 for the Bohman-Frieze process. We are not aware of any similar applications of this technique in understanding distance scaling for random graph models in the barely subcritical regime. It would be interesting to see if one can derive similar results in other models.

5.5. **Related results in the barely subcritical regime.** Coupled with structural results of components in the critical regime, a number of results in this paper deal with the barely subcritical regime, in particular precise estimates of the susceptibility functions $\bar{s}_2$ and $\bar{s}_3$ as well as the size of the largest component and maximal diameter all at time $t_n = t_c - \varepsilon_n$ where $\varepsilon_n = n^{-\delta}$ with $\delta \in (1/5, 1/6)$; see Theorems e.g. 4.10, 4.4 and Corollary 4.5. For these models, the behavior of the susceptibility functions at times $t_c - \varepsilon$ with $\varepsilon > 0$ fixed as $n \to \infty$ has previously been studied, see e.g. [38] for the configuration model, [43] for the IRG and [35] for Bohman-Frieze process. Since we need to understand these functions close to the regime where they explode with $\varepsilon_n \to 0$, this results in stronger assumptions on the degree sequence for the configuration model and finite type space for the IRG whilst the results in the above papers apply to more general models including configuration models with only moment assumptions on the degree sequence as well as inhomogeneous random graph models with general type space; however the analysis deals with fixed $\varepsilon > 0$. For the Erdős-Rényi random graph process, more precise estimates of the susceptibility function in the barely subcritical regime are derived in [40]. Also see [7] where similar estimates as in this paper were derived for a random graph model with immigration.

5.6. **Open Problems.** In the interest of keeping this paper to a manageable length, we considered the IRG model where the type space χ was finite and all connection intensities $\kappa(x, y) > 0$, whilst in the case of the configuration model we assumed finite exponential moments. We do not believe either of these restrictions are necessary and that all the main results in this paper can be extended under general moment conditions. Now for the IRG model, it is known [21] that when one admits an infinite state space then one can construct models for which the scaling exponents of the size of the maximal component in the barely supercritical regime are quite different from the Erdős-Rényi random graph. For these models, one does not expect analogous results as in this paper either for the critical scaling window or component sizes at criticality. However if one assumes various irreducibility conditions and moment conditions [21, Theorem 3.17] shows that the scaling exponents for general IRG model are similar to the Erdős-Rényi random graph. Thus under general conditions we expect that all the results and proof techniques in this paper can be extended to IRG models with general ground space $\mathscr{X}$. Similarly under high enough moment conditions the same should be true for the configuration model.

6. Proofs: Universality

This section proves Theorem 3.2 for the blob-level superstructure and 3.4 for the complete scaling limit starting with the random graph $\mathscr{G}(\mathbf{x}, q)$ as defined in Section 2.4.

6.1. **Outline of the proof.**

The framework of the proof is as follows.

(a) We start in Section 6.2 with some preliminary constructions related to the model including the important notion of size-biased reordering used in [5] to prove Theorem 2.2.
(b) Section 6.2.2 contains an elementary result decoupling the weights of connected components and the distribution of the components conditional on the weights of vertices in the components.
(c) Section 6.2.3 recalls some of the main results from [15] including scaling limits of connected components of rank one random graphs (Theorem 6.2) via constructing these connected components through tilts of random $\mathbf{p}$-trees and then adding permitted edges independently (Proposition 6.3). This leads to a simple proof of Theorem 3.2 in Section 6.3.
(d) Section 6.4 is the most technical part of this section and completes the proof of Theorem 3.4 via incorporating inter blob-level structure.

6.2. **Preliminaries.**

In this section we recall various constructions from [5, 15] regarding the random graph model $\mathscr{G}(\mathbf{x}, q)$.

6.2.1. **Size-biased re-ordering.**

Recall that given a vertex set $[m]$ with associated positive vertex weights $\{x_i : i \in [m]\}$, a *size biased reordering* of the vertex set is a random permutation $(v(1), v(2), \ldots, v(m))$ of $[m]$ where

$$\begin{aligned} &\mathbb{P}(v(1) = k) \propto x_i, \qquad k \in [m], \\ &\mathbb{P}(v(i) = k | v(1), v(2), \ldots, v(i-1)) \propto x_k, \qquad k \in [m] \setminus \{v(1), \ldots v(i-1)\} \text{ for } i \geqslant 2. \end{aligned} \tag{6.1}$$

An easy way to generate such an order is to first generate independent exponentials $\xi_i \sim \exp(x_i)$ and then consider the permutation generated by arranging these in increasing order namely

$$\xi_{v(1)} < \xi_{v(2)} < \cdots < \xi_{v(m)}.$$

To prove Theorem 2.2, Aldous constructed the random graph $\mathscr{G}(\mathbf{x}, q)$ simultaneously with an exploration of the graph in a size-biased random order. We give a succinct description referring the interested reader to [5, Section 3.1]. For $i \neq j$ let $\xi_{i,j}$ denote independent $\exp(qx_j)$ random variables. To start the construction, the exploration process initializes by selecting a vertex $v(1)$ with probability proportional to the vertex weights $\{x_i : i \in [m]\}$. Then the neighbors (sometimes referred to as children) of $v(1)$ are the vertices $\left\{v : \xi_{v(1),i} \leqslant x_{v(1)}\right\}$. Writing $c(1)$ for the number of children of $v(1)$, label the children of $v(1)$ as $v(2), v(3), \ldots v(c(1)+1)$ in increasing order of the $\xi_{v(1),v(i)}$ values. Now move to $v(2)$ and obtain the unexplored children of $v(2)$ through $\left\{\xi_{v(2),j} : j \neq v(1), \ldots v(c(i)+1)\right\}$, again labeling them in increasing order as $v(c(1)+2), \ldots v(c(1)+c(2)+1)$ in increasing order of their $\xi_{v(2),j}$ values. Proceed recursively until the component of $v(1)$ has been explored. Then select a new vertex amongst unexplored vertices with probability proportional to the weights and proceed until all vertices have been explored. It is easy to check that the order of vertices explored $(v(1), v(2), \ldots v(m))$ is in size-biased random order.

6.2.2. **Partitions of connected components.** Recall that $(\mathscr{C}_i : i \geqslant 1)$ denote the components of $\mathscr{G}(\mathbf{x}, q)$, where the size of a component is the sum of masses x_j in the component. Since we will relate the connected components to random $\mathbf{p}$-trees, which use a probability mass function $\mathbf{p}$ as the driving parameter for their distribution, it will be convenient to parametrize connected components via the relative masses of vertices in these components. We first need some notation. Fix $\mathcal{V} \subset [n]$ and write $\mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$ the space of all simple connected graphs with vertex set $\mathcal{V}$. For fixed $a > 0$, and probability mass function $\mathbf{p} = (p_v : v \in \mathcal{V})$, define the probability distribution on the space of connected graphs with vertex set $\mathcal{V}$, $\mathbb{P}_{\mathrm{con}}(\cdot; \mathbf{p}, a, \mathcal{V})$ on $\mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$ as follows. For $u, v \in \mathcal{V}$ let

$$q_{uv} := 1 - \exp(-a p_u p_v). \tag{6.2}$$

Consider the probability distribution on $\mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$ defined as

$$\mathbb{P}_{\mathrm{con}}(G; \mathbf{p}, a, \mathcal{V}) := \frac{1}{Z(\mathbf{p}, a)} \prod_{(u,v) \in E(G)} q_{uv} \prod_{(u,v) \notin E(G)} (1 - q_{uv}), \text{ for } G \in \mathbb{G}_{\mathcal{V}}^{\mathrm{con}}, \tag{6.3}$$

where $E(G)$ denotes the edge set of the graph G and $Z(\mathbf{p}, a)$ is the normalizing constant

$$Z(\mathbf{p}, a) := \sum_{G \in \mathbb{G}_{\mathcal{V}}^{\mathrm{con}}} \prod_{(u,v) \in E(G)} q_{uv} \prod_{(u,v) \notin E(G)} (1 - q_{uv}).$$

Now let $\mathcal{V}^{(i)} := \{v \in [n] : v \in \mathscr{C}_i\}$ for $i \in \mathbb{N}$ denote the vertex set of component $\mathscr{C}_i$ and note that $\left\{\mathcal{V}^{(i)} : i \geqslant 1\right\}$ denotes a random (finite) partition of the vertex set $[n]$. The following trivial proposition characterizes the distribution of the random graphs $(\mathscr{C}_i : i \geqslant 1)$ conditioned on the partition $\left\{\mathcal{V}^{(i)} : i \geqslant 1\right\}$.

Proposition 6.1. *For $i \geqslant 1$ define*

$$\mathbf{p}^{(i)} := \left(\frac{x_v}{\sum_{v \in \mathcal{V}^{(i)}} x_v} : v \in \mathcal{V}^{(i)} \right), \;\; a^{(i)} := q \left(\sum_{v \in \mathcal{V}^{(i)}} x_v \right)^2. \tag{6.4}$$

Then for any $k \in \mathbb{N}$ and $G_i \in \mathbb{G}_{\mathcal{V}^{(i)}}^{\mathrm{con}}$, we have

$$\mathbb{P}\left(\mathscr{C}_i = G_i, \;\; \forall i \geqslant 1 \mid \left\{\mathcal{V}^{(i)} : i \in \mathbb{N}\right\}\right) = \prod_{i \geqslant 1} \mathbb{P}_{\mathrm{con}}(G_i; \mathbf{p}^{(i)}, a^{(i)}, \mathcal{V}^{(i)}).$$

The above proposition says the random graph $\mathscr{G}(\mathbf{x},q)$ can be generated in two stages.

(i) First generate the partition of the vertices into different components, i.e. $\{\mathcal{V}^{(i)} : i \in \mathbb{N}\}$.
(ii) In the second stage, given the partition, we generate the internal structure of each component following the law of $\mathbb{P}_{\text{con}}(\cdot; \mathbf{p}^{(i)}, a^{(i)}, \mathcal{V}^{(i)})$, independently across different components.

The next Section studies the second aspect of this construction.

6.2.3. **Tilted p-trees and scaling limits of connected components.** Fix $m \geqslant 1$, a probability mass function $\mathbf{p}$ on $[m]$ and $a > 0$ and consider the distribution of the connected random graph in (6.3). Proposition 6.1 suggests that a major issue in understanding scaling limits of the components of $\mathscr{G}(\mathbf{x},q)$ is understanding the asymptotics of $\mathbb{P}_{\text{con}}(\cdot; \mathbf{p}, a, [m])$ as $m \to \infty$, under suitable regularity conditions on $\mathbf{p}, a$. Define

$$\sigma(\mathbf{p}) := \sqrt{\sum_{i\in[m]} p_i^2}, \qquad p_{\max} := \max_{i\in[m]} p_i, \qquad p_{\min} := \min_{i\in[m]} p_i.$$

The following was proved in [15].

Theorem 6.2 ([15, Theorem 7.3])**.** *Assume that there exist $\bar{\gamma} \in (0,\infty)$, $r_0 \in (0,\infty)$, and $\eta_0 \in (0,1/2)$ such that*

$$\lim_{m\to\infty} \sigma(\mathbf{p}) = 0, \ \lim_{m\to\infty} \frac{p_{\max}}{[\sigma(\mathbf{p})]^{3/2+\eta_0}} = 0, \ \lim_{m\to\infty} \frac{[\sigma(\mathbf{p})]^{r_0}}{p_{\min}} = 0, \ \lim_{m\to\infty} a\sigma(\mathbf{p}) = \bar{\gamma}. \tag{6.5}$$

Let $\mathscr{G}_m$ be a $\mathbb{G}_m^{\text{con}}$-valued random variable with law $\mathbb{P}_{\text{con}}$. Under Assumptions (6.5), as $m \to \infty$,

$$\operatorname{scl}\big(\sigma(\mathbf{p}),1\big)\cdot\mathscr{G}_m \xrightarrow{\mathrm{w}} \mathscr{G}(2\tilde{\mathbf{e}}^{\bar{\gamma}}, \bar{\gamma}\tilde{\mathbf{e}}^{\bar{\gamma}}, \mathscr{P}),$$

where the limit metric space $\mathscr{G}(2\tilde{\mathbf{e}}^{\bar{\gamma}}, \bar{\gamma}\tilde{\mathbf{e}}^{\bar{\gamma}}, \mathscr{P})$ is as defined in Section 2.3 using a tilted Brownian excursion of length $\bar{\gamma}$.

A key ingredient of the proof of this theorem is an algorithm for constructing random graphs with distribution $\mathbb{P}_{\text{con}}$ via tilts of $\mathbf{p}$ trees which we now describe. Write $\mathbb{T}_m^{\text{ord}}$ for the space of ordered rooted trees with vertex set m where we view the root as the original progenitor. By ordered we mean the children of every vertex v are given an order from "oldest" to "youngest" (alternatively these are viewed as planar trees). Given a tree $\mathbf{t} \in \mathbb{T}_m^{\text{ord}}$ and vertex $v \in [m]$, write $d_v(\mathbf{t}) \geqslant 0$ for the number of children of v in $\mathbf{t}$. A random $\mathbf{p}$-tree [6,53] is a random tree $\mathscr{T}^{\mathbf{p}}$ with distribution,

$$\mathbb{P}_{\text{ord}}(\mathscr{T}^{\mathbf{p}} = \mathbf{t}) = \prod_{v\in[m]} \frac{p_v^{d_v(\mathbf{t})}}{(d_v(\mathbf{t}))!}, \qquad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \tag{6.6}$$

Given an ordered tree $\mathbf{t} \in \mathbb{T}_m^{\text{ord}}$, write $E(\mathbf{t})$ for the edge set of the tree. Explore this in the depth-first order starting from the root. More precisely we will recursively build three sets of vertices, $\mathscr{A}(\cdot)$, the set of active vertices, $\mathscr{O}(\cdot)$ the set of explored vertices and $\mathscr{U}(\cdot)$ the set of unexplored vertices. We will view $\mathscr{A}(\cdot)$ as a vertical stack with the top most vertex to be explored at the next stage. Initialize with $\mathscr{A}(0) = v(1)$, where $v(1)$ is the root of $\mathbf{t}$, $\mathscr{U}(0) = [m] \setminus \rho$ and $\mathscr{O}(0) = \emptyset$. Having constructed the exploration process till step i, at step $i+1$, we will explore $v(i)$, the vertex on the **top** of the stack in $\mathscr{A}(i)$. Remove the children of $v(i)$ from $\mathscr{U}(i)$ to obtain $\mathscr{U}(i+1)$ and add them to $\mathscr{A}(i)$ in the order prescribed by the tree. Let $\mathscr{O}(i+1) = \mathscr{O}(i) \cup \{v(i)\}$.

Now write $\mathscr{P}(\mathbf{t})$ for the collection of *permitted edges*

$$\mathscr{P}(\mathbf{t}) := \left\{(v(i), j) : i \in [m], j \in \mathscr{A}(i-1)\{v(i)\}\right\},$$

namely the collection of pairs of vertices both of which were present in the active set at some time $i \in [m]$. Write $[m]_2$ for the collection of all possible edges on the vertex set $[m]$ and write $\mathscr{F}(\mathbf{t}) = [m]_2 \setminus ([m]_2 \cup \mathscr{P}(\mathbf{t}))$ for the collection of *forbidden edges.*

Define the function $L : \mathbb{T}_m^{\mathrm{ord}} \to \mathbb{R}_+$ as

$$L(\mathbf{t}) := \prod_{(i,j)\in E(\mathbf{t})} \left[\frac{\exp(ap_ip_j) - 1}{ap_ip_j}\right] \exp\left(\sum_{(i,j)\in\mathscr{P}(\mathbf{t})} ap_ip_j\right), \qquad \mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}. \tag{6.7}$$

Consider the *tilted* $\mathbf{p}$-tree distribution $\tilde{\mathbb{P}}_{\mathrm{ord}}$ as

$$\frac{d\tilde{\mathbb{P}}_{\mathrm{ord}}}{d\mathbb{P}_{\mathrm{ord}}}(\mathbf{t}) = \frac{L(\mathbf{t})}{\mathbb{E}_{\mathrm{ord}}[L(\mathscr{T}^{\mathbf{p}})]}, \text{ for } \mathbf{t} \in \mathbb{T}_m, \tag{6.8}$$

where as before $\mathscr{T}^{\mathbf{p}} \sim \mathbb{P}_{\mathrm{ord}}$ and $\mathbb{E}_{\mathrm{ord}}$ is the corresponding expectation operator with respect to $\mathbb{P}_{\mathrm{ord}}$.

Proposition 6.3 ([15, Proposition 7.4])**.** *Fix a probability mass function* $\mathbf{p}$ *and* $a > 0$. *Then a random graph* $\mathscr{G}_m \sim \mathbb{P}_{\mathrm{con}}$ *with distribution in* (6.3) *can be constructed via the following two step procedure:*

(a) *Generate a random planar tree* $\tilde{\mathscr{T}}$ *with tilted* $\mathbf{p}$*-tree distribution* (6.8).
(b) *Conditional on* $\tilde{\mathscr{T}}$, *add each of the permitted edges* $\{u, v\} \in \mathscr{P}(\tilde{\mathscr{T}})$ *independently with the appropriate probability* q_{uv}.

6.3. **Proof of Theorem 3.2.** The previous section analyzed Proposition 6.1(ii). To complete the proof, we need to analyze (i) of the Proposition and show that the partition of vertex weights satisfy good asymptotic properties at least for the maximal components. Then using Theorem 6.2 completes the proof.

Recall from Theorem 2.2 that $\boldsymbol{\xi}(\lambda) = (\gamma_i(\lambda) : i \geqslant 1)$ denoted limits of weighted component sizes in $\mathscr{G}(\mathbf{x}, q)$. The partition of vertices into different components follows via the size-biased breadth-first exploration used by Aldous in [5], described in Section 6.2.1 to construct the graph $\mathscr{G}(\mathbf{x}, q)$. Aldous used this construction to prove Theorem 2.2 on the weighted sizes of components. This breadth-first exploration generates the partitions of the components in Proposition 6.1 and Aldous used this to prove Theorem 2.2 via analyzing properties of this partition and in particular [5, Lemma 13] shows that for each fixed $i \geqslant 1$,

$$\frac{\sigma_2}{\sigma_3} \cdot \frac{\sum_{v\in\mathscr{C}_i^{(n)}} x_v^2}{\sum_{v\in\mathscr{C}_i^{(n)}} x_v} \xrightarrow{\mathrm{P}} 1, \qquad \text{as } n \to \infty. \tag{6.9}$$

Theorem 2.2 implies in particular that for each fixed $i \geqslant 1$, $\sum_{v\in\mathscr{C}_i} x_v \xrightarrow{\mathrm{w}} \gamma_i(\lambda)$. Assumptions 3.1 coupled with Aldous's original assumptions 2.1 now imply that for each fixed $i \geqslant 1$, (6.5) is satisfied with $a^{(i)}$ as defined in (6.4) satisfying

$$a^{(i)}\sigma(\mathbf{p}^{(i)}) \xrightarrow{\mathrm{w}} \gamma_i^{3/2}(\lambda), \qquad \text{as } n \to \infty. \tag{6.10}$$

Theorem 6.2 and a Brownian scaling argument now completes the result. ∎

6.4. **Proof of Theorem 3.4.** The aim of this section is to consider the case where each of the vertices in $\mathscr{G}(\mathbf{x}, q)$ is in fact a "small" compact connected metric space (a "blob") and $\mathscr{G}(\mathbf{x}, q)$ forms the blob-level superstructure connecting these metric spaces. Recall that in proving Theorem 3.2, we decoupled the problem into two parts (a) studying scaling limits of random graphs conditioned on being connected (Theorem 6.2); (b) Using Proposition 6.1 and the breadth-first exploration in [5] to understand properties of the partition of vertex weights formed by connected components. Similarly here we first start with understanding the case where the blob-level superstructure is connected.

6.4.1. **Connected random graphs with blob-level structure.** We need the following three ingredients analogous to the construction in Section 3.2.

(a) **Blob level superstructure:** Fix $m \geqslant 1$, $a > 0$ and a probability mass function $\mathbf{p} = (p_i : i \in [m])$. Let $\mathscr{G}^{\mathbf{p}}$ be a *connected* random graph with vertex set $[m]$ and distribution (6.3).
(b) **Blobs:** Fix compact connected measured metric spaces $\mathbf{M} = \{(M_i, d_i, \mu_i) : i \in [m]\}$ with μ_i assumed to be a probability measure for all i.
(c) **Blob to blob junction points:** $\mathbf{X} = (X_{i,j} : i, j \in [m])$ be independent random variables (and independent of $\mathscr{G}^{\mathbf{p}}$) such that for each fixed $i \in [m]$, $X_{i,j}$ takes values in M_i with distribution μ_i. Recall from Section 3.2 that for fixed vertex $i \in [m]$ $u_i := \mathbb{E}[d_i(X_{i,1}, X_{i,2})]$ denoted the first moment of the distance between two iid points in M_i with distribution μ_i.

Now define compact connected measured metric space $\bar{\mathscr{G}}^{\mathbf{p}}$ and constant A_m as

$$\bar{\mathscr{G}}^{\mathbf{p}} := \Gamma(\mathscr{G}^{\mathbf{p}}, \mathbf{p}, \mathbf{M}, \mathbf{X}), \qquad A_m := \sum_{i \in [m]} p_i u_i, \tag{6.11}$$

where the operation Γ is as in Section 3.2. The following is the main result of this section.

Theorem 6.4. *Assume* (6.5) *holds and further assume*

$$\lim_{m \to \infty} \frac{[\sigma(\mathbf{p})]^{1/2-\eta_0} d_{\max}}{A_m + 1} = 0. \tag{6.12}$$

Then we have

$$\frac{\sigma(\mathbf{p})}{A_m + 1} \bar{\mathscr{G}}^{\mathbf{p}} \xrightarrow{\mathrm{w}} \mathscr{G}(2\tilde{\mathbf{e}}^{\bar{\gamma}}, \bar{\gamma}\tilde{\mathbf{e}}^{\bar{\gamma}}, \mathscr{P}), \ \textit{as } m \to \infty.$$

Proof: Using Proposition 6.3, we assume that $\mathscr{G}^{\mathbf{p}}$ and $\bar{\mathscr{G}}^{\mathbf{p}}$ have been constructed as follows: **(a)** Generate the junction points $\mathbf{X}$; **(b)** Construct $\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}$ using the tilted $\mathbf{p}$-tree distribution in (6.8). Using the junction points $\mathbf{X}$, let $\bar{\mathscr{T}}^{\mathbf{p}}_{\mathrm{tilt}} := \Gamma(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}, \mathbf{p}, \mathbf{M}, \mathbf{X})$. **(c)** Obtain $\mathscr{G}^{\mathbf{p}}$ by adding permitted edges $\{u, v\} \in \mathscr{P}(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}})$ with probability q_{uv}, independent across edges. Call the collection of edges added at this stage:

$$\mathrm{Surplus}(\mathscr{G}^{\mathbf{p}}) := \{(i_k, j_k) : 1 \leqslant k \leqslant \ell\} = E(\mathscr{G}^{\mathbf{p}}) \setminus E(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}),$$

for the collection of *surplus* edges. Once again use the junction points $\mathbf{X}$ to obtain $\bar{\mathscr{G}}^{\mathbf{p}}$ using these surplus edges. Write $\mathrm{spls}(\mathscr{G}^{\mathbf{p}}) = |E(\mathscr{G}^{\mathbf{p}}) \setminus E(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}})|$ for the number of surplus edges in $\mathscr{G}^{\mathbf{p}}$.

Now note that by Theorem 6.2, under assumptions (6.5), the blob level superstructure $\mathscr{G}^{\mathbf{p}}$ satisfies

$$\sigma(\mathbf{p})\mathscr{G}^{\mathbf{p}} \xrightarrow{\mathrm{w}} \mathscr{G}(2\tilde{\mathbf{e}}^{\bar{\gamma}}, \bar{\gamma}\tilde{\mathbf{e}}^{\bar{\gamma}}, \mathscr{P}), \text{ as } m \to \infty.$$

Thus we only need to prove

$$d_{\mathrm{GHP}}\left(\sigma(\mathbf{p})\mathscr{G}^{\mathbf{p}}, \frac{\sigma(\mathbf{p})}{A_m+1}\bar{\mathscr{G}}^{\mathbf{p}}\right) \xrightarrow{\mathrm{P}} 0, \text{ as } m \to \infty. \tag{6.13}$$

By [4, Lemma 4.2], it suffices to show that the following two *pointed measured metric spaces* are close to each other:

$$\Xi_1^{(m)} := (\sigma(\mathbf{p})\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}; i_1, j_1, \ldots, i_\ell, j_\ell) \text{ and } \Xi_2^{(m)} := \left(\frac{\sigma(\mathbf{p})}{A_m+1}\bar{\mathscr{T}}^{\mathbf{p}}_{\mathrm{tilt}}; X_{i_1,j_1}, X_{j_1,i_1}, \ldots, X_{i_\ell,j_\ell}, X_{j_\ell,i_\ell}\right), \tag{6.14}$$

where as before $\{(i_k, j_k) : k = 1,2,\ldots,\ell\} = E(\mathscr{G}^{\mathbf{p}}) \setminus E(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}})$ denotes the set of surplus edges. More precisely, writing $\mathscr{T}^{\mathbf{p}} = ([m], d_{\mathscr{T}}, \mu)$ and $\bar{\mathscr{T}}^{\mathbf{p}} = (\bar{M} = \sqcup_{i\in[m]} M_i, d_{\bar{\mathscr{T}}}, \bar{\mu})$, consider the correspondence C_m and the measure ν_m on $\mathscr{T}^{\mathbf{p}} \times \bar{\mathscr{T}}^{\mathbf{p}}$

$$C_m = \{(i,x) : i \in [m], x \in M_i\}.$$
$$\nu_m(\{i\} \times A) = p_i \mu_i(A \cap M_i), \text{ for } i \in [m], A \subset \bar{M}.$$

Note that the correspondence C_m has the following property: for all $(i,j) \in E(\mathscr{G}^{\mathbf{p}}) \setminus E(\mathscr{T}^{\mathbf{p}})$ and the corresponding $(X_{i,j}, X_{j,i})$, we have $(i, X_{i,j}), (j, X_{j,i}) \in C_m$. Therefore C_m is actually a correspondence between the two pointed metric spaces. Since the limiting random metric spaces in Section 2.3 consist of identifying points in random real trees, it is more convenient to work with slight variants of the original metric spaces $\mathscr{G}^{\mathbf{p}}$ and $\bar{\mathscr{G}}^{\mathbf{p}}$. Write $\mathscr{G}^{\mathbf{p}}_*$ [$\bar{\mathscr{G}}^{\mathbf{p}}_*$] for the metric space obtained from $\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}$ by identifying all pairs of (i,j) [$(X_{i,j}, X_{j,i})$] for all surplus edges $(i,j) \in E(\mathscr{G}^{\mathbf{p}}) \setminus E(\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}})$, instead of putting an edge of length one between i,j as in $\mathscr{G}^{\mathbf{p}}$. Write $\mathrm{dis}(C_m)$ for the distortion of C_m, where we view C_m as a correspondence between $\Xi_1^{(m)}$ and $\Xi_2^{(m)}$ where $\Xi_i^{(m)}$ are as in (6.14). Write $\mathrm{dsc}(\nu_m)$ for the discrepancy of ν_m. By [4, Lemma 4.2], we have

$$d_{\mathrm{GHP}}\left(\sigma(\mathbf{p})\mathscr{G}^{\mathbf{p}}_*, \frac{\sigma(\mathbf{p})}{A_m+1}\bar{\mathscr{G}}^{\mathbf{p}}_*\right) \le (\mathrm{spls}(\mathscr{G}^{\mathbf{p}})+1)\max\left\{\frac{1}{2}\mathrm{dis}(C_m), \mathrm{dsc}(\nu_m), \nu_m(C_m^c)\right\}, \tag{6.15}$$

It is easy to check that for all m, $\mathrm{dsc}(\nu_m) = \nu_m(C_m^c) = 0$ and

$$d_{\mathrm{GHP}}(\mathscr{G}^{\mathbf{p}}_*, \mathscr{G}^{\mathbf{p}}) \le \mathrm{spls}(\mathscr{G}^{\mathbf{p}}), \;\; d_{\mathrm{GHP}}(\bar{\mathscr{G}}^{\mathbf{p}}_*, \bar{\mathscr{G}}^{\mathbf{p}}) \le \mathrm{spls}(\mathscr{G}^{\mathbf{p}}).$$

Since $\sigma(\mathbf{p}) \to 0$ as $m \to \infty$, by (6.15) in order to complete the proof of (6.13), we only need to show:

$$\mathrm{dis}(C_m) \xrightarrow{\mathrm{P}} 0 \text{ as } m \to \infty \tag{6.16}$$

$$\mathrm{spls}(\mathscr{G}^{\mathbf{p}}) \text{ is tight.} \tag{6.17}$$

Negligibility of the distortion: We first study $\mathrm{dis}(C_m)$ and prove (6.16). For $x \in \bar{M}$, write $i(x)$ for the unique $i \in [m]$ such that $x \in M_i$. Recall that $d_{\mathscr{T}}$ and $d_{\bar{\mathscr{T}}}$ are the distances on $\mathscr{T}^{\mathbf{p}}_{\mathrm{tilt}}$ and $\bar{\mathscr{T}}^{\mathbf{p}}_{\mathrm{tilt}}$, respectively. Then by definition

$$\mathrm{dis}(C_m) = \sup_{x,y\in\bar{M}}\left\{|\sigma(\mathbf{p})d_{\mathscr{T}}(i(x), i(y)) - \frac{\sigma(\mathbf{p})}{A_m+1}d_{\bar{\mathscr{T}}}(x,y)|\right\}. \tag{6.18}$$

Now independent of $(\mathscr{G}^{\mathbf{p}}, \mathbf{X}, \mathscr{T}_{\text{tilt}}^{\mathbf{p}})$ consider the following set of independent random variables: (a) Select two blobs I and J from $[m]$ with distribution $\mathbf{p}$; (b) For for each $i \in [m]$, let Y_i be M_i-valued random variable with distribution μ_i.

Proposition 6.5. *For fixed $\epsilon > 0$, there exist $m_0 = m_0(\varepsilon)$ such that for all $m > m_0$,*

$$\mathbb{P}(\text{dis}(C_m) > \epsilon) \leq \frac{1}{p_{\min}} \mathbb{P}\left(\left|\sigma(\mathbf{p}) d_{\mathscr{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(Y_I, Y_J)\right| > \frac{\epsilon}{5}\right).$$

Proof: First note that since $\sigma(\mathbf{p}) \to 0$, using (6.12) we have,

$$\frac{\sigma(\mathbf{p}) d_{\max}}{A_m+1} \to 0 \text{ as } m \to \infty. \tag{6.19}$$

We start by showing that for any fixed blob $i_0 \in [m]$ and point $x_0 \in M_{i_0}$, we have

$$\text{dis}(C_m) \leq 4 \sup_{y \in \bar{M}} \left\{ |\sigma(\mathbf{p}) d_{\mathscr{T}}(i_0, i(y)) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(x_0, y)| \right\} + \frac{2\sigma(\mathbf{p})}{A_m+1} d_{\max}. \tag{6.20}$$

For two fixed points $x, y \in \bar{M}$, there are two unique paths $(i_0, \ldots, i(x))$ and $(i_0, \ldots, i(y))$ in the tree $\mathscr{T}_{\text{tilt}}^{\mathbf{p}}$. Write $(i_0, \ldots, i_k)$ for the longest common path shared by these two paths. Let $i_* = i_k$ and $x_* = X_{i_k, i_{k-1}}$. Since $\mathscr{T}_{\text{tilt}}^{\mathbf{p}}$ is a tree, we have

$$d_{\mathscr{T}}(i(x), i(y)) = d_{\mathscr{T}}(i_0, i(x)) + d_{\mathscr{T}}(i_0, i(y)) - 2 d_{\mathscr{T}}(i_0, i_*). \tag{6.21}$$

By a similar observation but now for $\bar{\mathscr{T}}_{\text{tilt}}^{\mathbf{p}}$, we have

$$d_{\bar{\mathscr{T}}}(x,y) \leq d_{\bar{\mathscr{T}}}(x_0, x) + d_{\bar{\mathscr{T}}}(x_0, y) - 2 d_{\bar{\mathscr{T}}}(x_0, x_*) \leq d_{\bar{\mathscr{T}}}(x,y) + 2 d_{\max}. \tag{6.22}$$

Equation (6.20) then follows by using (6.21) and (6.22) in (6.18). Next, we replace every $y \in M_i$ in (6.20) with $Y_i \in M_i$, and this incurs an error of at most $4\sigma(\mathbf{p}) d_{\max}/(A_m+1)$ in the right hand side of (6.20). Therefore we have

$$\text{dis}(C_m) \leq 4 \sup_{i \in [m]} \left\{ \left| \sigma(\mathbf{p}) d_{\mathscr{T}}(i_0, i) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(x_0, Y_i) \right| \right\} + \frac{6\sigma(\mathbf{p}) d_{\max}}{A_m+1}.$$

Using (6.19), we can find m_0 such that $6\sigma(\mathbf{p}) d_{\max}/(A_m+1) < \epsilon/5$ for $m > m_0$. Thus,

$$\begin{aligned}
\mathbb{P}\{\text{dis}(C_m) > \epsilon\} &\leq \mathbb{P}\left(\sup_{i \in [m]} \left\{ |\sigma(\mathbf{p}) d_{\mathscr{T}}(i_0, i) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(x_0, Y_i)| \right\} > \frac{\epsilon}{5} \right) \\
&\leq \frac{1}{p_{\min}} \sum_{i \in [m]} p_i \mathbb{P}\left(|\sigma(\mathbf{p}) d_{\mathscr{T}}(i_0, i) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(x_0, Y_i)| > \frac{\epsilon}{5} \right) \\
&= \frac{1}{p_{\min}} \mathbb{P}\left(|\sigma(\mathbf{p}) d_{\mathscr{T}}(i_0, J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathscr{T}}}(x_0, Y_J)| > \frac{\epsilon}{5} \right),
\end{aligned}$$

Since x_0 and i_0 are arbitrary, we get the same bound but now taking the random blob $i_0 = I$ and the point x_0 in M_I now replaced by the random point Y_I. This completes the proof.

■

Now we continue with the proof of (6.16). Recall from (6.8) that $\mathscr{T}_{\text{tilt}}^{\mathbf{p}}$ was constructed via tilting the distribution of a random $\mathbf{p}$-tree using the function $L(\cdot)$. Let $\mathscr{T}^{\mathbf{p}}$ be a random

$\mathbf{p}$-tree with the (untilted) distribution (6.6). Using the bound in Lemma 6.5 and Holder's inequality, we have for fixed $\varepsilon > 0$ and $m > m_0$,

$$\begin{aligned}
&\mathbb{P}(\operatorname{dis}(C_m) > \epsilon) \\
\leq &\frac{1}{p_{\min}} \mathbb{P}\left(|\sigma(\mathbf{p}) d_{\mathcal{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathcal{T}}}(Y_I,Y_J)| > \frac{\epsilon}{5}\right) \\
= &\frac{1}{p_{\min}\mathbb{E}_{\mathrm{ord}}[L(\mathcal{T}^{\mathbf{p}})]} \mathbb{E}_{\mathrm{ord}}\left[L(\mathcal{T}^{\mathbf{p}}) \mathbb{1}\left\{|\sigma(\mathbf{p}) d_{\mathcal{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathcal{T}}}(Y_I,Y_J)| > \frac{\epsilon}{5}\right\}\right] \\
\leq &\frac{\left(\mathbb{E}_{\mathrm{ord}}[L^{q_1}(\mathcal{T}^{\mathbf{p}})]\right)^{1/q_1}}{p_{\min}\mathbb{E}_{\mathrm{ord}}[L(\mathcal{T}^{\mathbf{p}})]} \left(\mathbb{P}_{\mathrm{ord}}\left\{|\sigma(\mathbf{p}) d_{\mathcal{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathcal{T}}}(Y_I,Y_J)| > \frac{\epsilon}{5}\right\}\right)^{1/q_2}, \\
\leq &\frac{C(q_1)}{p_{\min}} \left(\mathbb{P}_{\mathrm{ord}}\left\{|\sigma(\mathbf{p}) d_{\mathcal{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathcal{T}}}(Y_I,Y_J)| > \frac{\epsilon}{5}\right\}\right)^{1/q_2}
\end{aligned} \tag{6.23}$$

where $q_1, q_2 > 1$ and $1/q_1 + 1/q_2 = 1$. Here to simplify notation we continue to use $d_{\mathcal{T}}$ and $d_{\bar{\mathcal{T}}}$ to represent distances but now for the random tree $\mathcal{T}^{\mathbf{p}}$ and $\bar{\mathcal{T}}^{\mathbf{p}} := \Gamma(\mathcal{T}^{\mathbf{p}}, \mathbf{p}, \mathbf{M}, \mathbf{X})$ respectively, where as before the junction points $\mathbf{X}$ have been generated independently of $\mathcal{T}^{\mathbf{p}}$. Further in arriving at the last inequality we have used the fact that $L(\mathcal{T}^{\mathbf{p}}) \geqslant 1$ and under Assumptions (6.5), by [15, Corollary 7.13],

$$\left(\mathbb{E}_{\mathrm{ord}}[L^{q_1}(\mathcal{T}^{\mathbf{p}})]\right)^{1/q_1} \leqslant C(q_1),$$

for some constant depending only on q_1. Now note that $d_{\mathcal{T}}(I,J)$ is the distance between two random vertices selected according to distribution $\mathbf{p}$ from the random $\mathbf{p}$-tree with distribution $\mathcal{T}^{\mathbf{p}}$. Write $R^* := d_{\mathcal{T}}(I,J)$, and let $(I_0 = I, I_1, ..., I_{R^*-1} = J)$ be the actual path between I and J in $\mathcal{T}^{\mathbf{p}}$. Define $\mathbb{R}^+$-valued random variables $\left\{\xi_i^* : 0 \leqslant i \leqslant R^*-1\right\}$ as follows: For the end points, let $\xi_0^* = d_I(Y_I, X_{I,I_1})$, $\xi_{R^*-1}^* = d_J(X_{J,I_{R^*-2}}, Y_J)$. Let

$$\xi_i^* = d_{I_i}(X_{I_i,I_{i-1}}, X_{I_i,I_{i+1}}), \qquad \text{for } 1 \leqslant i \leqslant R^*-2.$$

As before we remind the reader that the junction points $\mathbf{X}$ and the reference points $\mathbf{Y} = \{Y_i : i \in [m]\}$ are independent of $\mathcal{T}^{\mathbf{p}}$. With the above notation, for the distances in $\bar{\mathcal{T}}^{\mathbf{p}}$ and $\mathcal{T}^{\mathbf{p}}$ we have

$$d_{\bar{\mathcal{T}}}(Y_I,Y_J) = \sum_{i=0}^{R^*-1} \xi_i^* + (R^*-1) \qquad \text{and} \qquad d_{\mathcal{T}}(I,J) = R^*-1.$$

Thus

$$\sigma(\mathbf{p}) d_{\mathcal{T}}(I,J) - \frac{\sigma(\mathbf{p})}{A_m+1} d_{\bar{\mathcal{T}}}(Y_I,Y_J) = \frac{\sigma(\mathbf{p})}{A_m+1}\left(\sum_{i=0}^{R^*-1}(A_m - \xi_i^*) - A_m\right) \tag{6.24}$$

In order to estimate the probability in (6.23), the final ingredient we will need is a construction from [27] of the path between two vertices sampled according to $\mathbf{p}$ in a random $\mathbf{p}$-tree $\mathcal{T}^{\mathbf{p}}$. This construction coupled with extra randomization for the junction points $\mathbf{X}$ and the reference points $\{Y_i : i \in [m]\}$ allows us to explicitly construct the joint distribution as $d_{\bar{\mathcal{T}}}(Y_I,Y_J)$ and $d_{\mathcal{T}}(I,J)$. The construction is as follows:

Let $\mathbf{J} := \{J_i\}_{i\geqslant 0}$ be a sequence of i.i.d. $[m]$-valued random variables with law $\mathbf{p}$. For each fixed $j \in [m]$, let $\boldsymbol{\xi}^{(j)} := \left\{\xi_i^{(j)}\right\}_{i\geqslant 0}$ be a sequence i.i.d. copies of the random variable

$d_j(X_{j,1}, X_{j,2})$, independent across $j \in [m]$ and of the family $\mathbf{J}$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the probability space on which the collection of random variables $\{\mathbf{J}, \boldsymbol{\xi}^{(j)} : j \in [m]\}$ are defined.

Now note that for any fixed $k \geqslant 1$, the size $|\{J_i : 0 \leqslant i \leqslant k\}|$ gives the number of distinct elements in this set. Define $R := \inf\{k \geqslant 1 : |\{J_i : i = 0, 1, ..., k\}| < k+1\}$ for the first repeat time of the sequence $\mathbf{J}$. By [27, Corollary 3] the path between I and J in $\mathscr{T}^{\mathbf{p}}$ can be constructed as

$$(I_0, ..., I_{R^*-1}; R^*)_{\mathbb{P}_{\text{ord}}} \overset{d}{=} (J_0, ..., J_{R-1}; R)_{\mathbb{P}}.$$

From the construction of $(\mathscr{T}^{\mathbf{p}}, \bar{\mathscr{T}}^{\mathbf{p}})$ we have

$$(I_0, ..., I_{R^*-1}; R^*; \xi_0^*, ..., \xi_{R^*-1}^*)_{\mathbb{P}_{\text{ord}}} \overset{d}{=} (J_0, ..., J_{R-1}; R; \xi_0^{(J_0)}, ..., \xi_{R-1}^{(J_{R-1})})_{\mathbb{P}'}.$$

Using (6.24) and $\sigma(\mathbf{p}) \to 0$ as $m \to \infty$ we have for fixed $\varepsilon > 0$ and all large m,

$$\mathbb{P}_{\text{ord}}\left(\left|\sigma(\mathbf{p}) d_{\mathscr{T}}(I, J) - \frac{\sigma(\mathbf{p})}{A_m + 1} d_{\bar{\mathscr{T}}}(Y_I, Y_J)\right| > \frac{\epsilon}{5}\right) \leqslant \mathbb{P}\left(\frac{\sigma(\mathbf{p})}{A_m + 1}\left|\sum_{i=0}^{R-1} (\xi_i^{(J_i)} - A_m)\right| > \frac{\epsilon}{6}\right) \tag{6.25}$$

Note that $\xi_i^{(J_i)}$ for $i \geqslant 1$, is a collection of i.i.d. random variables with $0 \leqslant \xi_i^{(J_i)} \leqslant d_{\max}$ and with mean

$$\mathbb{E}[\xi_0^{J_0}] = \sum_{i \in [m]} p_i \mathbb{E}[\xi_0^{(i)}] = \sum_{i \in [m]} p_i u_i = A_m.$$

Thus the sequence $\left\{\sum_{i=0}^{k} (\xi_i^{(J_i)} - A_m)\right\}_{k \geqslant 0}$ is a martingale with respect to the natural filtration. Further

$$\mathbb{P}\left(\frac{\sigma(\mathbf{p})}{A_m + 1}\left|\sum_{i=0}^{R-1} (\xi_i^{(J_i)} - A_m)\right| > \frac{\epsilon}{6}\right) \leqslant \mathbb{P}(R \geqslant t) + \mathbb{P}\left(\sup_{0 \leqslant k \leqslant t-1}\left|\sum_{i=0}^{k} (\xi_i^{(J_i)} - A_m)\right| > \frac{\epsilon(A_m + 1)}{6\sigma(\mathbf{p})}\right). \tag{6.26}$$

The first term in the above display is bounded via the following lemma.

Lemma 6.6. *For any $t \in (0, 1/p_{\max})$, we have*

$$\mathbb{P}(R \geqslant t) \leqslant 2\exp\left(-\frac{t^2 \sigma^2(\mathbf{p})}{24}\right).$$

Proof: Following [27], we assume that the i.i.d. sequence $\mathbf{J} = \{J_i\}_{i \geqslant 0}$ has been constructed through embedding in a Poisson process as follows. Let $N = \{(S_i, U_i)\}_{i \geqslant 0}$ be a rate one Poisson point process on $[0, \infty)$ with points arranged as $0 < S_0 < S_1 < ....$ Partition the interval $[0, 1]$ into m intervals $\{B_i : i \in [m]\}$ such that the length of B_i is p_i. Now for $i \geqslant 0$ let

$$J_i = \sum_{j \in [m]} j \mathbb{1}_{\{U_i \in B_j\}}.$$

Write $N(t) := N([0, t] \times [0, 1])$ and $N(t-) := N([0, t) \times [0, 1])$. Define

$$T = \inf\{t \geqslant 0 : N(t) > R\}.$$

Thus $T = S_R$ and $N(T-) = R$. For any $t > 0$ we have

$$\begin{aligned}\mathbb{P}(R \geqslant t) &\leq \mathbb{P}\left(T > \frac{t}{2}\right) + \mathbb{P}\left(T \leqslant \frac{t}{2},\ R \geqslant t\right)\\ &\leq \mathbb{P}\left(T > \frac{t}{2}\right) + \mathbb{P}\left(N\left(\frac{t}{2}\right) \geqslant t\right)\end{aligned} \tag{6.27}$$

For the second term in (6.27), basic tail bounds for the Poisson distribution imply

$$\mathbb{P}\left(N\left(\frac{t}{2}\right) \geqslant t\right) \leqslant \exp\left(-(2\log 2 - 1/2)\,t\right) < e^{-.19t}, \tag{6.28}$$

For the first term in (6.27) using [27, Equations (26), (29)], we have for all $0 < t < 1/p_{\max}$ we have

$$\log \mathbb{P}(T > t) \leqslant -\frac{t^2}{2}\sigma^2(\mathbf{p}) + \frac{t^3}{3}\frac{p_{\max}\sigma^2(\mathbf{p})}{1 - t p_{\max}} \leqslant -\frac{t^2\sigma^2(\mathbf{p})}{6}.$$

For the first term in (6.27) using [27, Equations (26), (29)], we have for all $0 < t < 1/p_{\max}$

$$\log \mathbb{P}\left(T > \frac{t}{2}\right) \leqslant -\frac{t^2}{8}\sigma^2(\mathbf{p}) + \frac{t^3}{12}\cdot\frac{p_{\max}\sigma^2(\mathbf{p})}{2 - t p_{\max}} \leqslant -\frac{t^2\sigma^2(\mathbf{p})}{24}.$$

Using the above bound, (6.27), (6.28), and the fact that $t\sigma^2(\mathbf{p}) \le \sigma^2(\mathbf{p})/p_{\max} \le 1$ completes the proof of Lemma 6.6.

■

The second term on the right hand side of (6.26) can be bounded by using Markov inequality and the Burkholder-Davis-Gundy inequality. For $i \geqslant 0$ write $\Delta_i := \xi_i^{(J_i)} - A_m$ for the martingale differences. For fixed $r \geqslant 1$ we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{0\leqslant k\leqslant t-1}\left|\sum_{i=0}^{k}\Delta_i\right| > \frac{\epsilon(A_m+1)}{6\sigma(\mathbf{p})}\right) &\le \left(\frac{6\sigma(\mathbf{p})}{\epsilon(A_m+1)}\right)^{2r}\mathbb{E}\left[\sup_{0\leqslant k\leqslant t-1}\left|\sum_{i=0}^{k}\Delta_i\right|^{2r}\right] \\
&\le \left(\frac{6\sigma(\mathbf{p})}{\epsilon(A_m+1)}\right)^{2r}\cdot C(r)\,\mathbb{E}\left[\left(\sum_{i=1}^{t-1}\Delta_i^2\right)^{r}\right] \\
&\le \left(\frac{6\sigma(\mathbf{p})}{\epsilon(A_m+1)}\right)^{2r}\cdot C(r)\,t^r\,\mathbb{E}\left[\Delta_0^{2r}\right], \\
&\le \left(\frac{12\sigma(\mathbf{p})}{\epsilon(A_m+1)}\right)^{2r}\cdot C(r)\,t^r\,d_{\max}^{2r}
\end{aligned} \tag{6.29}$$

where the second inequality uses the Burkholder-Davis-Gundy inequality, the third inequality uses the Jenson's inequality, and the last bound uses the elementary bound $\mathbb{E}|\xi_0^{J_0} - \mathbb{E}\xi_0^{J_0}|^{2r} \leqslant \mathbb{E}|\xi_0^{J_0} + \mathbb{E}\xi_0^{J_0}|^{2r} \leqslant 2^{2r}\,\mathbb{E}[(\xi_0^{J_0})^{2r}] \leqslant 2^{2r} d_{\max}^{2r}$. Combining (6.26), (6.29) and Lemma 6.6, we have, for all $t < 1/p_{\max}$,

$$\begin{aligned}
\mathbb{P}\left\{\frac{\sigma(\mathbf{p})}{A_m+1}\left|\sum_{i=0}^{R-1}(\xi_i^{(J_i)} - A_m)\right| > \frac{\epsilon}{5}\right\} &\le 2\exp\left(-\frac{t^2\sigma^2(\mathbf{p})}{24}\right) + \left(\frac{12\sigma(\mathbf{p})}{\epsilon(A_m+1)}\right)^{2r}\cdot C(r)\,t^r\,d_{\max}^{2r} \\
&:= \mathscr{B}_1 + \mathscr{B}_2.
\end{aligned} \tag{6.30}$$

Taking $t = t_m := 4\sqrt{-r\log\sigma(\mathbf{p})}/\sigma(\mathbf{p})$ so that we have

$$t_m p_{\max} = o\left(\frac{p_{\max}}{[\sigma(\mathbf{p})]^{3/2}}\right) \to 0, \quad t_m\sigma(\mathbf{p}) = 4\sqrt{-r\log\sigma(\mathbf{p})} \to \infty, \text{ as } m \to \infty,$$

where the first convergence uses the assumption (6.5). Thus when m is large we have

$$\mathscr{B}_1 \leqslant 2\exp\left(-\frac{t_m^2}{16}\sigma^2(\mathbf{p})\right) = 2[\sigma(\mathbf{p})]^r.$$

Denoting $\alpha_m := 4\sqrt{-r\log\sigma(\mathbf{p})}$, we have, when m is large,

$$\mathscr{B}_2 = \frac{12^{2r}C(r)}{\epsilon^{2r}}\alpha_m^r \cdot \frac{\sigma^r(\mathbf{p})d_{\max}^{2r}}{(A_m+1)^{2r}} \leqslant C(r,\epsilon)\alpha_m^r[\sigma(\mathbf{p})]^{2r\eta_0},$$

where the last bound uses the assumption (6.12). Since $r > 2r\eta_0$ and $\alpha_m \to \infty$, combining the above two bounds with (6.30), we have, for m large,

$$\mathbb{P}\left\{\frac{\sigma(\mathbf{p})}{A_m+1}\left|\sum_{i=0}^{R-1}(\xi_i^{(J_i)} - A_m)\right| > \frac{\epsilon}{6}\right\} \leqslant 2C(r,\epsilon)\alpha_m^r[\sigma(\mathbf{p})]^{2r\eta_0}.$$

Combining the above bound, (6.25) and (6.23), we have

$$\mathbb{P}(\mathrm{dis}(C_m) > \epsilon) \leqslant C(q_2,r,\epsilon)\frac{1}{p_{\min}}\alpha_m^{r/q_2}[\sigma(\mathbf{p})]^{2r\eta_0/q_2},$$

where $C(q_1,r,\epsilon)$ is a constant depending only on q_2, r and ϵ. Let $q_2 = 2$, $r = \lfloor r_0/\eta_0\rfloor + 1$ and then letting $m \to \infty$, by assumption (6.5) and the fact that α_m is a power of $-\log\sigma(\mathbf{p})$ implies that the above expression goes to zero. This completes the proof of (6.16) and thus the negligibility of the distortion of the correspondence. ■

Tightness of the Surplus: Next, in order to complete the proof of (6.13), we only need to verify the tightness namely (6.17). Note that Proposition 6.3(b) implies that to obtain the surplus edges, we add all permitted edges in $\{u,v\} \in \mathscr{P}(\mathscr{T}_{\mathrm{tilt}}^{\mathbf{p}})$ independently with probability proportional to $q_{u,v} = 1 - \exp(-ap_up_v)$. Thus we have

$$\mathbb{E}[\mathrm{spls}(\mathscr{G}^{\mathbf{p}})] \leqslant \tilde{\mathbb{E}}_{\mathrm{ord}}\left[\sum_{(i,j)\in\mathscr{P}(\mathscr{T}_{\mathrm{tilt}}^{\mathbf{p}})} ap_ip_j\right],$$

where $\tilde{\mathbb{E}}_{\mathrm{ord}}$ is the expectation with respect to $\tilde{\mathbb{P}}_{\mathrm{ord}}$ as in (6.8).Using the definition of the tilted distribution $\tilde{\mathbb{P}}_{\mathrm{ord}}$ with reference to the original distribution $\mathbb{P}_{\mathrm{ord}}$ in (6.8) and the form of the tilt function $L(\mathbf{t})$ in (6.7), we have

$$\mathbb{E}[\mathrm{spls}(\mathscr{G}^{\mathbf{p}})] \leqslant \mathbb{E}_{\mathrm{ord}}\left[L(\mathscr{T}^{\mathbf{p}})\sum_{(i,j)\in\mathscr{P}(\mathscr{T}^{\mathbf{p}})} ap_ip_j\right] \leqslant \mathbb{E}_{\mathrm{ord}}[L^2(\mathscr{T}^{\mathbf{p}})] < C,$$

where $\mathscr{T}^{\mathbf{p}}$ as before is the untilted $\mathbf{p}$-tree with distribution (6.6) and where the last bound follows from [15, Corollary 7.13] under Assumptions (6.5). Here C is an absolute constant independent of m. Thus we have $\sup_{m\geqslant 1}\mathbb{E}[\mathrm{spls}(\mathscr{G}^{\mathbf{p}})] < \infty$, which implies tightness. This completes the proof of Theorem 6.4. ■

6.4.2. **Completing the proof of Theorem 3.4.** We prove the assertion of Theorem 3.4 for the maximal component $\mathscr{C}_1$; the same proof works for any component $\mathscr{C}_k$ for fixed $k \geqslant 1$ using Theorem 2.2. Write m for the number of blobs in $\mathscr{C}_1$ and let $\bar{\mathbf{M}} := \{(\bar{M}_i, \bar{d}_i, \bar{\mu}_i) : i \in [m]\}$ be the collection of blobs in $\mathscr{C}_1$. Recall that $\{u_i : i \in [m], k \geqslant 1\}$ for the moments of distances within these blobs (see (3.2)) and $\bar{\mathbf{X}}$ for the inter-blob junction points. Finally let $\bar{\mathscr{C}}_1 = \Gamma(\mathscr{C}_1, \bar{\mathbf{w}}, \bar{\mathbf{M}}, \bar{\mathbf{X}})$. Theorem 3.4 asserts that

$$\mathrm{scl}\left(\frac{(\sigma_2)^2}{\sigma_2 + \sum_{i\in[n]} x_i^2 u_i}, 1\right)\bar{\mathscr{C}}_1 \xrightarrow{\mathrm{w}} \mathscr{G}(2\tilde{\mathbf{e}}_{\gamma_1}, \tilde{\mathbf{e}}_{\gamma_1}, \mathscr{P}_1), \text{ as } n \to \infty, \tag{6.31}$$

Let us now prove this assertion. Recall that conditional on the weights and blobs in $\mathscr{C}_1$, we are in the setting of Theorem 6.4. To apply this theorem, we need to know that the regularity properties required by (6.5) and (6.12) hold as well as the scaling of the constant A_m when applied to $\mathscr{C}_1$; here $\mathbf{p} = \mathbf{p}^{(1)}$ as defined in (6.4). We start with an auxiliary result that plays the main role in relating the moments of the weights and distances in $\mathscr{C}_1$ with the moments of the entire sequence $\mathbf{x}$. Let $m := |\mathscr{C}_1|$ and $\{x_v : v \in \mathscr{C}_1\}$ be the set of vertex weights in $\mathscr{C}_1$. Also recall that $\sigma_r = \sum_i x_i^r$ denoted the moments of the complete weight sequence $\mathbf{x}$ used to construct the graph $\mathscr{G}(\mathbf{x}, q)$.

Proposition 6.7. *Under Assumptions* 2.1 *and* 3.3*, the weights and average inter-blob distances within* $\mathscr{C}_1$ *satisfy*

$$\frac{\sum_{v\in\mathscr{C}_1} x_v^2}{\sum_{v\in\mathscr{C}_1} x_v} \cdot \frac{\sigma_2}{\sigma_3} \xrightarrow{\mathrm{P}} 1, \ \textit{as } n \to \infty,$$

and

$$\frac{\sum_{v\in\mathscr{C}_1} x_v u_v}{\sum_{v\in\mathscr{C}_1} x_v} \cdot \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2 u_i} \xrightarrow{\mathrm{P}} 1, \ \textit{as } n \to \infty.$$

Proof: Recall the breadth-first exploration construction of $\mathscr{G}(\mathbf{x}, q)$ used by Aldous in [5], described in Section 6.2.1. The properties of this construction relevant for us are summarized as follows:

(a) The order in which vertices (blobs) are explored in this construction $(x_{v(i)} : i \in [n])$, is in the size-biased random order using the vertex weights $(x_i : i \in [n])$.
(b) Suppose the exploration of the maximal component $\mathscr{C}_1$ commences at time $m_L + 1$ and ends at m_R. Then the vertices in $\mathscr{C}_1$ are $\{v(i) : m_L + 1 \leqslant i \leqslant m_R\}$.
(c) Under assumptions 2.1, Aldous [5] shows that $\sum_{i=1}^{m_R} x_{v(i)}$ is tight.
(d) Finally Theorem 2.2 implies that $\sum_{i=m_L+1}^{m_R} x_{v(i)} \xrightarrow{\mathrm{w}} \gamma_1$ as $n \to \infty$, where as before γ_1 is the maximal excursion of $\bar{W}_\lambda(\cdot)$ from zero.

Note that in terms of this exploration process and the times of start and finish in the exploration of $\mathscr{C}_1$, Proposition 6.7 is equivalent to showing

$$\frac{\sum_{i=m_L+1}^{m_R} x_{v(i)}^2}{\sum_{i=m_L+1}^{m_R} x_{v(i)}} \cdot \frac{\sigma_2}{\sigma_3} \xrightarrow{\mathrm{P}} 1, \qquad \frac{\sum_{i=m_L+1}^{m_R} x_{v(i)} u_{v(i)}}{\sum_{i=m_L+1}^{m_R} x_{v(i)}} \cdot \frac{\sigma_2}{\sum_{i=1}^n x_i^2 u_i} \xrightarrow{\mathrm{P}} 1. \tag{6.32}$$

Thus here we are interested in the behavior of other functions of the vertices explored in a size-biased order, including squares of vertex weights and weighted average of the mean inter-blob distances. Such questions were studied in [15, Lemma 8.2] which we now quote. The sequences $\mathbf{x}$ and $\mathbf{u}$ could and in our situation do depend on n but we suppress this for ease of notation.

Lemma 6.8 ([15], Lemma 8.2)**.** *Let* $\mathbf{x} := (x_i : i \in [n])$ *be a sequence of vertex weights and let* $\mathbf{u} = (u_i \geqslant 0 : i \in [n])$ *be another function of the vertices. Let* $(v(i) \in [n] : i \in [n])$ *be a size-biased re-ordering using* $\mathbf{x}$. *Assume that for all n, the ratio* $c_n := \sum_{i\in[n]} x_i u_i / \sum_{i\in[n]} x_i > 0$. *Let* $x_{\max} := \max_{i\in[n]} x_i$ *and* $u_{\max} = \max_{i\in[n]} u_i$. *Let* $\ell = \ell(n) \in [n]$ *such that as* $n \to \infty$,

$$\frac{\ell x_{\max}}{\sum_{i\in[n]} x_i} \to 0, \ \frac{u_{\max}}{\ell c_n} \to 0. \tag{6.33}$$

Then we have, as $n\to\infty$,

$$\sup_{k\leqslant \ell}\left|\frac{\sum_{i=1}^{k} u_{v(i)}}{\ell c_n}-\frac{k}{l}\right| \xrightarrow{\mathrm{P}} 0.$$

Now we are ready to prove Proposition 6.7. We first show that the average weight of vertices in $\mathscr{C}_1$ satisfies

$$\frac{\sum_{i=m_L+1}^{m_R} x_{v(i)}}{m_R-m_L}\cdot\frac{\sigma_1}{\sigma_2}\xrightarrow{\mathrm{P}} 1, \text{ as } n\to\infty. \tag{6.34}$$

Fix $\eta>0$. Since $\sum_{i=1}^{m_R} x_{v(i)}$ is tight, there exists $T>0$ such that for all n,

$$\mathbb{P}\left(\sum_{i=1}^{m_R} x_{v(i)}\geqslant T\right)<\eta. \tag{6.35}$$

Let $m_0:=\sigma_1/\sigma_2$. We now apply Lemma 6.8 with $\ell=2Tm_0$ and $u_i\equiv x_i$. The assumptions in (6.33) are equivalent to $x_{\max}/\sigma_2\to 0$ (note that in this case $m_0c_n=1$), which directly follows from Assumption 2.1. By Lemma 6.8, there exists N_η such that when $n>N_\eta$,

$$\mathbb{P}\left(\sup_{k\leqslant 2Tm_0}\left|\sum_{i=1}^{k} x_{v(i)}-\frac{k}{m_0}\right|>\eta\right)<\eta. \tag{6.36}$$

On the set

$$\left\{\sup_{k\leqslant 2Tm_0}\left|\sum_{i=1}^{k} x_{v(i)}-k/m_0\right|\leq\eta\right\}\cap\left\{\sum_{i=1}^{m_R} x_{v(i)}\leqslant T\right\},$$

we have $m_L<m_R<2T$ (assuming $\eta<T$), and therefore $|\sum_{i=m_L+1}^{m_R} x_{v(i)}-(m_R-m_L)/m_0|<2\eta$. Since η can be arbitrarily small, using (6.35) and (6.36) we have

$$\left|\sum_{i=m_L+1}^{m_R} x_{v(i)}-\frac{m_R-m_L}{m_0}\right|\xrightarrow{\mathrm{P}} 0 \text{ as } n\to\infty.$$

Using Property (d) above on the properties of the size-biased exploration, we have $(m_R-m_L)/m_0\xrightarrow{\mathrm{w}}\gamma_1$. Thus multiplying the above expression by $m_0/(m_R-m_L)$, we have shown (6.34). Similarly, replacing $x_{v(i)}$ in (6.34) with $\bar{x}^2_{v(i)}$ and $x_{v(i)}\bar{u}_{v(i)}$ respectively and using Lemma 6.8, assuming

$$\frac{\sigma_2 x_{\max}^2}{\sigma_3}\to 0,\ \ \frac{\sigma_2 x_{\max} d_{\max}}{\sum_{i\in[n]} x_i^2 u_i}\to 0 \text{ as } n\to\infty, \tag{6.37}$$

then as $n\to\infty$,

$$\frac{\sum_{i=m_{L+1}}^{m_R} x_{v(i)}^2}{m_R-m_L}\cdot\frac{\sigma_1}{\sigma_3}\xrightarrow{\mathrm{P}} 1,\ \ \frac{\sum_{i=m_{L+1}}^{m_R} x_{v(i)}u_{v(i)}}{m_R-m_L}\cdot\frac{\sigma_1}{\sum_{i\in[n]} x_i^2 u_i}\xrightarrow{\mathrm{P}} 1. \tag{6.38}$$

Equation (6.37) follow from Assumptions 2.1 and 3.3. Combining (6.34) and (6.38) completes the proof of (6.32) and thus Proposition 6.7. ■

Proof of Theorem 3.4: Now we give the proof of Theorem 3.4. Without loss of generality we work with the maximal component, $\bar{\mathscr{C}}_1$. The same proof works for any fixed k. Let $m:=|\mathscr{C}_1|$. Denote the vertices $\{v: v\in\mathscr{C}_1\}=\{v(i): i\in[m]\}$ and relabel the vertices by $1,2,\dots,m$ so that $\mathscr{C}_1$ can be viewed as a graph on the vertex set $[m]$. Write $\bar{w}_i=x_{v(i)}$ and $\bar{u}_i=u_{v(i)}$

for $i \in [m]$. By Theorem 2.2 and Proposition 6.7, without loss of generality, we consider the probability space on which the following convergences hold almost surely: as $n \to \infty$,

$$\sum_{i\in[m]} \bar{w}_i \xrightarrow{\text{a.e.}} \gamma_1,\ \frac{\sum_{i\in[m]} \bar{w}_i^2}{\sum_{i\in[m]} \bar{w}_i} \cdot \frac{\sigma_2}{\sigma_3} \xrightarrow{\text{a.e.}} 1,\ \frac{\sum_{i\in[m]} \bar{u}_i}{\sum_{i\in[m]} \bar{w}_i} \cdot \frac{\sigma_2}{\sum_{i=1}^n x_i^2 u_i} \xrightarrow{\text{a.e.}} 1, \tag{6.39}$$

where $\gamma_1 = \gamma_1(\lambda)$ is a random variable as defined in Theorem 2.2. Let $\mathscr{F}^{(n)}$ be the sigma-field generated by the random partition $\{\mathcal{V}^{(i)} : i \geqslant 1\}$ as defined before Proposition 6.1. Then conditioned on $\mathscr{F}^{(n)}$, $\mathscr{C}_1$ has the law of $\mathbb{P}_{\text{con}}(\cdot; \mathbf{p}, a, [m])$ (see (6.3)) with

$$\mathbf{p} = (p_i : i \in [m]) := \left(\frac{\bar{w}_i}{\sum_{j\in[m]} \bar{w}_j} : i \in [m] \right),\ a := q \cdot (\sum_{i\in[m]} \bar{w}_i)^2. \tag{6.40}$$

By Assumption 2.1 we have $\sigma_3/\sigma_2^3 \to 1$ and $q\sigma_2 \to 1$ as $n \to \infty$. Combining this and (6.39) we have

$$a \cdot \sigma(\mathbf{p}) \sim \gamma_1^{3/2} \text{ as } n \to \infty,$$

where $\sim$ means the ratio of the left hand side to the right hand side converges to one almost surely. Thus assuming we can apply Theorem 6.4, we have

$$\operatorname{scl}\left(\frac{\sigma(\mathbf{p})}{1 + \sum_{i\in[m]} p_i \bar{u}_i}, \frac{1}{\gamma_1} \right) \bar{\mathscr{C}}_1 \xrightarrow{\text{w}} \mathscr{G}(2\tilde{\mathbf{e}}^{\gamma_1^{3/2}}, \gamma_1^{3/2} \tilde{\mathbf{e}}^{\gamma_1^{3/2}}, \mathscr{P}_1).$$

Again, by (6.39) and the fact $\sigma_3 \sim \sigma_2^3$, we have

$$\frac{\sigma(\mathbf{p})}{1 + \sum_{i\in[m]} p_i \bar{u}_i} \sim \frac{\sqrt{\sigma_3/\sigma_2\gamma_1}}{1 + (\sum_{i\in[n]} x_i^2 u_i)/\sigma_2} \sim \frac{\sigma_2^2}{\sigma_2 + \sum_{i\in[n]} x_i^2 u_i} \cdot \frac{1}{\gamma_1^{1/2}}.$$

Note that for any excursions h and g and Poisson point process $\mathscr{P}$, for $\alpha, \beta > 0$, we have

$$\operatorname{scl}(\alpha, \beta) \mathscr{G}(h, g, \mathscr{P}) \stackrel{d}{=} \mathscr{G}(\alpha h(\cdot/\beta), \frac{1}{\beta} g(\cdot/\beta), \mathscr{P}).$$

Therefore we have

$$\operatorname{scl}(\gamma_1^{1/2}, \gamma_1) \operatorname{scl}\left(\frac{\sigma(\mathbf{p})}{1 + \sum_{i\in[m]} p_i \bar{u}_i}, \frac{1}{\gamma_1} \right) \bar{\mathscr{C}}_1 \xrightarrow{\text{w}} \mathscr{G}(2\gamma_1^{1/2} \tilde{\mathbf{e}}^{\gamma_1^{3/2}}(\cdot/\gamma_1), \gamma_1^{1/2} \tilde{\mathbf{e}}^{\gamma_1^{3/2}}(\cdot/\gamma_1), \mathscr{P}_1).$$

By the Brownian scaling, we have $\gamma_1^{1/2} \tilde{\mathbf{e}}^{\gamma_1^{3/2}}(\cdot/\gamma_1) \stackrel{d}{=} \tilde{\mathbf{e}}_{\gamma_1}(\cdot)$. Combining this fact and the above convergence, we have proved (6.31).

Then we only need to verify the assumptions in Theorem 6.4. Not that by (6.39), the corresponding quantities in assumptions (6.5) and (6.12) satisfies

$$\sigma(\mathbf{p}) \sim \sqrt{\frac{\sigma_3}{\sigma_2\gamma_1}} \sim \frac{\sigma_2}{\gamma_1^{1/2}},\ A_m \sim \frac{\sum_{i=1}^n x_i^2 u_i}{\sigma_2}.$$

Thus the assumptions (6.5) and (6.12) follow from Assumptions 2.1, 3.3 and the fact that $\mathbb{P}(\gamma_1 \in (0, \infty)) = 1$. This completes the proof of Theorem 3.4. ■

7. PROOFS: SCALING LIMITS OF INHOMOGENEOUS RANDOM GRAPHS

This section contains the proof of all the results for the IRG model. Recall the kernel κ_n^- (4.2) for fixed $\delta \in (1/5, 1/6)$. There is a natural coupling between $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ and $\mathscr{G}_{\mathrm{IRG}}^{(n)}$ such that $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ is a subgraph of $\mathscr{G}_{\mathrm{IRG}}^{(n)}$. The main idea is to use the universality result, Theorem 3.4 where the blobs correspond to the connected components in $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$. These are the main steps in the proof:

(a) Recall from Section 4.1 that we introduced a model $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}(\kappa_n^-, [n])$ that was closely related to $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$. This model turns out to be technically easier to study an in particular prove that the configuration of components satisfy good properties (Theorem 4.4) that are required to apply Theorem 3.4. In Section 7.1, assuming Theorem 4.4, we show how to complete the proof of all the other results starting with the proof of Corollary 4.5 relating $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ to $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}(\kappa_n^-, [n])$. We will then use Corollary 4.5 to prove Theorem 4.3 on the continuum limit of the metric structure and 4.2 on the actual sizes of connected components.
(b) In Section 7.2, we will build all the technical machinery to prove Theorem 4.4 through a detailed study of the associated multitype branching process.
(c) Finally in Section 7.3, we use this technical machinery to complete the proof of Theorem 4.4.

7.1. **Scaling limit for the IRG model.** This section contains the proof of all the other results assuming Theorem 4.4.

Proof of Corollary 4.5: Define $p_{ij}^- = p_{ij}^{(n),-} = 1 - \exp(-\kappa_n^-(x_i, x_j)/n)$. Then note that $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ and $\mathscr{G}_{\mathrm{IRG}^{(n),\star}}$ are both models of random graphs where we place edges independently between different vertices $i, j \in [n]$, using p_{ij}^- for $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ and $p_{ij}^\star$ for $\mathscr{G}_{\mathrm{IRG}}^{(n),\star}$ where

$$p_{ij}^- := 1 - \exp\left(-\frac{\kappa_n^-(x_i, y_i)}{n}\right), \qquad p_{ij}^\star := 1 \wedge \left(\frac{\kappa_n^-(x_i, x_j)}{n}\right), \tag{7.1}$$

where as before $x_i \in [K]$ denotes the type of vertex K. Thus we have

$$\sum_{1\le i<j\le n} \frac{\left(p_{ij}^- - p_{ij}^\star\right)^2}{p_{ij}^\star} \le \sum_{1\le i<j\le n} {p_{ij}^\star}^3 = O\left(\frac{1}{n}\right). \tag{7.2}$$

Thus, the claim follows from Theorem 4.4 and the asymptotic equivalence between the two random graph models $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ and $\mathscr{G}_{\mathrm{IRG}^{(n),\star}}$ under 7.2 using [37, Corollary 2.12]. ■

Now note that there is a natural coupling between $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ and $\mathscr{G}_{\mathrm{IRG}}^{(n)}$ such that $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ is a subgraph of $\mathscr{G}_{\mathrm{IRG}}$. Furthermore, conditioned on $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$, $\mathscr{G}_{\mathrm{IRG}}$ can be obtained by putting edges between each pair of vertices independently with probability $1-\exp(1/n^{1+\delta})$. Therefore, given two distinct components $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$ in $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$, the number of edges added between them in $\mathscr{G}_{\mathrm{IRG}}$, say N_{ij}, is distributed as $\mathrm{Binomial}(|\mathscr{C}_i^-||\mathscr{C}_j^-|, 1-\exp[n^{-(1+\delta)}])$. In addition, given N_{ij}, the endpoints of these edges that link the two components are chosen uniformly among the vertices of $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$ respectively. Also, for any component $\mathscr{C}_i^-$ in

$\mathscr{G}_{\rm IRG}^{(n),-}$, Binomial$\left(\binom{|\mathscr{C}_i^-|}{2}, 1-\exp[n^{-(1+\delta)}]\right)$ many edges are added between the vertices of $\mathscr{C}_i^-$ in $\mathscr{G}_{\rm IRG}$.

Our plan is to apply Theorem 3.4 to the components of $\mathscr{G}_{\rm IRG}^{(n),-}$ where the blobs consist of the connected components of $\mathscr{G}_{\rm IRG}^{(n),-}$ with the usual graph distance and the measure μ_i on each blob $\mathscr{C}_i^-$ is just the uniform measure. In the setup of Theorem 3.4, however, we place **one** edge between two distinct components $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$ in $\mathscr{G}_{\rm IRG}^{(n),-}$ with probability $1-\exp(-|\mathscr{C}_i^-||\mathscr{C}_j^-|/n^{1+\delta})$. If such an edge is added, its endpoints are chosen uniformly and independently from the vertices of $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$. Compare this with the Binomial distribution of edges between blobs in the original model. Let d' be the resulting graph distance. Let $\bar{\mathscr{C}}_i(\mathscr{G}_{\rm IRG}^{(n)})$ be the i-th largest component of $\mathscr{G}_{\rm IRG}^{(n)}$ endowed with the metric d'. We will assume that $\bar{\mathscr{C}}_i(\mathscr{G}_{\rm IRG}^{(n)})$ and $\mathscr{C}_i(\mathscr{G}_{\rm IRG}^{(n)})$ are coupled in a way so that $d \leqslant d'$. In order to apply Theorem 3.4, we need to show that $\bar{\mathscr{C}}_i(\mathscr{G}_{\rm IRG}^{(n)})$ and $\mathscr{C}_i(\mathscr{G}_{\rm IRG}^{(n)})$ are "close" with respect to Gromov-Hausdorff metric. The following lemma serves this purpose.

Lemma 7.1. *For each $k \geqslant 1$,*

$$n^{-1/3} d_{\rm GHP}\left(\mathscr{C}_k(\mathscr{G}_{\rm IRG}^{(n)}), \bar{\mathscr{C}}_k(\mathscr{G}_{\rm IRG}^{(n)})\right) \xrightarrow{\rm P} 0.$$

We will first prove Theorems 4.3 and 4.2 assuming Lemma 7.1 and then give a proof of this lemma.

Proof of Theorem 4.3 and Theorem 4.2: Define,

$$x_i = \frac{\beta^{1/3}|\mathscr{C}_i^-|}{n^{2/3}},\ q = \frac{n^{1/3-\delta}}{\beta^{2/3}},\ M_i = \mathrm{scl}(1, 1/|\mathscr{C}_i^-|)\mathscr{C}_i^-. \tag{7.3}$$

Writing $\sigma_k = \sum_i x_i^k$ for $k \geqslant 1$, we have

$$\sigma_2 = \frac{\beta^{2/3}\bar{s}_2}{n^{1/3}},\ \sigma_3 = \frac{\beta \bar{s}_3}{n},\ x_{\max} = \frac{\beta^{1/3}|\mathscr{C}_1^-|}{n^{2/3}},\ \text{and}\ x_{\min} \geqslant \frac{\beta^{1/3}}{n^{2/3}}.$$

By Corollary 4.5 (more precisely the analogue of (4.5) for $\mathscr{G}_{\rm IRG}^{(n),-}$), Assumption 2.1 holds with $\lambda = \zeta\beta^{-2/3}$. Theorem 4.3 now follows from Theorem 2.2.

In view of Lemma 7.1, it is enough to check that the conditions in Assumption 3.3 hold. Now, by definition, $u_\ell = \sum_{i,j\in\mathscr{C}_\ell^-} d^-(i,j)/|\mathscr{C}_\ell^-|^2$ where d^- denotes the graph distance in $\mathscr{G}_{\rm IRG}^{(n),-}$. Therefore,

$$\sum_{\ell\geq 1} x_\ell^2 u_\ell = \frac{\beta^{2/3}}{n^{4/3}} \sum_{\ell\geq 1}\sum_{i,j\in\mathscr{C}_\ell^-} d^-(i,j) = \frac{\beta^{2/3}\bar{\mathscr{D}}}{n^{1/3}}.$$

Corollary 4.5 together with the above observations ensures that the conditions in Assumption 3.3 hold and further,

$$\frac{\sigma_2^2}{\sigma_2 + \sum_{\ell\geq 1} x_\ell^2 u_\ell} \sim \frac{\beta^{2/3}}{\alpha n^{1/3}}.$$

This completes the proof of Theorem 4.2. ■

Proof of Lemma 7.1: Recall that $\mathscr{C}_1^-, \mathscr{C}_2^-, \ldots$ are the components of $\mathscr{G}_{\rm IRG}^{(n),-}$ arranged in decreasing order of size. For $i \neq j$, let N_{ij} be the number of edges between $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$ in $\mathscr{G}_{\rm IRG}^{(n)}$. Let N_{ii} be the number of edges added between vertices of $\mathscr{C}_i^-$ while going from

$\mathcal{G}_{\mathrm{IRG}}^{(n),-}$ to $\mathcal{G}_{\mathrm{IRG}}^{(n)}$. Let $\mathscr{F}_-$ denote the σ-field generated by $\mathcal{G}_{\mathrm{IRG}}^{(n),-}$. Define $X_n = \sum_{i\neq j} \mathbb{1}\{N_{ij} \geqslant 2\} + \sum_i \mathbb{1}\{N_{ii} \geqslant 1\}$. Then, for any $k \geqslant 1$ and $x, y \in \mathscr{C}_k(\mathcal{G}_{\mathrm{IRG}}^{(n)})$,

$$|d(x,y) - d'(x,y)| \leqslant 2X_n \mathscr{D}_{\max}^- \text{ which implies } d_{\mathrm{GH}}\left(\mathscr{C}_k(\mathcal{G}_{\mathrm{IRG}}^{(n)}), \bar{\mathscr{C}}_k(\mathcal{G}_{\mathrm{IRG}}^{(n)})\right) \leqslant X_n \mathscr{D}_{\max}^-.$$

From Corollary 4.5, $n^{-1/3}\mathscr{D}_{\max}^- \xrightarrow{\mathrm{P}} 0$. So it is enough to show that X_n is tight. To this end, note that

$$\mathbb{P}(N_{ij} \geqslant 2|\mathscr{F}_-) \leqslant |\mathscr{C}_i^-|^2 |\mathscr{C}_j^-|^2 / n^{2+2\delta} \text{ and } \mathbb{P}(N_{ii} \geqslant 1|\mathscr{F}_-) \leqslant |\mathscr{C}_i^-|^2 / n^{1+\delta}.$$

Hence, $\mathbb{E}[X_n|\mathscr{F}_-] \leqslant \bar{s}_2^2/n^{2\delta} + \bar{s}_2/n^{\delta}$. Now, an application of Corollary 4.5 will show that $\bar{s}_2/n^{\delta} \xrightarrow{\mathrm{P}} 1$. This proves tightness of X_n. Hence, we have shown that $d_{\mathrm{GH}}\left(\mathscr{C}_k(\mathcal{G}_{\mathrm{IRG}}^{(n)}), \bar{\mathscr{C}}_k(\mathcal{G}_{\mathrm{IRG}}^{(n)})\right) \to 0$ for fixed $k \geqslant 1$. Now the corresponding statement for d_{GHP} follows trivially. ■

7.2. **Branching process approximation.** As has been observed in [21], one key tool in study the IRG model is a closely related multitype branching process. The aim of this section, is to introduce this object and study its properties in the barely subcritical regime. For any graph G and a vertex $v \in \mathcal{V}(G)$, define $\mathscr{C}(v;G)$ to be the component in G that contains the vertex v. Denote d for the graph distance on G. Define

$$\mathscr{D}(v;G) := \sum_{i \in \mathscr{C}(v;G)} d(v,i).$$

Recall the definition of $\mathcal{G}^\star = \mathcal{G}_{\mathrm{IRG}}^{(n),\star}$. Let

$$\mathscr{C}(i) = \mathscr{C}(i;\mathcal{G}_{\mathrm{IRG}}^\star) \text{ and } \mathscr{D}(i) = \mathscr{D}(i;\mathcal{G}_{\mathrm{IRG}}^\star).$$

Let v and u be two uniformly chosen vertices from $[n]$, independent of each other and of $\mathcal{G}_{\mathrm{IRG}}^\star$. Suppose v, u and $\mathcal{G}_{\mathrm{IRG}}^\star$ are defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\mathscr{F}_\star \subset \mathscr{F}$ be the σ-field generated by $\mathcal{G}_{\mathrm{IRG}}^\star$. Then we have

$$\bar{s}_{k+1}^\star = \mathbb{E}[|\mathscr{C}(v)|^k|\mathscr{F}_\star],\ \bar{\mathscr{D}}^\star = \mathbb{E}[\mathscr{D}(v)|\mathscr{F}_\star], \text{ for } k = 1,2,\ldots. \tag{7.4}$$

Furthermore, we have

$$(\bar{s}_{k+1}^\star)^2 = \mathbb{E}\left[|\mathscr{C}(v)|^k|\mathscr{C}(u)|^k\middle|\mathscr{F}_\star\right],\ (\bar{\mathscr{D}}^\star)^2 = \mathbb{E}\left[\mathscr{D}(v)\mathscr{D}(u)\middle|\mathscr{F}_\star\right], \text{ for } k = 1,2,\ldots. \tag{7.5}$$

The goal of this section is to prove the following result.

Proposition 7.2. *We have,*

$$\lim_n \frac{\mathbb{E}[\bar{s}_2^\star] - n^\delta}{n^{2\delta-1/3}} = \zeta,\ \lim_n \frac{\mathbb{E}[\bar{s}_3^\star]}{n^{3\delta}} = \beta, \text{ and } \lim_n \frac{\mathbb{E}[\bar{\mathscr{D}}^\star]}{n^{2\delta}} = \alpha. \tag{7.6}$$

In addition, there exist positive constants $C_1 = C_1(k,\kappa,\mu), C_2 = C_2(\kappa,\mu)$, and some positive integer n_0 such that for all $n \geqslant n_0$,

$$\mathrm{Var}(\bar{s}_{k+1}^\star) \leq C_1 n^{(4k+1)\delta-1} \text{ and} \tag{7.7}$$

$$\mathrm{Var}(\bar{\mathscr{D}}^\star) \leq C_2 n^{(8\delta-1)}. \tag{7.8}$$

The cut-off n_0 depends only on the sequences μ_n, κ_n.

It turns out that bounding $\mathrm{Var}(\bar{\mathscr{D}}^\star)$ is the most difficult part. We will prove the other asymptotics first and leave this to the end of this section. We start with the following lemma.

Lemma 7.3. *For all $k \geq 1$ and all n, we have*

$$\mathrm{Var}(\bar{s}^\star_{k+1}) \leq \frac{1}{n}\mathbb{E}[|\mathscr{C}(v)|^{2k+1}] \text{ and} \tag{7.9}$$

$$\mathrm{Cov}(\bar{s}^\star_{k+1}, \bar{\mathscr{D}}^\star) \leq \frac{1}{n}\mathbb{E}[|\mathscr{C}(v)|^{k+1}\mathscr{D}(v)]. \tag{7.10}$$

Proof: By (7.5), we have $\mathbb{E}[(\bar{s}^\star_{k+1})^2] = \mathbb{E}[|\mathscr{C}(v)|^k|\mathscr{C}(u)|^k]$ and $\mathbb{E}[(\bar{\mathscr{D}}^\star)^2] = \mathbb{E}[\mathscr{D}_k(v)\mathscr{D}_k(u)]$. Let $\mathscr{V}(\mathscr{C}(v))$ denote the vertex set of $\mathscr{C}(v)$. Write $\mathscr{G}'$ for the graph induced by $\mathscr{G}^\star_{\mathrm{IRG}}$ on the vertex set $[n]\setminus\mathscr{V}(\mathscr{C}(v))$, we have

$$\begin{aligned}\mathbb{E}\left[|\mathscr{C}(u)|^k \,\middle|\, \{\mathscr{C}(v), v\}\right] &= \frac{|\mathscr{C}(v)|}{n}\cdot|\mathscr{C}(v)|^k + \frac{1}{n}\mathbb{E}\left[\sum_{\mathscr{C}\subset\mathscr{G}'}|\mathscr{C}|^{k+1}\,\middle|\,\{\mathscr{C}(v), v\}\right],\\ &\leq \frac{1}{n}|\mathscr{C}(v)|^{k+1} + \mathbb{E}[\bar{s}_{k+1}]\\ &= \frac{1}{n}|\mathscr{C}(v)|^{k+1} + \mathbb{E}[|\mathscr{C}(v)|^k].\end{aligned}$$

where $\sum_{\mathscr{C}\subset\mathscr{G}'}$ denotes sum over all components in $\mathscr{G}'$. Therefore,

$$\begin{aligned}\mathbb{E}\left[(\bar{s}^\star_{k+1})^2\right] &= \mathbb{E}\left[|\mathscr{C}(u)|^k|\mathscr{C}(v)|^k\right]\\ &= \mathbb{E}\left[|\mathscr{C}(v)|^k\,\mathbb{E}\left(|\mathscr{C}(u)|^k \,\middle|\, \{\mathscr{C}(v), v\}\right)\right] \leq \frac{1}{n}\mathbb{E}[\mathscr{C}^{2k+1}(v)] + (\mathbb{E}[|\mathscr{C}(v)|^k])^2,\end{aligned}$$

which gives the bound on $\mathrm{Var}(\bar{s}^\star_{k+1})$. Similarly

$$\begin{aligned}\mathbb{E}\left[\bar{s}^\star_{k+1}\bar{\mathscr{D}}^\star\right] &= \mathbb{E}\left[|\mathscr{C}(u)|^k\mathscr{D}(v)\right] = \mathbb{E}\left[\mathscr{D}(v)\,\mathbb{E}\left(|\mathscr{C}(u)|^k \,\middle|\, \{\mathscr{C}(v), v\}\right)\right]\\ &\leq \frac{1}{n}\mathbb{E}\left[\mathscr{D}(v)|\mathscr{C}(v)|^{k+1}\right] + \mathbb{E}\left[|\mathscr{C}(v)|^k\right]\mathbb{E}[\mathscr{D}(v)].\end{aligned}$$

This completes the proof of Lemma 7.3. ■

Recall the definition of κ^-_n from (4.2). We will now consider a K-type branching process in which each particle of type $j \in [K]$ in k-th generation has $\mathrm{Binomial}\big(n\mu_n(i),\, \kappa^-_n(i,j)/n\big)$ number of type i children in the next generation for $i \in [K]$ and the number of children of different types are independent. Suppose in the 0-th generation, there is only one particle and its type is $x \in [K]$. Define $G_k(x) = G_k(x; n, \mu_n, \kappa^-_n)$ to be the total number of particles in the k-th generation of such a branching process, $k = 0, 1, 2, \ldots$. Then $G_0(x) \equiv 1$. Define

$$T_k(x) = T_k(x; n, \mu_n, \kappa^-_n) := \sum_{\ell=0}^{\infty} \ell^k G_\ell(x), \text{ for } k \geq 0.$$

Denote by $T_0(\mu_n)$ and $T_k(\mu_n)$, the corresponding quantities for the branching process when the type of the first particle follows the distribution μ_n. We define $T_0(x, y)$ to be the total number of type-y particles in the branching process starting from a particle of type x.

Given a random vector $\boldsymbol{w} = (w_1, \ldots, w_K)^t$ with $w_y \geqslant 0$, let

$$G_k(\boldsymbol{w}) = \sum_{y \in [K]} \sum_{i=1}^{w_y} G_k^{(i)}(y) \text{ and let } H_k(\boldsymbol{w}) = \sum_{s=1}^{k} G_s(\boldsymbol{w}) \text{ for } k \geqslant 1 \tag{7.11}$$

where $\left\{G_k^{(i)}(y) : k \geqslant 0\right\}$ has the same distribution as $\{G_k(y) : k \geqslant 0\}$ for $1 \leqslant i \leqslant w_y$ and the collections of random variables $\left\{G_k^{(i)}(y) : k \geqslant 0\right\}_{y \in [K], 1 \leqslant i \leqslant w_y}$ are independent conditional on $\boldsymbol{w}$. Analogously define

$$T_0(\boldsymbol{w}) = \sum_{y \in [K]} \sum_{i=1}^{w_y} T_0^{(i)}(y) \tag{7.12}$$

where $T_0^{(i)}(y)$ is distributed as $T_0(y)$ and the random variables $T_0^{(i)}(y),\ y \in [K], 1 \leqslant i \leqslant w_y$ are independent conditional on $\boldsymbol{w}$.

In the following lemma, we will study certain asymptotic properties of this K-type branching process.

Lemma 7.4. (a) **Growth rates for $T_k(\mu_n)$:** *For any $r, k \geqslant 0$, there exists a constant $C_1 = C_1(k, r; \kappa, \mu)$ such that*

$$\sup_{n \geqslant n_0} \mathbb{E}[T_0^r(\mu_n) T_k(\mu_n)] \leqslant C_1 n^{(2r+k+1)\delta} \tag{7.13}$$

where n_0 depends only on k, r and the sequences μ_n and κ_n. In particular, for any $r \geqslant 1$ and $x \in [K]$, there exists a constant $C_2 = C_2(r; \kappa, \mu)$ such that

$$\sup_{n \geqslant n_1} \frac{\mathbb{E}[T_0^r(x)]}{n^{(2r-1)\delta}} \leqslant C_2 \tag{7.14}$$

for some n_1 depending only on r and the sequences μ_n and κ_n. Further, for any $J > 0$ and integers $r, k \geqslant 0$, there exists a constant $C_3 = C_3(J, r, k; \kappa, \mu)$ such that

$$\sup_{n \geqslant n_2} \mathbb{E}\left[T_0^r(x; n, \mu_n', \kappa_n^-) \times T_k(x; n, \mu_n', \kappa_n^-)\right] \leqslant C_3 n^{(2r+k+1)\delta} \tag{7.15}$$

for any $x \in [K]$ and any sequence of measures μ_n' on $[K]$ satisfying $\sum_{x \in [K]} |\mu_n'(x) - \mu_n(x)| \leqslant J \log n / n^{3/5}$ for all n. The cut-off n_2 depends only on J, k, r and the sequences μ_n and κ_n.

(b) **Exact asymptotics for $T_0(\mu_n)$ and $T_0^2(\mu_n)$:** *We have,*

$$\lim_n \frac{\mathbb{E}[T_0(\mu_n)] - n^\delta}{n^{2\delta - 1/3}} = \zeta, \text{ and } \lim_n \frac{\mathbb{E}[T_0^2(\mu_n)]}{n^{3\delta}} = \beta. \tag{7.16}$$

(c) **Exact asymptotics for $T_1(\mu_n)$:** *We have,*

$$\lim_n \frac{\mathbb{E}[T_1(\mu_n)]}{n^{2\delta}} = \alpha. \tag{7.17}$$

(d) **Tail bound on height and component size:** *For $x \in [K]$, let* $\mathrm{ht}(x)$ *denote the the height of the K-type branching process started from one initial particle of type x. Then, there* *exist constants C_1, C_2, C_3 depending only on κ an μ such that for all $x \in [K]$ and $m \geqslant 1$, we have,*

$$\mathbb{P}(\mathrm{ht}(x) \geqslant m) \leqslant C_1 \exp\left(-C_2 m / n^{\delta}\right) \text{ and} \tag{7.18}$$

$$\mathbb{P}\left(T_0(x) \geqslant m\right) \leq 2 \exp\left(-C_3 m / n^{2\delta}\right) \tag{7.19}$$

for $n \geqslant n_3$ where the cut-off n_3 depends only on the sequences $\{\mu_n\}$ and $\{\kappa_n\}$.

While proving (7.8), we will need an analogue of (7.13) for the setup where the empirical distribution of types on $[K]$ may be different from μ_n but is sufficiently concentrated around μ_n. This is the only part where we will use (7.15). However, to avoid introducing additional notation, we will only prove (7.13). The proof for any sequence μ_n' as in the statement of Lemma 7.4 follows the exact same steps. We will continue to write $G_k(x), T_k(x)$ etc. without any ambiguity as the underlying empirical measure will always be μ_n.

We will need the following elementary lemma.

Lemma 7.5. *Let A_1, A_2 and A_3 be square matrices of order K. Assume that the entries of A_1 are positive. Let $\mathbf{w}_\ell$ and $\mathbf{w}_r$ be left and right eigenvectors of A_1 corresponding to $\rho(A_1)$ subject to $\mathbf{w}_\ell \cdot \mathbf{w}_r = 1$. Then,*

$$\lim_{\substack{x \to 0 \\ y \to 0}} \frac{\rho(A_1 + xA_2 + yA_3) - \rho(A_1 + xA_2)}{y} = \mathbf{w}_\ell^t A_3 \mathbf{w}_r.$$

Proof: Since the entries of A_1 are positive, $\rho(A_1)$ is a simple eigenvalue of A_1. An application of implicit function theorem shows that $\rho(x, y) := \rho(A_1 + xA_2 + yA_3)$ is a C^∞ function of x, y in a small neighborhood of $(0,0)$. So the required limit is simply $\partial \rho(0,0)/\partial y$.

For some small $\epsilon > 0$, let $\mathbf{w}_\ell(y)$ $\left(\text{resp. } \mathbf{w}_r(y)\right) : [-\epsilon, \epsilon] \to \mathbb{R}^K$ be a C^∞ function such that $\mathbf{w}_\ell(0) = \mathbf{w}_\ell$ (resp. $\mathbf{w}_r(y) = \mathbf{w}_r$) and for each $y \in [-\epsilon, \epsilon]$, $\mathbf{w}_\ell(y)$ (resp. $\mathbf{w}_r(y)$) is a left (resp. right) eigenvector of $A_1 + yA_3$ corresponding to $\rho(0, y)$. We further assume that $\mathbf{w}_\ell(y) \cdot \mathbf{w}_r(y) = 1$ for $y \in [-\epsilon, \epsilon]$. Hence, we have

$$\left(\frac{\partial}{\partial y} \mathbf{w}_\ell(y)\right)^t \mathbf{w}_r(y) + \mathbf{w}_\ell(y)^t \left(\frac{\partial}{\partial y} \mathbf{w}_r(y)\right) = 0 \text{ for } y \in [-\epsilon, \epsilon]. \tag{7.20}$$

Note that, $\mathbf{w}_\ell(y)^t (A_1 + yA_3) \mathbf{w}_r(y) = \rho(0, y)$. Hence,

$$\begin{aligned} y \mathbf{w}_\ell(y)^t A_3 \mathbf{w}_r(y) &= \rho(0, y) - \mathbf{w}_\ell(y)^t A_1 \mathbf{w}_r(y) \\ &= \rho(0, y) - \rho(0,0) + \mathbf{w}_\ell^t A_1 \mathbf{w}_r - \mathbf{w}_\ell(y)^t A_1 \mathbf{w}_r(y). \end{aligned}$$

The result follows upon dividing by y and taking limits in the last equation and using (7.20). ■

Proof of Lemma 7.4: For $x, y \in [K]$, define

$$m_{xy}^{(n)} = \mu_n(y) \kappa_n^-(x, y) \text{ and let } M_n = \left(m_{xy}^{(n)}\right)_{K \times K}.$$

Note that for large n, M_n is a matrix with positive entries. Let $\rho_n = \rho(M_n)$ and let $\mathbf{u}_n$ and $\mathbf{v}_n$ be the associated right and left eigenvectors of M_n respectively subject to $\mathbf{v}_n^t \mathbf{u}_n = \mathbf{u}_n^t \mathbf{1} = 1$.

Now,

$$\mathbb{E}\,T_0(x) = \sum_{\ell=0}^{\infty} \mathbb{E}\,G_\ell(x) = \sum_{\ell=0}^{\infty} \mathbf{e}_x^t M_n^\ell \mathbf{1} \tag{7.21}$$

where $\mathbf{e}_x$ denotes the unit vector with one at the x-th coordinate. From Frobenius theorem for positive matrices (see e.g. [8]), it follows that there exists $c > 0$ and $0 < r < 1$ such that

$$M = \mathbf{u}\mathbf{v}^t + R \text{ where } R^t\mathbf{v} = R\mathbf{u} = \mathbf{0} \text{ and } \max_{i,j}\left|R^\ell(i,j)\right| \leqslant cr^\ell$$

for every $\ell \geqslant 1$. A similar decomposition holds for M_n:

$$M_n = \rho_n \mathbf{u}_n \mathbf{v}_n^t + R_n \text{ where } R_n^t \mathbf{v}_n = R_n \mathbf{u}_n = \mathbf{0}. \tag{7.22}$$

Since $\max_{i,j} |m_{ij}^{(n)} - m_{ij}| = O(n^{-\delta})$ and similar statements are true for $\|\mathbf{u}_n - \mathbf{u}\|$, $\|\mathbf{v}_n - \mathbf{v}\|$ and $(1-\rho_n)$, it follows that $\max_{i,j} |(R_n - R)(i,j)| = O(n^{-\delta})$. Hence, there exist positive constants c_1, c_2 such that

$$\max_{i,j}\left|R_n^\ell(i,j)\right| \leqslant c_1(r + c_2 n^{-\delta})^\ell \text{ for } \ell \geqslant 1. \tag{7.23}$$

Using this decomposition, (7.21) yields

$$\lim_n \frac{\mathbb{E}\,T_0(\mu_n) - n^\delta}{n^{2\delta-1/3}} = \lim_n \frac{\big((\boldsymbol{\mu}_n^t \mathbf{u}_n)(\mathbf{v}_n^t \mathbf{1})/(1-\rho_n)\big) - n^\delta}{n^{2\delta-1/3}}. \tag{7.24}$$

We can write

$$\begin{aligned} M_n = \kappa_n^- D_n &= \kappa D + \kappa(D_n - D) + (\kappa_n - \kappa)D_n + (\kappa_n^- - \kappa_n)D_n \\ &= M + \frac{\kappa B}{n^{1/3}} + \frac{AD}{n^{1/3}} - \frac{\mathbf{1}\boldsymbol{\mu}^t}{n^\delta} + o(n^{-1/3}). \end{aligned} \tag{7.25}$$

Note that

$$n^{1/3}\rho_n = n^{1/3}\rho\left(M + \frac{\kappa B}{n^{1/3}} + \frac{AD}{n^{1/3}} - \frac{\mathbf{1}\boldsymbol{\mu}^t}{n^\delta}\right) + o(1). \tag{7.26}$$

Lemma 7.5 coupled with (7.26) and (7.25) gives

$$\lim_n n^\delta(1-\rho_n) = \lim_n n^\delta\left(\rho(M) - \rho(M_n)\right) = (\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1}). \tag{7.27}$$

Using (7.27) together with the facts $\|\boldsymbol{\mu}_n - \boldsymbol{\mu}\| = O(n^{-1/3})$, $\|\mathbf{u}_n - \mathbf{u}\| = O(n^{-\delta})$, $\|\mathbf{v}_n - \mathbf{v}\| = O(n^{-\delta})$ and $\delta > 1/6$, we conclude from (7.24) that

$$\begin{aligned} \lim_n \frac{\mathbb{E}\,T_0(\mu_n) - n^\delta}{n^{2\delta-1/3}} &= \lim_n \frac{\big((\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})/(1-\rho_n)\big) - n^\delta}{n^{2\delta-1/3}} = \lim_n \left[n^{1/3-\delta} - \frac{n^{1/3}(1-\rho_n)}{(\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})}\right] \\ &= \lim_n \left[n^{1/3-\delta} - \frac{n^{1/3}\left(1 - \rho\left(M + \frac{\kappa B}{n^{1/3}} + \frac{AD}{n^{1/3}} - \frac{\mathbf{1}\boldsymbol{\mu}^t}{n^\delta}\right)\right)}{(\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})}\right], \end{aligned} \tag{7.28}$$

the last inequality being a consequence of (7.26).

Since $f(x) := x^{-1}\left(1-\rho(M-x\mathbf{1}\boldsymbol{\mu}^t)\right)$ is C^∞ on a compact interval around zero and $f(0) = (\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})$, $|f(0)-f(x)| = O(|x|)$ on an interval around zero. We thus have,

$$\begin{aligned} n^{1/3-\delta} &= \frac{n^{1/3-\delta}f(n^{-\delta})}{f(0)} + \frac{n^{1/3-\delta}\left[f(0)-f(n^{-\delta})\right]}{f(0)} \\ &= \frac{n^{1/3}\left(1-\rho(M-n^{-\delta}\mathbf{1}\boldsymbol{\mu}^t)\right)}{(\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})} + O(n^{1/3-2\delta}). \end{aligned}$$

Plugging this in (7.28) and using Lemma 7.5, we get

$$\lim_n \frac{\mathbb{E}\, T_0(\mu_n) - n^\delta}{n^{2\delta-1/3}} = \left(\mathbf{v}^t(\kappa B + AD)\mathbf{u}\right) / \left((\boldsymbol{\mu}^t\mathbf{u})(\mathbf{v}^t\mathbf{1})\right).$$

This proves the first part of (7.16). Here, we make note of the following fact:

$$\lim_n \frac{\mathbb{E}\, T_0(x)}{n^\delta} = \frac{u_x}{\boldsymbol{\mu}^t\mathbf{u}} \tag{7.29}$$

which is a direct consequence of (7.21), (7.22), (7.23), (7.27) and the facts that $\mathbf{u}_n \to \mathbf{u}$ and $\mathbf{v}_n \to \mathbf{v}$. We will need this result later.

To get the other part of (7.16), recall that for a random variable $\mathbf{Y} = (Y_1,\ldots,Y_r)^t$, the p-th order cumulants are given by

$$\begin{aligned} \mathrm{cumu}_{\mathbf{Y}}(\ell_1,\ldots,\ell_p) &= \mathrm{cumu}(Y_{\ell_1},\ldots,Y_{\ell_p}) \\ &:= \sum_{q=1}^p \sum\nolimits_1 (-1)^{q-1}(q-1)! \prod_{i=1}^q \mathbb{E}\left(\prod_{j\in I_i} Y_{\ell_j}\right) \end{aligned} \tag{7.30}$$

where $1 \leqslant \ell_i \leqslant r$ and $\sum_1$ denotes the sum over all partitions of $I = \{1,\ldots,p\}$ into q subsets $I_1,\ldots,I_q$. Moments of $\mathbf{Y}$ can be expressed in terms of the cumulants as follows:

$$\mathbb{E}\left[\prod_{i=1}^p Y_{\ell_i}\right] = \sum_{q=1}^p \sum\nolimits_1 \prod_{i=1}^q \mathrm{cumu}_{\mathbf{Y}}\left(\{\ell_j\}_{j\in I_i}\right), \tag{7.31}$$

where $\sum_1$ has the same meaning as in (7.30).

For $x \in [K]$, let $\{Z(x,y) : y \in [K]\}$ be independent random variables having Binomial$(n\mu_n(y), \kappa_n^-(x,y)/n)$ distribution and let $a_x(y_1,\ldots,y_q) = \mathrm{cumu}(Z(x,y_1),\ldots,Z(x,y_q))$ where $y_1,\ldots,y_q \in [K]$. Recall the definition of $T_0(x,y)$ from right before Lemma 7.4. Then it follows from (13) of [46] that

$$\begin{aligned} \mathrm{cumu}(T_0(x,y_1),\ldots,T_0(x,y_p)) &= \sum_{y\in[K]} m_{x,y}^{(n)}\,\mathrm{cumu}(T_0(y,y_1),\ldots,T_0(y,y_p)) \\ &\quad + \sum_{q=2}^p \sum\nolimits_1 \sum_{k_1,\ldots,k_q} a_x(k_1,\ldots,k_q) \prod_{m=1}^q \mathrm{cumu}\left(\{T_0(k_m,y_j)\}_{j\in I_m}\right) \end{aligned} \tag{7.32}$$

where $\sum_1$ is sum over all partitions of $I = \{1,\ldots,p\}$ into q subsets $I_1,\ldots,I_q$. For $p=2$, (7.32) reduces to

$$\begin{aligned} \mathrm{Cov}\left(T_0(x,y_1),T_0(x,y_2)\right) &= \sum_{u\in[K]} m_{xu}^{(n)}\,\mathrm{Cov}(T_0(u,y_1),T_0(u,y_2)) \\ &\quad + \sum_{u\in[K]} \mathrm{Var}(Z(x,u))(\mathbb{E}\, T_0(u,y_1))(\mathbb{E}\, T_0(u,y_2)). \end{aligned}$$

Summing both sides over all $y_1, y_2 \in [K]$ and using the relations $\mathrm{Var}(Z(x,u)) = m_{xu}^{(n)} + O(n^{-1})$ and $\max_{y\in[K]} \mathbb{E}\, T_0(y) = O(n^{\delta})$, we get

$$\begin{aligned}\mathrm{Var}(T_0(x)) &= \sum_{u\in[K]} m_{xu}^{(n)}\,\mathrm{Var}(T_0(u)) + \sum_{u\in[K]} m_{xu}^{(n)}\,[\mathbb{E}\, T_0(u)]^2 + O(n^{2\delta}/n)\\ &= \sum_{u\in[K]} m_{xu}^{(n)}\left[\mathbb{E}(T_0(u)^2)\right] + O(n^{2\delta}/n).\end{aligned}$$

Letting $\mathbf{w}_1(n) = \left[\mathbb{E}\, T_0^2(x)\right]_{x\in[K]}$ and $\mathbf{w}_2(n) = \left[(\mathbb{E}\, T_0(x))^2\right]_{x\in[K]}$, we have

$$(I - M_n)\mathbf{w}_1(n) = \mathbf{w}_2(n) + O(n^{2\delta}/n), \tag{7.33}$$

where the second term represents a vector with each coordinate $O(n^{2\delta}/n)$. Since $\rho_n = \rho(M_n) < 1$ for large n,

$$(I - M_n)^{-1} = I + \sum_{k=1}^{\infty} M_n^k. \tag{7.34}$$

Note also that, $\mathbb{E}\, T_0^2(\mu_n) = \boldsymbol{\mu}_n^t \mathbf{w}_1(n)$. The statement $n^{-3\delta}\,\mathbb{E}\, T_0^2(\mu_n) \to \beta$ now follows from (7.29), (7.22), (7.23) and (7.27).

Suppose we have proved that all cumulants (and hence all moments via (7.31)) of order r are $O(n^{(2r-1)\delta})$ for $r \leqslant p-1$. To prove the same for $r = p$, note that the second term on the right side of (7.32) is $O(\prod_{m=1}^{q} n^{(2|I_m|-1)\delta}) = O(n^{(2p-q)\delta}) = O(n^{(2p-2)\delta})$. From (7.34), (7.22) and (7.27), it is clear that every entry of $(I - M_n)^{-1}$ is $O(n^{\delta})$. These two observations combined yield (7.14).

Next, note that

$$\mathbb{E}\, T_1(\mu_n) = \sum_{\ell=1}^{\infty} \ell\, \mathbb{E}\, G_\ell(x) = \sum_{\ell=1}^{\infty} \ell \boldsymbol{\mu}_n^t \left(\rho_n^{\ell} \mathbf{u}_n \mathbf{v}_n^t + R_n\right) \mathbf{1}.$$

From Assumption 4.1 (b), (7.23) and the facts $\|\mathbf{u}_n - \mathbf{u}\| + \|\mathbf{v}_n - \mathbf{v}\| = O(n^{-\delta})$, it follows that

$$\lim_n \frac{\mathbb{E}\, T_1(\mu_n)}{n^{2\delta}} = (\boldsymbol{\mu}^t \mathbf{u}\mathbf{v}^t \mathbf{1}) \lim_n \frac{1}{n^{2\delta}} \left(\sum_{\ell=1}^{\infty} \ell \rho_n^{\ell}\right) = \lim_n \frac{(\boldsymbol{\mu}^t \mathbf{u}\mathbf{v}^t \mathbf{1})}{n^{2\delta}(1-\rho_n)^2} = \alpha,$$

where the last equality is a consequence of (7.27). This proves (7.17).

To prove (7.18), notice that (7.27) ensures the existence of C_2 and n_3 as in the statement of Lemma 7.4 such that for $n \geqslant n_3$, we have $\rho_n \leqslant 1 - C_2/n^{\delta} \leqslant \exp(-C_2/n^{\delta})$. (In fact $C_2 = 2(\boldsymbol{\mu}^t \mathbf{u}\mathbf{v}^t \mathbf{1})$ works.) Now (7.22) yields for each $m \geqslant 1$,

$$\begin{aligned}\mathbb{P}(\mathrm{ht}(x) \geqslant m) &= \mathbb{P}(G_m(x) \geqslant 1) \leqslant \mathbb{E}\, G_m(x)\\ &\leqslant C_1 \rho_n^m \leqslant C_1 \exp\left(-C_2 m/n^{\delta}\right)\end{aligned}$$

for $n \geqslant n_3$ where C_1 is as in the statement of Lemma 7.4.

Now, (7.19) can be proved by imitating the proof of [12, Lemma 6.13], and using (7.27) and the fact: $\|\kappa_n - n^{-\delta}\|_{L^2(\mu_n)} = \rho(M_n)$. Since no new idea is involved, we omit the proof.

Finally, we prove (7.13). Note that

$$\mathbb{E}[T_0^r(\mu_n) T_k(\mu_n)] = \sum_{x\in[K]} \mu_n(x)\, \mathbb{E}[T_0^r(x) T_k(x)],$$

so we only need to prove the bound for all $x \in [K]$. Consider one initial particle of type x. Let $N = G_1(x)$ and for $i = 1,2,\ldots,N$, denote by x_i, the type of the i-th particle in generation one. Let $(T_k^{(\ell)}(y), G_k^{(\ell)}(y) : k = 0,1,\ldots, y \in [K])$, $\ell = 1,2,\ldots$, be the corresponding random variables defined on independent copies of the same branching process. By the branching structure, we have

$$
\begin{aligned}
&T_0(x) \stackrel{d}{=} 1 + \sum_{i=1}^{N} T_0^{(i)}(x_i).\\
&T_k(x) \stackrel{d}{=} \sum_{i=1}^{N}\sum_{j=0}^{\infty}(j+1)^k G_j^{(i)}(x_i) = \sum_{i=1}^{N}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k}\binom{k}{\ell} j^\ell G_j^{(i)}(x_i) \text{ (with the convention } 0^0 = 1)\\
&\qquad = \sum_{i=1}^{N}\sum_{\ell=0}^{k}\binom{k}{\ell}\left[\sum_{j=0}^{\infty} j^\ell G_j^{(i)}(x_i)\right] = \sum_{i=1}^{N}\sum_{\ell=0}^{k}\binom{k}{\ell} T_\ell^{(i)}(x_i), \text{ for } k = 1,2,\ldots
\end{aligned}
$$

The above distributional equalities also hold jointly. Observe that for $k \geqslant 0$,

$$
\mathbb{E}\, T_k(x) = \sum_{\ell \geqslant 0} \ell^k \mathbf{e}_x^t M_n^k \mathbf{1} = O\Big(\sum_{\ell\geqslant 0} \ell^k \rho_n^k\Big) = O(n^{(k+1)\delta}).
$$

So it is enough to prove (7.13) for $r \geqslant 1, k \geqslant 1$. We will prove this by induction on $r+k$. First, we show the inductive step as follows. Assume (7.14) and that

$$
\mathbb{E}[T_0^{r'}(x) T_{k'}(x)] = O(n^{(2r'+k'+1)\delta}), \text{ for all } \left\{(r',k') : r' < r \text{ or } k' < k\right\}. \tag{7.35}
$$

Then for (r,k), observe that

$$
\begin{aligned}
(T_0(x)-1)^r T_k(x) &\stackrel{d}{=} \left[\sum_{i=1}^{N} T_0^{(i)}(x_i)\right]^r \left[\sum_{i=1}^{N}\sum_{\ell=0}^{k}\binom{k}{\ell} T_\ell^{(i)}(x_i)\right]\\
&= \left[\sum_{r_1,\ldots,r_N}\binom{r}{r_1,\ldots,r_N}\prod_{j=1}^{N}(T_0^{(j)})^{r_j}\right]\left[\sum_{i=1}^{N}\sum_{\ell=0}^{k}\binom{k}{\ell} T_\ell^{(i)}(x_i)\right]\\
&= \sum_{r_1,\ldots,r_N}\sum_{i=1}^{N}\sum_{\ell=0}^{k}\left[\binom{r}{r_1,\ldots,r_N}\binom{k}{\ell} T_\ell^{(i)}(x_i)\prod_{j=1}^{N}(T_0^{(j)})^{r_j}\right]\\
&= \sum_{r_1,\ldots,r_N}\sum_{i=1}^{N}\sum_{\ell=0}^{k}\left[\binom{r}{r_1,\ldots,r_N}\binom{k}{\ell}[T_0^{(i)}(x_i)]^{r_i} T_\ell^{(i)}(x_i)\prod_{j\neq i}(T_0^{(j)}(x_j))^{r_j}\right],
\end{aligned}
\tag{7.36}
$$

where the summation $\sum_{r_1,\ldots,r_N}$ is over the set

$$
\left\{(r_1,r_2,\ldots,r_N) \in \mathbb{N}_0^N : \sum_{i=1}^{N} r_i = r\right\}.
$$

By independence, denoting $\mathscr{F}_1$ for the σ-field generated by $x_1, x_2, \ldots, x_N$, we have

$$
\mathbb{E}\left[(T_0^{(i)}(x_i))^{r_i} T_\ell^{(i)}(x_i)\prod_{j\neq i}(T_0^{(j)}(x_j))^{r_j}\,\middle|\,\mathscr{F}_1\right] = \mathbb{E}[(T_0^{(i)}(x_i))^{r_i} T_\ell^{(i)}(x_i)|\mathscr{F}_1]\prod_{j\neq i}\mathbb{E}[(T_0^{(j)}(x_j))^{r_j}|\mathscr{F}_1].
$$

Then whenever $l < k$ or $r_i < r$, we can apply the assumptions (7.35). Therefore we have

$$
\mathbb{E}\left[(T_0(x_i))^{r_i} T_\ell^{(i)}(x_i)\prod_{j\neq i}(T_0^{(j)}(x_j))^{r_j}\,\middle|\,\mathscr{F}_1\right] = O(n^{\phi(\mathbf{r},\ell,i)\delta}), \tag{7.37}
$$

where $\phi(\mathbf{r},\ell,i) = 2r+\ell-|\{j : r_j > 0\}| + \mathbb{1}_{\{r_i>0\}} + \mathbb{1}_{\{\ell>0\}}$. One can check that, when $r_i < r$ or $\ell < k$, we have $\phi(\mathbf{r},\ell,i) \le 2r+k$. Further, using (7.35) again, we get

$$\mathbb{E}[(T_0(x)-1)^r T_k(x)] = \mathbb{E}[(T_0(x))^r T_k(x)] + O(n^{(2r+k-1)\delta}).$$

Therefore, from (7.36), we have

$$\begin{aligned}\mathbb{E}[T_0(x)^r T_k(x)] &= \mathbb{E}\left[\sum_{i=1}^{N} (T_0^{(i)}(x_i))^r T_k^{(i)}(x_i)\right] + O(n^{(2r+k)\delta})\\ &= \sum_{y\in[K]} m_{xy}^{(n)}\, \mathbb{E}[T_0(y)^r T_k(y)] + O(n^{(2r+k)\delta}).\end{aligned}$$

This induction step is completed upon noting that each entry of $(I-M_n)^{-1}$ is $O(n^\delta)$. Now we only need to bound $E[T_0(x)T_1(x)]$. By an expansion similar to (7.36), we have

$$T_0(x)T_1(x) - T_1(x) \overset{d}{=} \left[\sum_{i=1}^{N} T_0^{(i)}(x_i)\right]\left[\sum_{j=1}^{N} (T_0^{(j)}(x_j) + T_1^{(j)}(x_j))\right].$$

Now we can use the facts $\mathbb{E}[(T_0(x))^2] = O(n^{3\delta})$, $\mathbb{E}\, T_0(x) = O(n^\delta)$ and $\mathbb{E}[T_1(x)] = O(n^{2\delta})$ to conclude that $\mathbb{E}\, T_0(x)T_1(x) = O(n^{4\delta})$. This proves the starting point of the induction and thus finishes the proof of (7.13). This completes the proof of Lemma 7.4. ■

The following lemma shows how closely we can approximate $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$ by the branching process.

Lemma 7.6. *We have,*

$$\left|\mathbb{E}\, T_0(\mu_n) - \mathbb{E}|\mathscr{C}(v)|\right| = O(n^{4\delta-1}),\ \left|\mathbb{E}\, T_0(\mu_n)^2 - \mathbb{E}|\mathscr{C}(v)|^2\right| = O(\sqrt{n^{9\delta-1}}) \text{ and} \tag{7.38}$$

$$\left|\mathbb{E}\, T_1(\mu_n) - \mathbb{E}\mathscr{D}(v)\right| = O(n^{4\delta-1}). \tag{7.39}$$

Further, for $r \geqslant 0$ and $n \geqslant 1$, we have

$$\mathbb{E}\left[|\mathscr{C}(v)|^r \mathscr{D}(v)\right] \leqslant \mathbb{E}\left[T_0^r(\mu_n) T_1(\mu_n)\right], \tag{7.40}$$

$$|\mathscr{C}(i)| \leqslant_{st} T_0(x_i) \text{ and } \operatorname{diam}(\mathscr{C}(i)) \leqslant_{st} 2\times \operatorname{ht}(T_0(x_i)) \tag{7.41}$$

where $X \leqslant_{st} Y$ means Y dominates X stochastically.

We now set some notation which we will follow throughout the rest of this section. For real numbers a and b, we will write "$a \preceq b$" if there exists a positive constant c depending only on κ and μ such that $a \leqslant cb$. For sequences $\{a_m\}$ and $\{b_m\}$, we will write "$a_m \preceq_m b_m$" if there exists a positive constant c depending only on κ and μ and an integer m_0 depending only on the sequences $\{\mu_n\}$ and $\{\kappa_n\}$ such that $a_m \leqslant cb_m$ for $m \geqslant m_0$. If we have two sequences $\{a_m(k)\}_{m\geqslant 1}$ and $\{b_m(k)\}_{m\geqslant 1}$ for each $k \geqslant 1$, we will write "$a_m(k) \preceq_m b_m(k)$ for $k \geqslant 1$" if $a_m(k) \leqslant cb_m(k)$ for $m \geqslant m_0$ and all $k \geqslant 1$ where c and m_0 are as before, we emphasize that the same c and m_0 work for all k.

We will use the following lemma in the proof of Lemma 7.6.

Lemma 7.7. *We have,*

$$\mathbb{E}[G_\ell(x)^2] \preceq_n \rho_n^\ell/(1-\rho_n) \text{ for } x \in [K] \text{ and } \ell \geqslant 1, \tag{7.42}$$

and for any non-random vector $\boldsymbol{w} = (w_y : y \in [K])$ *with* $w_y \geqslant 0$ *for each* $y \in [K]$*,*

$$\mathbb{E}[G_\ell(\boldsymbol{w}) H_\ell(\boldsymbol{w})] \preceq_n \frac{1}{1-\rho_n}\left[(\mathbf{1}\cdot\boldsymbol{w})\,\ell\rho_n^\ell + (\mathbf{1}\cdot\boldsymbol{w})^2\rho_n^\ell\right] \text{ for } \ell \geqslant 1. \tag{7.43}$$

(Recall the definitions of $G_\ell(\boldsymbol{w})$ *and* $H_\ell(\boldsymbol{w})$ *from* (7.11)*.)*

Proof: Let $G_k(x,y)$ denote the number of type-y particles in the k-th generation of the multitype branching process started from a single particle of type x. Let $\mathbf{G}_k(x)^t = \left(G_k(x,y) : y \in [K]\right)$. Define the vector $\mathbf{G}_k(\boldsymbol{w})$ in a similar fashion.

Let $\mathscr{F}_s = \sigma\{\mathbf{G}_k(x) : 0 \leqslant k \leqslant s\}$ for $s \geqslant 0$. For any vector $\mathbf{w} = (w_1,\ldots,w_K)^t$, let $\mathbf{w}^{(2)} = (w_1^2,\ldots,w_K^2)^t$ and $\|\mathbf{w}\|_\infty = \max_j w_j$. Also define $\mathbf{w}_k = M_n^k\mathbf{w}$ for $k \geqslant 0$. From (7.22) and (7.23), it follows that

$$\mathbf{1}^t M_n^k \mathbf{1} \preceq_n \rho_n^k \text{ and } \|\mathbf{w}_k\|_\infty \preceq_n \rho_n^k \|\mathbf{w}\|_\infty \text{ for } k \geqslant 1. \tag{7.44}$$

Now,

$$\begin{aligned}
\mathbb{E}(\mathbf{G}_\ell(x)\cdot\mathbf{w})^2 &= \mathbb{E}\left[\mathbb{E}\left((\mathbf{G}_\ell(x)\cdot\mathbf{w})^2\,\middle|\,\mathscr{F}_{\ell-1}\right)\right]\\
&= \mathbb{E}\left[\mathrm{Var}\left(\mathbf{G}_\ell(x)\cdot\mathbf{w}\,\middle|\,\mathscr{F}_{\ell-1}\right)\right] + \mathbb{E}\left(\mathbf{G}_{\ell-1}(x)^t M_n\mathbf{w}\right)^2\\
&= \sum_{y\in[K]} w_y^2\,\mathrm{Var}\left(G_\ell(x,y)\,\middle|\,\mathscr{F}_{\ell-1}\right) + \mathbb{E}(\mathbf{G}_{\ell-1}(x)\cdot\mathbf{w}_1)^2\\
&\leqslant \mathbb{E}\left[\mathbf{G}_{\ell-1}(x)^t M_n \mathbf{w}^{(2)}\right] + \mathbb{E}(\mathbf{G}_{\ell-1}(x)\cdot\mathbf{w}_1)^2 = \mathbf{e}_x^t M_n^\ell \mathbf{w}^{(2)} + \mathbb{E}(\mathbf{G}_{\ell-1}(x)\cdot\mathbf{w}_1)^2.
\end{aligned}$$

Proceeding in this fashion and making use of (7.44), we get

$$\begin{aligned}
\mathbb{E}(\mathbf{G}_\ell(x)\cdot\mathbf{w})^2 &\leqslant \mathbf{e}_x^t M_n^\ell \mathbf{w}^{(2)} + \mathbf{e}_x^t M_n^{\ell-1}\mathbf{w}_1^{(2)} + \ldots + \mathbf{e}_x^t M_n \mathbf{w}_{\ell-1}^{(2)} + (\mathbf{e}_x\cdot\mathbf{w}_\ell)^2\\
&\preceq_n \|\mathbf{w}\|_\infty^2\left(\rho_n^\ell + \rho_n^{\ell+1} + \ldots + \rho_n^{2\ell-1} + \rho_n^{2\ell}\right) \leqslant \|\mathbf{w}\|_\infty^2 \rho_n^\ell/(1-\rho_n).
\end{aligned}$$

We get (7.42) by taking $\mathbf{w} = \mathbf{1}$. Next, note that

$$\begin{aligned}
\mathbb{E}[G_\ell(\boldsymbol{w})H_\ell(\boldsymbol{w})] &= \mathbb{E}[G_\ell(\boldsymbol{w})^2] + \mathbb{E}[G_\ell(\boldsymbol{w})H_{\ell-1}(\boldsymbol{w})]\\
&= \mathbb{E}[G_\ell(\boldsymbol{w})^2] + \mathbb{E}\left[\left(\mathbf{G}_{\ell-1}(\boldsymbol{w})^t M_n\mathbf{1}\right)\times H_{\ell-1}(\boldsymbol{w})\right]\\
&\leqslant \mathbb{E}[G_\ell(\boldsymbol{w})^2] + \mathbb{E}\left[\left(\mathbf{1}^t M_n\mathbf{1}\right)\times G_{\ell-1}(\boldsymbol{w})^2\right] + \mathbb{E}\left[\left(\mathbf{G}_{\ell-1}(\boldsymbol{w})^t M_n\mathbf{1}\right)\times H_{\ell-2}(\boldsymbol{w})\right].
\end{aligned}$$

Proceeding in this way, we get

$$\mathbb{E}[G_\ell(\boldsymbol{w})H_\ell(\boldsymbol{w})] \leqslant \mathbb{E}[G_\ell(\boldsymbol{w})^2] + \sum_{k=1}^{\ell-1}\left(\mathbf{1}^t M_n^k\mathbf{1}\right)\times\mathbb{E}\left[G_{\ell-k}(\boldsymbol{w})^2\right].$$

Since

$$\begin{aligned}
\mathbb{E}[G_k(\boldsymbol{w})^2] &= \mathrm{Var}[G_k(\boldsymbol{w})] + [\mathbb{E}\,G_k(\boldsymbol{w})]^2 \\
&\leqslant (\mathbf{w}\cdot\mathbf{1})\max_{y\in[K]}\mathbb{E}[G_k(y)^2] + (\boldsymbol{w}\cdot\mathbf{1})^2(\mathbf{1}^t M_n^k\mathbf{1})^2,
\end{aligned} \tag{7.45}$$

(7.43) follows by an application of (7.42) and (7.44). ■

Proof of Lemma 7.6: Fix $i \in [n]$. Recall that x_i is the type of i and $\mathscr{C}(i)$ is the component of i in $\mathscr{G}_{\mathrm{IRG}}^{(n),-}$. It is enough to get bounds on

$|\mathbb{E}\, T_0(x_i) - \mathbb{E}|\mathscr{C}(i)||$, $\left|\mathbb{E}\, T_0(x_i)^2 - \mathbb{E}|\mathscr{C}(i)|^2\right|$ and $|\mathbb{E}\, T_1(x_i) - \mathbb{E}\mathscr{D}(i)|$ which are uniform in i. The proof proceeds via a coupling between a K-type branching process and the breadth-first construction of $\mathscr{C}(i)$. A similar coupling in the Erdős-Rényi case is standard and can be found in, for example, [30].

Let ϵ^{ℓ}_{jk} be independent Bernoulli$(\kappa_n^-(x_j, x_k)/n)$ random variables for $1 \leqslant j, k \leqslant n$ and $\ell \geqslant 1$. Let $\mathfrak{Z}_0 = \mathfrak{I}_0 = \{i\}$ and $\mathfrak{S}_0 = [n] \setminus \{i\}$. Assume that we have defined $\mathfrak{I}_t, \mathfrak{Z}_t$ and $\mathfrak{S}_t$ for $1 \leqslant t \leqslant \ell - 1$. For each $j \in \mathfrak{I}_{\ell-1}$ and $k \in \mathfrak{S}_{\ell-1}$, place an edge between j and k iff $\epsilon^{\ell}_{jk} = 1$. Let

$$\mathfrak{I}_\ell = \left\{k \in \mathfrak{S}_{\ell-1} : \epsilon^{\ell}_{jk} = 1 \text{ for some } j \in \mathfrak{I}_{\ell-1}\right\} \text{ and } \mathfrak{S}_\ell = \mathfrak{S}_{\ell-1} \setminus \mathfrak{I}_\ell.$$

Note that $\mathscr{C}(i) = \bigcup_{\ell \geqslant 0} \mathfrak{I}_\ell$. To define $\mathfrak{Z}_\ell$, we need to consider three kinds of excess particles.

(a) For each $u \in \mathfrak{Z}_{\ell-1} \setminus \mathfrak{I}_{\ell-1}$ and $y \in [K]$, create a collection $\mathfrak{E}^{(\ell)}_{uy}$ of type-y particles independently where $\mathfrak{E}^{(\ell)}_{uy}$ has Binomial$(n\mu_n(x_u), \kappa_n^-(x_u, y)/n)$ distribution. As usual, $x_u \in [K]$ denotes the type of u.

(b) For each $j \in \mathfrak{I}_{\ell-1}$ and $k \in \mathfrak{S}^c_{\ell-1}$, create a particle of type x_k iff $\epsilon^{(\ell)}_{jk} = 1$. Call this collection of newly created particles $\mathfrak{B}_\ell$.

(c) For each $k \in \mathfrak{S}_{\ell-1}$, create $\left[\sum_{j \in \mathfrak{I}_{\ell-1}} \epsilon^{(\ell)}_{jk} - \mathbb{1}\left\{\sum_{j \in \mathfrak{I}_{\ell-1}} \epsilon^{(\ell)}_{jk} \geqslant 1\right\}\right]$ many particles of type x_k. Call this collection of newly created particles $\mathfrak{C}_\ell$.

Set

$$\mathfrak{Z}_\ell = \bigcup_{\substack{u \in \mathfrak{Z}_{\ell-1} \setminus \mathfrak{I}_{\ell-1} \\ y \in [K]}} \mathfrak{E}^{(\ell)}_{uy} \bigcup (\mathfrak{I}_\ell \cup \mathfrak{B}_\ell \cup \mathfrak{C}_\ell).$$

Thus, we have constructed a K-type branching process starting from one particle of type x_i as described right after Lemma 7.3. For $\ell \geqslant 1$, define $\mathfrak{A}_\ell = \mathfrak{B}_\ell \cup \mathfrak{C}_\ell$ and write

$$Z_\ell = |\mathfrak{Z}_\ell|, I_\ell = |\mathfrak{I}_\ell|, A_\ell = |\mathfrak{A}_\ell|, B_\ell = |\mathfrak{B}_\ell|, C_\ell = |\mathfrak{C}_\ell|, S_\ell = |\mathfrak{S}_\ell| \text{ and } R_\ell = \sum_{j=0}^{\ell} I_j. \tag{7.46}$$

We will write $\boldsymbol{A}_\ell$ to denote the $K \times 1$ vector $(A_\ell(y) : y \in [K])$ where $A_\ell(y)$ is the number of type-y particles in $\mathfrak{A}_\ell$. Similarly define the $K \times 1$ vector $\boldsymbol{I}_\ell$.

Recall the definition of $T_0(\boldsymbol{w})$ from (7.12). Note that,

$$\sum_{\ell=1}^{\infty} T_0(\boldsymbol{A}_\ell) \overset{\mathrm{d}}{=} \sum_{\ell=2}^{\infty} \sum_{y \in [K]} \sum_{u \in \mathfrak{Z}_{\ell-1} \setminus \mathfrak{I}_{\ell-1}} |\mathfrak{E}^{(\ell)}_{uy}| + \sum_{\ell=1}^{\infty} (B_\ell + C_\ell).$$

Hence,

$$\begin{aligned} |\mathbb{E}\, T_0(x_i) - \mathbb{E}|\mathscr{C}(i)|| &= \sum_{\ell=1}^{\infty} \mathbb{E}(Z_\ell - I_\ell) = \sum_{\ell=1}^{\infty} \mathbb{E}\, T_0(\boldsymbol{A}_\ell) \\ &\preceq_n n^\delta \sum_{\ell=1}^{\infty} \mathbb{E}(A_\ell) = n^\delta \sum_{\ell=1}^{\infty} \mathbb{E}(B_\ell + C_\ell), \end{aligned} \tag{7.47}$$

the third step being a consequence of (7.29). Let $\mathscr{F}_0$ be the trivial σ-field and for $\ell \geqslant 1$, define

$$\mathscr{F}_\ell = \sigma\left\{\mathfrak{I}_s, \mathfrak{B}_s, \mathfrak{C}_s, \mathfrak{E}^{(s)}_{uy} : u \in \mathfrak{Z}_{s-1} \setminus \mathfrak{I}_{s-1}, y \in [K], 1 \leqslant s \leqslant \ell\right\}.$$

Then, $\mathbb{E}(B_\ell|\mathscr{F}_{\ell-1}) \preceq_n I_{\ell-1}R_{\ell-1}/n \leqslant Z_{\ell-1}(\sum_{s=0}^{\ell-1} Z_s)/n$. Hence,

$$\sum_{\ell=1}^{\infty} \mathbb{E}\, B_\ell \preceq_n \frac{1}{n}\sum_{\ell=0}^{\infty} \mathbb{E}\left[Z_\ell(1+H_\ell(\mathbf{e}_{x_i}))\right] = \frac{1}{n}\mathbb{E}\left[T_0(x_i) + \sum_{\ell=0}^{\infty} G_\ell(\mathbf{e}_{x_i})H_\ell(\mathbf{e}_{x_i})\right] \preceq_n \frac{n^{3\delta}}{n}, \tag{7.48}$$

the last inequality being a consequence of (7.14) and (7.43). We also have

$$\mathbb{E}(C_\ell|\mathscr{F}_{\ell-1}) \preceq_n I_{\ell-1}^2 \times S_{\ell-1} \times \frac{1}{n^2} \leqslant \frac{Z_{\ell-1}^2}{n}.$$

Hence,

$$\sum_{\ell=1}^{\infty} \mathbb{E}\, C_\ell \preceq_n \frac{1}{n}\sum_{\ell=0}^{\infty} \mathbb{E}(Z_\ell^2) \preceq_n \frac{n^{2\delta}}{n}. \tag{7.49}$$

Combining (7.47), (7.48) and (7.49), we get the first half of (7.38). To get the other inequality, note that

$$\begin{aligned} \left|\mathbb{E}\, T_0(x_i)^2 - \mathbb{E}\,|\mathscr{C}(i)|^2\right| &= \mathbb{E}\left(\sum_{\ell=1}^{\infty} Z_\ell\right)^2 - \mathbb{E}\left(\sum_{\ell=1}^{\infty} I_\ell\right)^2 \qquad (7.50)\\ &\leqslant \left[\mathbb{E}\left(\sum_{\ell=1}^{\infty}(Z_\ell - I_\ell)\right)^2\right]^{\frac{1}{2}} \times \left[\mathbb{E}\left(\sum_{\ell=1}^{\infty}(Z_\ell + I_\ell)\right)^2\right]^{\frac{1}{2}} \preceq_n n^{3\delta/2}\left[\mathbb{E}\left(\sum_{\ell=1}^{\infty}(Z_\ell - I_\ell)\right)^2\right]^{\frac{1}{2}}, \end{aligned}$$

where the last step follows from (7.14). Now,

$$\mathbb{E}\left(\sum_{\ell=1}^{\infty}(Z_\ell - I_\ell)\right)^2 = \mathbb{E}\left[\sum_{\ell=1}^{\infty} T_0(\mathbf{A}_\ell)\right]^2 \leqslant 2\,\mathbb{E}\left[\sum_{\ell=1}^{\infty} T_0(\mathbf{A}_\ell)^2 + \sum_{1\leqslant \ell\leqslant s} T_0(\mathbf{A}_\ell)T_0(\mathbf{A}_{s+1})\right]. \tag{7.51}$$

By an argument similar to the one used in (7.45) and the estimate from (7.14),

$$\mathbb{E}\left(T_0(\mathbf{A}_\ell)^2\middle|\mathscr{F}_\ell\right) \preceq_n A_\ell n^{3\delta} + A_\ell^2 n^{2\delta}. \tag{7.52}$$

Also

$$\mathbb{E}\,[T_0(\mathbf{A}_\ell)T_0(\mathbf{A}_{s+1})] \preceq_n n^{2\delta}\,\mathbb{E}(A_\ell A_{s+1}). \tag{7.53}$$

From (7.48) and (7.49), we have $\sum_{\ell\geqslant 1}\mathbb{E}\, A_\ell \preceq_n n^{3\delta}/n$. Further, $A_\ell^2 \leqslant 2(B_\ell^2 + C_\ell^2)$ and

$$\mathbb{E}\left(B_\ell^2\middle|\mathscr{F}_{\ell-1}\right) = \mathrm{Var}(B_\ell|\mathscr{F}_{\ell-1}) + (\mathbb{E}(B_\ell|\mathscr{F}_{\ell-1}))^2 \preceq_n \frac{Z_{\ell-1}}{n}\left(\sum_{s=0}^{\ell-1} Z_s\right) + \left[\frac{Z_{\ell-1}}{n}\left(\sum_{s=0}^{\ell-1} Z_s\right)\right]^2.$$

Hence,

$$\begin{aligned} \sum_{\ell=1}^{\infty}\mathbb{E}(B_\ell^2) &\preceq_n \frac{1}{n}\mathbb{E}\left[\sum_{\ell=0}^{\infty} Z_\ell\left(\sum_{s=0}^{\infty} Z_s\right)\right] + \frac{1}{n^2}\mathbb{E}\left[\sum_{\ell=0}^{\infty} Z_\ell^2\left(\sum_{s=0}^{\infty} Z_s\right)^2\right]\\ &\leqslant \frac{1}{n}\mathbb{E}\left[T_0(x_i)^2\right] + \frac{1}{n^2}\mathbb{E}\left[T_0(x_i)^4\right] \preceq_n n^{3\delta-1}, \end{aligned}$$

the final inequality follows from (7.14) and the fact that $\delta < 1/5$. Similarly,

$$\begin{aligned} \mathbb{E}(C_\ell^2|\mathscr{F}_{\ell-1}) &= \mathrm{Var}(C_\ell|\mathscr{F}_{\ell-1}) + [\mathbb{E}(C_\ell|\mathscr{F}_{\ell-1})]^2\\ &\preceq_n I_{\ell-1}^2 S_{\ell-1}/n^2 + \left[I_{\ell-1}^2 S_{\ell-1}/n^2\right]^2 \leqslant Z_{\ell-1}^2/n + Z_{\ell-1}^4/n^2. \end{aligned}$$

Using (7.42) and (7.14), we conclude that

$$\sum_{\ell=1}^{\infty} \mathbb{E}(C_\ell^2) \preceq_n \frac{1}{n}\sum_{\ell=0}^{\infty} \mathbb{E}\, Z_\ell^2 + \frac{1}{n^2}\mathbb{E}\, T_0(x_i)^4 \preceq_n n^{2\delta-1} + n^{7\delta-2}.$$

Combining these observations with (7.51), (7.52) and (7.53), we get

$$\mathbb{E}\left(\sum_{\ell=1}^{\infty}(Z_\ell - I_\ell)\right)^2 \preceq_n n^{6\delta-1} + n^{2\delta}\sum_{\ell=1}^{\infty}\sum_{s=\ell}^{\infty}\mathbb{E}\left[A_\ell B_{s+1} + A_\ell C_{s+1}\right]. \tag{7.54}$$

Now,

$$\begin{aligned}\mathbb{E}[A_\ell B_{s+1}] &\preceq_n \mathbb{E}\left[A_\ell\left(\frac{I_s R_s}{n}\right)\right] \leqslant \frac{1}{n}\mathbb{E}\left[A_\ell G_{s-\ell}(\boldsymbol{I}_\ell)\left(R_\ell + H_{s-\ell}(\boldsymbol{I}_\ell)\right)\right]\\ &\preceq_n \frac{1}{n}\cdot\rho_n^{s-\ell}\,\mathbb{E}\left[A_\ell I_\ell R_\ell\right] + \frac{1}{n(1-\rho_n)}\mathbb{E}\left[A_\ell\left((s-\ell)\rho_n^{s-\ell} I_\ell + \rho_n^{s-\ell} I_\ell^2\right)\right],\end{aligned}$$

where the last step is a consequence of (7.43) and the observation: $\mathbb{E}\left[G_{s-\ell}(\boldsymbol{I}_\ell)\middle|\mathscr{F}_\ell\right] \preceq_n I_\ell \times \max_{x\in[K]}\mathbb{E}\, G_{s-\ell}(x)$. Using (7.27), a simple computation yields

$$\sum_{\ell=1}^{\infty}\sum_{s=\ell}^{\infty}\mathbb{E}\, A_\ell B_{s+1} \preceq_n \sum_{\ell=1}^{\infty}\left[\frac{n^{\delta}}{n}\mathbb{E}(A_\ell I_\ell R_\ell) + \frac{n^{3\delta}}{n}\mathbb{E}(A_\ell I_\ell) + \frac{n^{2\delta}}{n}\mathbb{E}(A_\ell I_\ell^2)\right]. \tag{7.55}$$

We can write,

$$\sum_{\ell=1}^{\infty}\mathbb{E}(A_\ell I_\ell R_\ell) = \sum_{\ell=1}^{\infty}\left[\mathbb{E}(B_\ell I_\ell^2) + \mathbb{E}(B_\ell I_\ell R_{\ell-1}) + \mathbb{E}(C_\ell I_\ell^2) + \mathbb{E}(C_\ell I_\ell R_{\ell-1})\right]. \tag{7.56}$$

To bound the first term on the right side, note that

$$\mathbb{E}(B_\ell I_\ell^2) = \mathbb{E}\left[\mathbb{E}\left(B_\ell\middle|\mathscr{F}_{\ell-1}\right)\mathbb{E}\left(I_\ell^2\middle|\mathscr{F}_{\ell-1}\right)\right] \preceq_n \mathbb{E}\left[\frac{I_{\ell-1}R_{\ell-1}}{n}\cdot\left[\mathrm{Var}\left(I_\ell\middle|\mathscr{F}_{\ell-1}\right) + \left(\mathbb{E}\left(I_\ell\middle|\mathscr{F}_{\ell-1}\right)\right)^2\right]\right],$$

where the first equality holds because of independence between B_ℓ and I_ℓ conditional on $\mathscr{F}_{\ell-1}$. Thus,

$$\begin{aligned}\sum_{\ell=1}^{\infty}\mathbb{E}(B_\ell I_\ell^2) &\preceq_n \sum_{\ell=1}^{\infty}\mathbb{E}\left[\frac{I_{\ell-1}R_{\ell-1}}{n}\cdot(I_{\ell-1} + I_{\ell-1}^2)\right] \leqslant \frac{2}{n}\sum_{\ell=1}^{\infty}\mathbb{E}\left[I_{\ell-1}^3 R_{\ell-1}\right]\\ &\leqslant \frac{2}{n}\sum_{\ell=0}^{\infty}\mathbb{E}\left[Z_\ell^3\left(\sum_{\ell=0}^{\infty} Z_\ell\right)\right] \leqslant \frac{2}{n}\mathbb{E}\left[T_0(x_i)^4\right] \preceq_n \frac{n^{7\delta}}{n},\end{aligned} \tag{7.57}$$

by an application of (7.14). By a similar argument,

$$\sum_{\ell=1}^{\infty}\mathbb{E}\left[B_\ell I_\ell R_{\ell-1}\right] \preceq_n n^{7\delta-1}. \tag{7.58}$$

Next,

$$\mathbb{E}\left[C_\ell I_\ell^2\middle|\mathscr{F}_{\ell-1}\right] \leqslant \sum\nolimits_1 \mathbb{E}\left[\mathbb{1}\left\{\epsilon_{j_1,k}^{(\ell)} = \epsilon_{j_2,k}^{(\ell)} = 1\right\}\cdot I_\ell^2\middle|\mathscr{F}_{\ell-1}\right]$$

where $\sum_1$ stands for sum over all $j_1, j_2 \in \mathfrak{I}_{\ell-1}$ and $k \in \mathfrak{S}_{\ell-1}$ such that $j_1 \neq j_2$. For any such j_1, j_2, k, we can write

$$I_\ell = V_{j_1,j_2;k}^{(\ell)} + \mathbb{1}\left\{\epsilon_{j_1,k}^{(\ell)} = 1 \text{ or } \epsilon_{j_2,k}^{(\ell)} = 1\right\}$$

where $V^{(\ell)}_{j_1,j_2;k}$ is independent of $(\epsilon^{(\ell)}_{j_1,k}, \epsilon^{(\ell)}_{j_2,k})$ conditional on $\mathscr{F}_{\ell-1}$. Hence,

$$\mathbb{E}\left[C_\ell I_\ell^2 \middle| \mathscr{F}_{\ell-1}\right] \leqslant \sum\nolimits_1 \mathbb{E}\left[\mathbb{1}\left\{\epsilon^{(\ell)}_{j_1,k} = \epsilon^{(\ell)}_{j_2,k} = 1\right\} \middle| \mathscr{F}_{\ell-1}\right] \cdot \mathbb{E}\left[\left(1 + V^{(\ell)}_{j_1,j_2;k}\right)^2 \middle| \mathscr{F}_{\ell-1}\right]$$
$$\preceq_n \frac{1}{n^2} \times I_{\ell-1}^2 S_{\ell-1} \times \left[1 + \mathbb{E}\left(I_\ell^2 \middle| \mathscr{F}_{\ell-1}\right)\right] \preceq_n \frac{I_{\ell-1}^4}{n}.$$

We thus have

$$\sum_{\ell=1}^{\infty} \mathbb{E}\left[C_\ell I_\ell^2\right] \preceq_n \sum_{\ell=0}^{\infty} \frac{1}{n} \mathbb{E}\left[I_\ell^4\right] \leqslant \frac{1}{n} \mathbb{E}\, T_0(x_i)^4 \preceq_n \frac{n^{7\delta}}{n}. \tag{7.59}$$

We can similarly argue that

$$\sum_{\ell=1}^{\infty} \mathbb{E}\left[C_\ell I_\ell R_{\ell-1}\right] \preceq_n n^{7\delta-1}. \tag{7.60}$$

Combining (7.56), (7.57), (7.58), (7.59) and (7.60), we have

$$\sum_{\ell=1}^{\infty} \mathbb{E}\left[A_\ell I_\ell R_\ell\right] \preceq_n n^{7\delta-1}. \tag{7.61}$$

We can use similar reasoning to bound the second and third terms on the right side of (7.55) and the term $\mathbb{E}(A_\ell C_{s+1})$ appearing on the right side of (7.54), we omit the details. The final estimates will be:

$$\sum_{\ell=1}^{\infty} \mathbb{E}\left[A_\ell I_\ell\right] \preceq_n n^{5\delta-1}, \; \sum_{\ell=1}^{\infty} \mathbb{E}\left[A_\ell I_\ell^2\right] \preceq_n n^{7\delta-1} \text{ and } \sum_{\ell=1}^{\infty}\sum_{s=\ell}^{\infty} \mathbb{E}\left[A_\ell C_{s+1}\right] \preceq_n n^{8\delta-2}. \tag{7.62}$$

We get the second inequality in (7.38) by combining (7.50), (7.54), (7.55), (7.61) and (7.62).

Next, notice that

$$|\mathbb{E}\, T_1(x_i) - \mathbb{E}\mathscr{D}(i)| = \sum_{\ell=1}^{\infty} \mathbb{E}\left[\ell Z_\ell - \ell I_\ell\right] = \sum_{\ell=1}^{\infty} \ell\, \mathbb{E} B_\ell + \sum_{\ell=1}^{\infty} \ell\, \mathbb{E} C_\ell,$$

so we can again argue similarly to get the estimate (7.39). Finally, (7.40) and (7.41) are immediate from the coupling between the branching process and the breadth-first construction of a component. This completes the proof of Lemma 7.6. ■

We will need the following lemma to prove (7.8).

Lemma 7.8. *Fix $n \geqslant 1$ and as before, let $\mathcal{V} = \mathcal{V}^{(n)} = [n]$ and define $\mathcal{V}^- = [n] \setminus \{1\}$. Recall that for each $i \in [n]$, $x_i \in [K]$ denotes the type of the vertex i. Let $\bar\kappa$ be a kernel on $[K] \times [K]$. Let $\mathscr{G}_1$ be the IRG model on the vertex set $\mathcal{V}$ where we place an edge between $i, j \in [n], i \neq j$, independently with probability $(\bar\kappa(x_i, x_j)/n \wedge 1)$. Let $\mathscr{G}_0$ be the graph on the vertex set $\mathcal{V}^-$ induced by $\mathscr{G}_1$. Define $A := \max_{x,y\in[K]} \bar\kappa(x, y)$. Then, we have,*

$$\mathbb{E}[\mathscr{D}(\mathscr{G}_0)] \leq \mathbb{E}[\mathscr{D}(\mathscr{G}_1)] + \frac{A^2}{2n^2} \mathbb{E}[\mathscr{D}(\mathscr{G}_0)\mathscr{S}_2(\mathscr{G}_0)].$$

Proof: Define the event

$$E := \{\exists \text{ a component } \mathscr{C} \text{ of } \mathscr{G}_0 \text{ such that there are at least}$$
$$\text{two edges between vertex 1 and the component } \mathscr{C}\}.$$

One important observation is that, on the event E^c, $\mathscr{D}(\mathscr{G}_0) \le \mathscr{D}(\mathscr{G}_1)$. Note also that the connection probability of each pair of vertices is bounded by A/n. Hence,

$$\mathbb{P}(E|\mathscr{G}_0) \le \sum_{\mathscr{C}\subset\mathscr{G}_0} \binom{|\mathscr{C}|}{2}\left(\frac{A}{n}\right)^2 \le \frac{A^2\mathscr{S}_2(\mathscr{G}_0)}{2n^2}.$$

Thus we have,

$$\mathbb{E}[\mathscr{D}(\mathscr{G}_0)] = \mathbb{E}\left[\mathscr{D}(\mathscr{G}_0)\mathbb{1}_{E^c}\right] + \mathbb{E}\left[\mathscr{D}(\mathscr{G}_0)\,\mathbb{E}(\mathbb{1}_E \mid \mathscr{G}_0)\right] \le \mathbb{E}[\mathscr{D}(\mathscr{G}_1)] + \frac{A^2}{2n^2}\mathbb{E}[\mathscr{D}(\mathscr{G}_0)\mathscr{S}_2(\mathscr{G}_0)].$$

This completes the proof. ∎

Proof of Proposition 7.2: Most of our work is already done, (7.6) follows from (7.4), (7.38), (7.39), (7.16) and (7.17). (7.7) is a consequence of (7.9), (7.41) and (7.14). So we turn directly to

Proof of (7.8)**:** Our goal is to bound $\mathbb{E}[\mathscr{D}(u)\mathscr{D}(v)]$. Let $N = |\mathscr{C}(v)|$ and let $\mathscr{V}(\mathscr{C}(v))$ be the vertex set of $\mathscr{C}(v)$. Define $\mathscr{V}_0 := [n] \setminus \mathscr{V}(\mathscr{C}(v))$ and consider a sequence of sets $\mathscr{V}_0 \subset \mathscr{V}_1 \subset ... \subset \mathscr{V}_N = [n]$ such that $\mathscr{V}_i = \mathscr{V}_{i-1} \cup \{v_i\}$ and $\mathscr{V}(\mathscr{C}(v)) = \{v_1, v_2, ..., v_N\}$. Notice that

$$\mathbb{E}[\mathscr{D}(u)|\{\mathscr{C}(v), v\}] = \frac{1}{n}\sum_{i,j\in\mathscr{C}(v)} d(i,j) + \frac{1}{n}\mathbb{E}\left(\sum_{i,j\in\mathscr{V}_0} d(i,j)\mathbb{1}_{\{d(i,j)<\infty\}}\,\middle|\,\{\mathscr{C}(v), v\}\right). \qquad (7.63)$$

Define $\mathscr{G}_i = \mathscr{G}^{(n),\star}_{\mathrm{IRG}}(\kappa_n^-, \mathscr{V}_i)$, $i = 0, 1, \ldots, N$. Applying Lemma 7.8 repeatedly, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j\in\mathscr{V}_0} d(i,j)\mathbb{1}_{\{d(i,j)<\infty\}}\,\middle|\,\{\mathscr{C}(v), v\}\right] &\le \mathbb{E}[\mathscr{D}(\mathscr{G}_1)] + \frac{A^2}{2}\mathbb{E}[\bar{\mathscr{D}}(\mathscr{G}_0)\bar{s}_2(\mathscr{G}_0)] \\
&\le \ldots \le \mathbb{E}[\mathscr{D}(\mathscr{G}_N)] + \frac{A^2}{2}\sum_{i=0}^{N-1}\mathbb{E}[\bar{\mathscr{D}}(\mathscr{G}_i)\bar{s}_2(\mathscr{G}_i)] \\
&= n\mathbb{E}[\mathscr{D}(v)] + \frac{A^2}{2}\sum_{i=0}^{N-1}\mathbb{E}[\bar{\mathscr{D}}(\mathscr{G}_i)\bar{s}_2(\mathscr{G}_i)]. \qquad (7.64)
\end{aligned}$$

Here, $A = 2\max_{x,y\in[K]}\kappa(x,y)$ and each inequality holds for $n \geqslant n_0$ where n_0 depends only on the sequence $\{\kappa_n\}$.

Define the event $F_n = \left\{|\mathscr{C}(v)| \leqslant Bn^{2\delta}\log n\right\}$ where B is a positive constant such that $n^5\,\mathbb{P}(F_n^c) \to 0$. (This can be done because of (7.19) and (7.41).) On F_n, the empirical distribution of types of vertices in $\mathscr{V}_i$ will satisfy the conditions required for (7.15) to hold. Now, note that $\mathrm{Cov}(\bar{\mathscr{D}}(\mathscr{G}_i), \bar{s}_2(\mathscr{G}_i))$ can be bounded by following the proof techniques of (7.10) and (7.40). Then an application of (7.15) will yield $\mathrm{Cov}(\bar{\mathscr{D}}(\mathscr{G}_i), \bar{s}_2(\mathscr{G}_i)) \preceq_n n^{6\delta-1}$ whenever $|\mathscr{C}(v)| \leqslant Bn^{2\delta}\log n$. Similarly, $\mathbb{E}\,\bar{s}_2(\mathscr{G}_i) \preceq_n n^{\delta}$ and $\mathbb{E}\bar{\mathscr{D}}(\mathscr{G}_i) \preceq_n n^{2\delta}$. Thus,

$$\mathbb{E}\left[\bar{\mathscr{D}}(\mathscr{G}_i)\bar{s}_2(\mathscr{G}_i)\right] = \mathrm{Cov}\left(\bar{\mathscr{D}}(\mathscr{G}_i), \bar{s}_2(\mathscr{G}_i)\right) + \mathbb{E}\left[\bar{\mathscr{D}}(\mathscr{G}_i)\right]\mathbb{E}\left[\bar{s}_2(\mathscr{G}_i)\right] \preceq_n (n^{6\delta-1} + n^{3\delta}) \preceq_n n^{3\delta}$$

for $i = 1, \ldots, N$ whenever $N \leqslant Bn^{2\delta}\log n$. Since $\mathbb{E}\mathscr{D}(v) \preceq_n n^{2\delta}$ (by an application of (7.40) and (7.13)), we conclude from (7.64) that on the event F_n,

$$\mathbb{E}\left[\sum_{i,j\in\mathscr{V}_0} d(i,j)\mathbb{1}_{\{d(i,j)<\infty\}}\,\middle|\,\{\mathscr{C}(v), v\}\right] \leqslant n\mathbb{E}\mathscr{D}(v) + \epsilon_n \text{ where } \epsilon_n \preceq_n n^{5\delta}\log n. \qquad (7.65)$$

Hence,

$$\begin{aligned}\mathbb{E}[\mathscr{D}(v)\mathscr{D}(u)] &= \mathbb{E}[\mathscr{D}(v)\mathscr{D}(u)\mathbb{1}_{F_n^c}] + \mathbb{E}[\mathscr{D}(v)\mathscr{D}(u)\mathbb{1}_{F_n}] \\ &\leqslant n^4\,\mathbb{P}(F_n^c) + E\left[\mathscr{D}(v)\mathbb{1}_{F_n}\mathbb{E}[\mathscr{D}(u)\,\middle|\,\{\mathscr{C}(v),v\}]\right] \\ &\preceq_n n^4\,\mathbb{P}(F_n^c) + \frac{1}{n}\mathbb{E}\left[\mathscr{D}(v)\sum_{i,j\in\mathscr{C}(v)} d(i,j)\right] + [\mathbb{E}\mathscr{D}(v)]^2 + \mathbb{E}\mathscr{D}(v)\frac{\epsilon_n}{n} \\ &=: n^4\,\mathbb{P}(F_n^c) + Q_n + [\mathbb{E}\mathscr{D}(v)]^2 + \epsilon_n\mathbb{E}\mathscr{D}(v)/n\end{aligned} \tag{7.66}$$

the third line being a consequence of (7.63) and (7.65). Since $n^5\,\mathbb{P}(F_n^c)\to 0$ and $\epsilon_n\mathbb{E}\mathscr{D}(v)\preceq_n n^{7\delta}\log n$, we just need to show $Q_n \preceq_n n^{8\delta-1}$. To this end, note that

$$\sum_{i,j\in\mathscr{C}(v)} d(i,j) \leqslant 2\sum_{i,j\in\mathscr{C}(v)} d(i,v) = 2|\mathscr{C}(v)|\sum_{i\in\mathscr{C}(v)} d(i,v) = 2|\mathscr{C}(v)|\mathscr{D}(v).$$

Further, we trivially have $\mathscr{D}(v)\leqslant|\mathscr{C}(v)|^2$. Thus,

$$Q_n \leqslant \frac{2}{n}\mathbb{E}\left[|\mathscr{C}(v)|^3\mathscr{D}(v)\right] \preceq_n n^{8\delta-1},$$

by an application of (7.40) and (7.13). This completes the proof of (7.8). ■

7.3. **Proof of Theorem 4.4.** By a simple union bound, $\mathbb{P}(|\mathscr{C}_1^\star|\geqslant m)\leqslant\sum_{i\in[n]}\mathbb{P}(|\mathscr{C}(i)|\geqslant m)$ for $m\geqslant 1$. Hence, the tail bound on $|\mathscr{C}_1^\star|$ is immediate from (7.19) and (7.41). Similarly, the tail bound on $\mathscr{D}^\star_{\max}$ follows from (7.18) and (7.41).

Since $2\delta > 1/3$, the first convergence in (7.6) shows that

$$\lim_n n^{-\delta}\,\mathbb{E}\,\bar{s}_2^\star = 1 \tag{7.67}$$

which in turn implies

$$\lim_n n^{1/3}\left(\frac{1}{n^\delta} - \frac{1}{\mathbb{E}\,\bar{s}_2^\star}\right) = \zeta. \tag{7.68}$$

Further, for each $\epsilon>0$, $\mathbb{P}(|\bar{s}_2^\star - \mathbb{E}\,\bar{s}_2^\star| > \epsilon n^\delta)\leqslant \epsilon^{-2}n^{-2\delta}\,\mathrm{Var}(\bar{s}_2^\star)\to 0$ by (7.7). Hence,

$$n^{-\delta}\,\bar{s}_2^\star \xrightarrow{\mathrm{P}} 1. \tag{7.69}$$

Now, for each $\epsilon>0$,

$$\begin{aligned}\mathbb{P}\left(\left|n^{1/3}\left(\frac{1}{\bar{s}_2^\star} - \frac{1}{\mathbb{E}\,\bar{s}_2^\star}\right)\right| > \epsilon\right) &\leqslant \mathbb{P}\left(2n^{1/3}\frac{\left|\bar{s}_2^\star - \mathbb{E}\,\bar{s}_2^\star\right|}{n^\delta\,\mathbb{E}\,\bar{s}_2^\star} > \epsilon\right) + \mathbb{P}\left(\bar{s}_2^\star \leqslant \frac{n^\delta}{2}\right) \\ &\leqslant \frac{4n^{2/3}\,\mathrm{Var}(\bar{s}_2^\star)}{\epsilon^2 n^{2\delta}(\mathbb{E}\,\bar{s}_2^\star)^2} + \mathbb{P}\left(\bar{s}_2^\star\leqslant\frac{n^\delta}{2}\right)\to 0\end{aligned} \tag{7.70}$$

where the last convergence is due to (7.7), (7.67) and (7.69). The second convergence in (4.5) now follows from (7.68) and (7.70).

The other two claims, namely $\bar{s}_3^\star/(\bar{s}_2^\star)^3 \xrightarrow{\mathrm{P}} \beta$ and $n^{-2\delta}\bar{\mathscr{D}}^\star \xrightarrow{\mathrm{P}} \alpha$ are simple consequences of Proposition 7.2, so we omit the details. ■

8. PROOFS: SCALING LIMITS OF THE CONFIGURATION MODEL

This section is organized as follows:

(a) In Section 8.1 we start with some simple preliminary estimates about the model including Lemma 8.2 on exponential concentration of the density of free edges $\bar{s}_1$ about a limit function s_1 (8.1). We also describe an equivalence between percolation on $\mathrm{CM}_n(\infty)$ and the dynamic construction of CM_n at a fixed time t.

(b) In Section 8.2 we start proving the main results starting with Theorem 4.11 on the maximal component size and diameter in the barely subcritical regime. In Section 8.3, using Theorem 4.11 we prove properties of the susceptibility functions namely Theorem 4.10.

(c) The dynamic version of the CM model does not have the exact merger dynamics as the multiplicative coalescent. In Section 8.4 we define a modification of the configuration model starting from the configuration $\mathrm{CM}_n(t_n)$ run from time $t_n = t_c - n^{-\delta}$ to $t_c + \lambda/n^{1/3}$ which has the same dynamics as the multiplicative coalescent. Section 8.5 derives properties of this modified process including component sizes and scaling limits of the associated metric spaces using the general universality result Theorem 3.4 that can be applied since Theorems 4.11, 4.10 guarantee that the configuration of blobs $\mathrm{CM}_n(t_n)$ satisfy the assumptions of Theorem 3.4.

(d) In Section 8.6 we study a coupling between the modified process and the original process to complete the proof of Theorem 4.9. Then in Section 8.7 we complete the proof of the scaling limits of the percolation clusters of the CM namely Theorem 4.7.

8.1. Preliminaries.

We start with some simple properties of the dynamic construction $\{\mathrm{CM}_n(t) : t \geqslant 0\}$. Recall that the degrees of the vertices are generated in an iid fashion from a distribution $\mathbf{p} = \{p_k : k \geqslant 0\}$ satisfying finite exponential tails as in Assumption 4.6. For $d \sim \mathbf{p}$ and $r \geqslant 1$ let $\sigma_r = \mathbb{E}(d^r)$ denote the r-th moment of the degree sequence. It will be convenient to work with a deterministic degree sequence satisfying some regularity conditions.

Assumption 8.1. *Assume that the degree sequence $\{\mathbf{d}(n)\}_{n\geqslant 1}$ with $\mathbf{d}(n) := \{d_i : i \in [n]\}$ satisfies the following regularity properties : There exists some $N < \infty$ such that for all $n > N$ the following assertions are true.*

(a) **Max degree:** *There exists $\lambda > 0$ such that the maximal degree* $\deg_{\max} = \max_{i\in[n]} d_i < \lambda \log n$.

(b) **First four moments:** *There exist constants $\sigma_i > 0$ and $q > 4$ such that for $1 \leqslant r \leqslant 4$,*

$$\left| \frac{1}{n} \sum_{i=1}^{n} d_i^r - \sigma_r \right| \leqslant \frac{(\log n)^q}{\sqrt{n}}.$$

(c) **Supercriticality:** *We have*

$$\nu = \frac{\sigma_2 - \sigma_1}{\sigma_1} > 1.$$

Write $\sigma_1 = \mu$.

Obviously degrees generated in an iid fashion under Assumption 4.6 satisfy these assumptions almost surely. As before let $\mathscr{C}_i(t)$ denotes the i-th largest component at time t

and

$$f_i(t) \leqslant \sum_{v \in \mathscr{C}_i(t)} d_v,$$

the number of alive (used interchangeably with free) half-edges in this component at time t; the inequality in the above equation arises owing to the fact that by time t some of the half edges at $t = 0$ would have been used up to form full edges thus creating the component $\mathscr{C}_i(t)$. Let $\bar{s}_1(t) = \sum_i f_i(t)/n$ be the density free edges at time t. Define the function

$$s_1(t) = \mu \exp(-2t), \qquad t \geqslant 0. \tag{8.1}$$

Lemma 8.2. *Under Assumption* 8.1*, for any* $T > 0$*,* $\exists C(T) > 0$ *and integer* $n_0 = n_0(T) < \infty$ *such that for all* $n > n_0$ *and all* $\gamma \in [0, 1/2)$*,*

$$\mathbb{P}(\sup_{0 \leqslant t \leqslant T} |\bar{s}_1(t) - s_1(t)| > n^{-\gamma}) \leqslant \exp(-C(T)n^{1-2\gamma}).$$

Proof: First note that $\bar{s}_1(\cdot)$ is a pure death process with jumps $\Delta\bar{s}_1(t) = -2/n$ at rate $n\bar{s}_1(t)$, the total rate at which the exponential clock of one of the alive half-edges at time t rings. Writing $\mu^{(n)} = n^{-1}\sum_{i\in[n]} d_i$, note that $\bar{s}_1(0) = \mu^{(n)}$. By [32, 45] this implies that $\bar{s}_1$ can be constructed as the unique solution of the stochastic equation

$$\bar{s}_1(t) := \mu^{(n)} - \frac{2}{n} Y\left(n \int_0^t \bar{s}_1(u)\,du\right), \qquad t \geqslant 0, \tag{8.2}$$

where $\{Y(t) : t \geqslant 0\}$ is a rate one Poisson process. Analogous to the asserted limit s_1 as in (8.1), let $s_1^*(t) = \mu^{(n)} \exp(-2t)$ and note that by Assumption 8.1 for all large n,

$$\sup_{t \geqslant 0} |s_1^*(t) - s_1(t)| \leqslant \frac{\log^q n}{\sqrt{n}}. \tag{8.3}$$

Further, $s_1^*(t)$ satisfies the integral equation

$$s_1^*(t) := \mu^{(n)} - 2\int_0^t s_1^*(u)\,du. \tag{8.4}$$

Using (8.2), (8.4) and $\bar{s}_1(\cdot) \leqslant \mu^{(n)}$, we have

$$\begin{aligned}
|\bar{s}_1(t) - s_1^*(t)| &= \left|\frac{2}{n} Y\left(n \int_0^t \bar{s}_1(u)\,du\right) - 2\int_0^t s_1^*(u)\,du\right| \\
&\leqslant \sup_{0 \leqslant t \leqslant \mu^{(n)} T} \left|\frac{2}{n} Y(nt) - 2t\right| + 2\int_0^t |\bar{s}_1(u) - s_1^*(u)|\,du.
\end{aligned}$$

Gronwall's lemma [32, Page 498] implies

$$\sup_{0 \leqslant t \leqslant T} |\bar{s}_1(t) - s_1^*(t)| \leqslant e^{2T} \sup_{t \leqslant \mu^{(n)} T} \left|\frac{2}{n} Y(nt) - 2t\right|$$

Standard large deviations for the Poisson process and (8.3) completes the proof. ■

The next two results describe an equivalence between the dynamic configuration model at finite times t and percolation on the full graph $\mathrm{CM}_n(\infty)$. We start with the following trivial Lemma. Recall that d_i denoted the degree of vertex i and $\sum_{i\in[n]} d_i$ denoted the total number of half-edges at time $t = 0$.

Lemma 8.3. *Fix any time $t > 0$ and $k \geqslant 1$. Then conditional on the number of full edges $|E(\mathrm{CM}_n(t))| = k$, as a random graph $\mathrm{CM}_n(t)$ is equivalent to the following construction:*

(a) *Choose $2k$ half-edges amongst all $\sum_{i\in[n]} d_i$ edges uniformly at random.*
(b) *Perform the configuration model with these $2k$ half-edges namely perform a uniform matching of these chosen $2k$ half-edges.*

Proof: Note that Part(a) is obvious by symmetry. Again conditional on the half-edges in $\mathrm{CM}_n(t)$ being $\{e_1, e_2, \dots e_{2k}\}$ again (b) follows by symmetry. ■

The next result is much more non-trivial and follows from [34], also see [36].

Proposition 8.4 ([34, Lemma 3.1 and 3.2])**.** *Fix edge probability p and consider percolation on $\mathrm{CM}_n(\infty)$ with edge probability p. Write $\mathrm{Perc}_n(p)$ for the resulting random graph. Then conditional on the number of edges $|E(\mathrm{CM}_n(t))| = k$, as a random graph $\mathrm{Perc}_n(t)$ is equivalent to the the construction in Lemma 8.3.*

Remark 4. Let us now briefly describe how we will use the above results. In the next section, we will use Proposition 8.4 to read off results about $\mathrm{CM}_n(t)$ for various choices of t, especially $t = t_c - \varepsilon_n$ for appropriate sequences $\varepsilon_n \to 0$. Further, for fixed time t, we will choose $p(t)$ appropriately and rather than conditioning on the total number of half-edges as in Proposition 8.4, we will retain each half-edge with probability $p(t)$ and remove them with probability $1 - p(t)$ and then create the random graph $\mathscr{G}_n(p(t))$ by performing uniform matching with this percolated degree sequence. This will then be used to derive bounds on the maximal component and diameter in $\mathrm{CM}_n(t)$. Then in Section 8.7 we will use these results in the other direction, using the established results on $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$ to then read off results for percolation on $\mathrm{CM}_n(\infty)$.

8.2. **Bounds on the maximal diameter and component size.** In this section we will prove Theorem 4.11. We start with a result about the configuration model constructed from a prescribed degree sequence at a fixed time and then describe how this can be used to prove Theorem 4.11 that proves uniform bounds over all times before time $t_c - n^{-\delta}$. We first need some notation. Fix $\delta < 1/4$ and $\alpha > 0$. For each vertex, retain each half-stub attached to it with probability p and remove it with probability $1 - p$ where

$$p = \frac{1}{\nu} - a, \qquad \text{where } \frac{\alpha}{n^{\delta}} < a < \frac{4C_2}{\lambda \nu \log n}, \tag{8.5}$$

where $C_2 = \sigma_3/\sigma_1$. Now let $\mathscr{G}_n(p)$ denote the configuration model formed with the above percolated degree sequence. Abusing notation, let $\mathscr{C}_n^{(1)}(p)$ and $\mathrm{diam}(p)$ denote the size of the largest component and diameter of $\mathscr{G}_n(p)$.

Theorem 8.5. *For p as in (8.5) and fixed $\kappa > 0$, there exists $\beta = \beta(\kappa)$ and $n_0 = n_0(\kappa)$ independent of the choice of a such that for all $n \geqslant n_0(\kappa)$,*

$$\mathbb{P}\left(\left\{\mathscr{C}_n^{(1)}(p) \geqslant \frac{\beta(\log n)}{a^2}\right\} \cup \left\{\mathrm{diam}(p) \geqslant \frac{\beta(\log n)}{a}\right\}\right) \leqslant n^{-\kappa},$$

Remark 5. Note that in Theorems such as the above, the constants n_0 and β will obviously depend on the degree distribution $\mathbf{p}$ and the various parameters of this distribution such as μ, ν etc. However for the rest of the proof we assume that the degree sequence is fixed and has good properties (Assumption 8.1) and derive error bounds which are uniform in the range of p of interest.

Proof of Theorem 4.11: Assuming Theorem 8.5 let us complete the proof of the bounds on the maximal component size and diameter continuous time version in the barely subcritical regime namely Theorem 4.11. The proof of Theorem 8.5 is given in the next section. To ease notation we show how Theorem 8.5 implies there exists $\beta > 0$ such that

$$\mathbb{P}\left(\exists\, t \in [0, t_n] \text{ such that } \mathscr{C}_1(t) \geqslant \frac{\beta \log^2 n}{(t_c - t)^2}\right) \to 0, \tag{8.6}$$

as $n \to \infty$. The same proof but now applied to the diameter instead of size of the largest component completes the proof of the Theorem.

First note that since the degrees of the vertices are generated in an iid fashion from a distribution having exponential tails (Assumption 4.6), they satisfy Assumption 8.1 a.s. for all large n, without loss of generality, for the rest of the proof we work with a deterministic degree sequence satisfying Assumption 8.1. Let $\sum_{i\in[n]} d_i := n\mu_n$. Fix $\beta > 0$ and let

$$m(\beta; n, t) := \frac{\beta \log^2 n}{(t_c - t)^2}$$

We start with the following Proposition.

Proposition 8.6. *Under Assumptions* 8.1 *and for* $\delta < 1/4$ *we have*

(a) *There exists a constant* C *such that for any* $\beta > 0$,

$$\mathbb{P}\left(\mathscr{C}_1(t) \geqslant 2m(\beta; n, t) \text{ for some } t \in [0, t_n]\right) \leqslant Cn^2 \sup_{0\leqslant t\leqslant t_n} [m(\beta; n, t)(\lambda \log n)^2]\, \mathbb{P}(\mathscr{C}_1(t) > m(\beta; n, t)).$$

(b) *Fix* $\kappa > 0$. *Let* C_2 *as in* (8.5). *The there exists a constant* $\beta = \beta(\kappa)$ *such that uniformly for all time* t *with*

$$t_* := t_c - \frac{8(\nu - 1)C_2}{\lambda \nu^2} \frac{1}{\log n} \leqslant t < t_c - \frac{1}{n^\delta},$$

we have

$$\mathbb{P}\left(\mathscr{C}_1(t) > \frac{\beta \log n}{(t_c - t)^2}\right) \leqslant \frac{1}{n^\kappa}. \tag{8.7}$$

Note that this completes the proof of Theorem 4.11 since we first fix $\kappa > 2 + 2\delta$ and then choose $\beta = \beta(\kappa)$ such that (8.7) holds with the chosen κ. Then choosing $\beta' > \beta$ appropriately we have for $t < t_*$,

$$\mathbb{P}(\mathscr{C}_1(t) > m(\beta'; n, t)) \leqslant \mathbb{P}(\mathscr{C}_1(t_*) > \beta \log n) \leqslant \frac{1}{n^\kappa},$$

by (8.7). Now using part(a) of the Proposition, the choice of κ, and the fact that $m(\beta; n, t) \leqslant \beta n^{2\delta} \log^2 n$ completes the proof of (8.6) and thus the Theorem.

Proof of Proposition 8.6: Let us first prove (a). Note that the total rate of edge formation at any time $t \geqslant 0$ is bounded by $n\mu_n$. Let $\mathscr{N}_n = (\tau_1, \tau_2, \ldots)$ be a rate $n\mu_n$ Poisson process and let $N(\cdot)$ be the associated counting process. Let $\left\{\mathrm{CM}_n^{(i)}(t), t \geqslant 0, i \geqslant 1\right\}$ be an iid family of continuous time constructions of the configuration model using the same degree sequence $\mathbf{d}$. Write $\mathrm{First}_n(t)$, respectively $\mathrm{First}_n^{(i)}(t)$ for the component of the containing the vertex whose stub was the first to have run and formed a full edge in CM_n, respectively $\mathrm{CM}_n^{(i)}$. Consider the new (hyper)-graph

$$\mathbf{CM}_n^t = \cup_{\tau_i \leqslant t} \mathrm{First}_n^{(i)}(t).$$

Then note that for any deterministic function $\alpha : [0, t_c] \to \mathbb{R}_+$

$$\begin{aligned}
&\mathbb{P}(\mathscr{C}_1(t) > \alpha(t) \text{ for some } t \in [0, t_n]) \\
&\qquad \leqslant \sum_{k=1}^{\infty} \mathbb{P}(\mathrm{First}_n^{(i)}(t) > \alpha(t) \text{ for some } t \in [0, t_n], \text{ and } i \leqslant k)\, \mathbb{P}(N(t_n) = k) \\
&\qquad \leqslant \sum_{k=1}^{\infty} k\, \mathbb{P}(\mathrm{First}_n(t) > \alpha(t) \text{ for some } t \in [0, t_n])\, \mathbb{P}(N(t_n) = k) \\
&\qquad \leqslant n\mu_n t_c\, \mathbb{P}(\mathrm{First}_n(t) > \alpha(t) \text{ for some } t \in [0, t_n])
\end{aligned} \tag{8.8}$$

Next let us show that for any continuous and increasing function α with $\alpha(0) > 1$,

$$\mathbb{P}(\mathrm{First}_n(t) > \alpha(t) \text{ for some } t \in [0, t_n]) \leqslant Cn[\lambda \log n]^2 \sup_{0 \leqslant s \leqslant t_n} \mathbb{P}(\mathrm{First}_n(s) > \alpha(s)), \tag{8.9}$$

where the constant C is independent of the function α. Combining (8.9) and (8.8) and using $\mathbb{P}(\mathrm{First}_n(s) > \alpha(s)) \leqslant \mathbb{P}(\mathscr{C}_1(s) > \alpha(s))$ for any s completes the proof of part(a) of the Proposition. To prove (8.9) we will need some notation. For fixed time t, write the components of $\mathrm{CM}_n(t)$ as $\{\mathscr{C}_n^s(t) : s \leqslant t\}$, where we label each component according to the time s of the first half-stub in that component to have rung and formed a full edge (the "originator" of that component). Abusing notation write, $\mathrm{First}_n(t) = \mathscr{C}_n^0(t)$. Let $\tau = \inf\{t \geqslant 0 : \mathrm{First}_n(s) \geqslant 2\alpha(t)\}$. Write $\mathscr{C}_n^0 \leftrightarrow_t \mathscr{C}_n^s$ for the event that a full edge is formed between the components $\mathscr{C}_n^0$ and $\mathscr{C}_n^s$ at time t. Then the event $\{\tau = t\}$ can be written as

$$\left\{\mathscr{C}_n^0(t-) < 2\alpha(t)\right\} \cap \left\{\mathscr{C}_n^0(t-) + \mathscr{C}_n^s(t-) \geqslant 2\alpha(t); \mathscr{C}_n^0 \leftrightarrow_t \mathscr{C}_n^s, \text{ for some} s < t\right\}.$$

Now note that

(i) The total rate of creation of edges by a half-edge ringing and forming a full edge at any time t is bounded by $n\mu_n$.

(ii) Conditional on a half-edge ringing at time instant t, the chance that this forms a connection $\mathscr{C}_n^0 \leftrightarrow_t \mathscr{C}_n^s \leqslant 2(\lambda \log n)^2 \mathscr{C}_n^0 \mathscr{C}_n^s / n\bar{s}_1(t)$ since the rate at which a half-edge completes a full edge between $\mathscr{C}_n^0$ and $\mathscr{C}_n^s$ is proportional to twice the number of **alive** half-stubs in $\mathscr{C}_n^0$ and $\mathscr{C}_n^s$ which are bounded by $\lambda \log n$ times the size of the component by our assumption on the maximal degree. Using $\mathscr{C}_n^0 \leqslant n$ and the fact that on the event of interest $\mathscr{C}_n^0 < 2\alpha(t)$ we get that the conditional on a half-edge ringing, the chance that this leads to a connection between between $\mathscr{C}_n^0$ and $\mathscr{C}_n^s$ is bounded by $4\alpha(t)[\lambda \log n]^2 / \bar{s}_1(t)$.

(iii) $\mathbb{P}(\mathscr{C}_n^0(t) + \mathscr{C}_n^s(t) \geqslant 2\alpha(t)) \leqslant 2\,\mathbb{P}(\mathscr{C}_n^0(t) \geqslant \alpha(t))$.

Combining and using Lemma 8.2 gives

$$\begin{aligned}
\mathbb{P}(\tau \leqslant t_n) &\leqslant 2e^{2t_c} \int_0^{t_n} n\mu_n 4\alpha(t)[\lambda \log n]^2 2\,\mathbb{P}(\mathscr{C}_n^0(t) \geqslant \alpha(t))\, dt \\
&\leqslant 16 e^{2t_c} \mu_n n [\lambda \log n]^2 \sup_{0 \leqslant t \leqslant t_n} \alpha(t)\, \mathbb{P}(\mathscr{C}_n^0(t) \geqslant \alpha(t)).
\end{aligned}$$

Using Assumptions 8.1 to replace μ_n by μ now completes the proof.

Proof of (b): It will be easier to reparametrize with $s = t_c - t$. Writing $C_* = 8(\nu-1)C_2/\lambda\nu^2$ then (b) is equivalent to showing that given any κ we can choose $\beta = \beta(\kappa)$ such that for all $n^{-\delta} < s < C_*/\log n$ we have

$$\mathbb{P}\left(\mathscr{C}_1(t_c - s) > \frac{\beta \log n}{s^2}\right) \leqslant \frac{1}{n^\kappa}.$$

Write $\mathscr{U}_n(t_c - s) = n\mu_n - n\bar{s}_1(t_c - s)$ for the number of half-edges *used up* by time $t_c - s$. By Lemma 8.2 we have for any $\gamma < 1/2$ with probability $\geqslant 1 - \exp(-Cn^{1-\gamma})$ we have

$$|\mathscr{U}_n(t_c - s) - n\mu\left(\frac{1}{\nu} - \frac{\nu - 1}{\nu}2s + O(s^2)\right)| \leqslant n^{1-\gamma}. \tag{8.10}$$

Further by symmetry, conditional on $\mathscr{U}_n(t_c - s)$ one obtains $\mathrm{CM}_n(t_c - s)$ by selecting $\mathscr{U}_n(t_c - s)$ of the $n\mu_n$ total half-edges and performing a perfect matching. Now consider the percolation model, the content of Theorem 8.5 where we first retain each half-edge with probability parameter $p(s)$ and delete it other wise where

$$p(s) = \frac{1}{\nu} - \frac{\nu - 1}{\nu}\frac{s}{8}.$$

Abusing notation and writing $\mathscr{U}_n(p(s))$ for the number of retained half-edges in this model Binomial tail bounds imply

$$\mathbb{P}\left(\mathscr{U}_n(p(s)) > n\mu\left(\frac{1}{\nu} - \frac{\nu - 1}{\nu}s\right)\right) \leqslant \exp(-Cn^{1-2\delta}), \tag{8.11}$$

where again C is independent of s. Now [34, 36] implies that for any fixed $l < k$, starting from the same degree sequence if one creates a random graph $\mathrm{CM}(l)$ by first selecting $2l$ of the available half edges and then creating a perfect matching amongst the selected half edges and similarly $\mathrm{CM}(k)$ then as random graphs we have $\mathrm{CM}(l) \leqslant_{st} \mathrm{CM}(k)$. Now using Theorem 8.5, (8.10) and (8.11), for appropriate choice of β (independent of s) we have

$$\mathbb{P}\left(\mathscr{C}_1(t_c - s) > \frac{\beta \log n}{s^2}\right) \leqslant n^{-\kappa} + \exp(-Cn^{1-2\delta}) + \exp(-Cn^{1-\gamma}).$$

This completes the proof. ■

8.2.1. **Proof of Theorem 8.5:** There are two steps in the construction of the random graph $\mathscr{G}_n(p)$: (a) perform percolation on the original degree sequence $\mathbf{d}(n)$ using half-edge rention probability p, and (b) construct the configuration model with the percolated degrees. Write $\{\tilde{d}_i : i \in [m]\}$ for the percolated degree sequence. Let us first show that the percolated degree sequence satisfies good properties and then show that conditional on these properties, the random graph satisfies the assertions of Theorem 8.5. Note that for a random variable $X \sim \mathrm{Bin}(d, p)$, the r-th factorial moment satisfies $\mathbb{E}((X)_r) = (d)_r p^r$. Since the percolated degrees satisfy $\tilde{d}_i \sim \mathrm{Bin}(d_i, p)$, the following easily follows from Azuma-Hoeffding using Assumption 8.1. We omit the proof.

Lemma 8.7. *Given $\kappa > 0$, there exists A independent of a in the range postulated by* (8.5) *such that for* $1 \leqslant r \leqslant 4$ *with probability greater than* $1 - 1/n^\kappa$ *we have*

$$\left|\frac{1}{n}\sum_i (\tilde{d})_r - \frac{1}{n}\sum_{i=1}^n (d_i)_r p^r\right| \leqslant \frac{A \log^4 n}{\sqrt{n}}.$$

In all the terms below, the constants will be independent of a. For the rest of the proof, we will assume the degree sequence satisfies Lemma 8.7 and work conditional on this degree sequence. As an immediate upshot, using Assumptions 8.1 on the original degree sequence, this implies that

$$\nu_p = \frac{\sum_{i=1}^n \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i=1}^n \tilde{d}_i} = p\nu + O\left(\frac{A\log^q n}{\sqrt{n}}\right) = (1 - a\nu) + O\left(\frac{A\log^q n}{\sqrt{n}}\right). \tag{8.12}$$

Further, there exist two constants $C_1, C_2 = \sigma_3/\sigma_1 > 0$ (here σ_i refer to the moments of the original unpercolated degree distribution) independent of a such that

$$\sum_i \tilde{d}_i > n\sigma_1/2\nu, \qquad C_1 < \frac{\sum_{i=1}^m (\tilde{d}_i - 2)^2 \tilde{d}_i}{\sum_i \tilde{d}_i} < C_2 \tag{8.13}$$

We now start the proof of Theorem 8.5 starting with the maximal component.

Analysis of the largest component: Fix $\kappa > 0$ as in Theorem 8.5 and choose $\beta > 0$ such that

$$\frac{\nu^2 \beta C_1}{8C_2^2} > \kappa, \qquad \frac{3\beta\lambda\nu}{512C_2^2} > \kappa. \tag{8.14}$$

We will show that this β works in the assertion of Theorem 8.5 for given κ. Note that for the configuration model, we can start from any vertex and construct the graph by starting at that vertex and exploring the component of the vertex chosen first. Pick a vertex $v(1)$ with probability proportional to the (percolated) degree $\{\tilde{d}_i : i \in [n]\}$ and let $\mathscr{C}_{v(1)}$ denote the component of this vertex. We will show the following.

Lemma 8.8. *With the choice of β in* (8.14) *we have*

$$\mathbb{P}\left(\mathscr{C}_{v(1)} \geqslant \frac{\beta(\log n)}{a^2}\right) \leqslant 1/n^\kappa.$$

Assuming this lemma, this completes the analysis of the maximal component size in Theorem 8.5 since

$$\begin{aligned}\mathbb{P}\left(\mathscr{C}_n^{(1)}(p) \geqslant \frac{\beta(\log n)}{a^2}\right) &\leqslant \left(\sum_i d_i\right) \sum_{i=1}^n \frac{\tilde{d}_i}{\sum_j \tilde{d}_j} \mathbb{P}\left(\mathscr{C}_i \geqslant \frac{\beta(\log n)}{a^2}\right), \\ &\leqslant 2\sigma_1 n \mathbb{P}\left(\mathscr{C}_{v(1)} \geqslant \frac{\beta(\log n)}{a^2}\right).\end{aligned}$$

Now to construct $\mathscr{C}_{v(1)}$, first select a vertex with probability proportional to its percolated degree $\tilde{d}$. Then sequentially attach all the half stubs $\tilde{d}_{v(1)}$ attached to this vertex by choosing available half-stubs uniformly at random. Each new vertex found is considered alive and all half-stubs of an alive . For understanding the size of the component the order in which neighbors of half stubs is unimportant though later, for the diameter, the breadth first attachment scheme will be used. Note that in the exploration process:

(a) Every time a half-stub selects a vertex not already in the active cluster the vertex is selected with probability proportional to the (percolated) degree. At this stage kill the two half edges that were merged to from the full edge. All *remaining* half-edges are now designated to be *alive*

(b) When an alive half-edge selects one of the alive half-edges then two half edges die (become inactive) and no new vertices are added to the cluster.

Write $S_n(i)$ for the number of alive half edges at step i, where $S_n(1) = \tilde{d}_{v(1)}$. Write $s(i) \leqslant i$ for the number of vertices found by the exploration process by time i. Then note that the transitions of this walk are

$$S_n(i+1) = \begin{cases} S_n(i) + \tilde{d}_{v(s(i)+1)} - 2 & \text{if a new vertex is found,} \\ S_n(i) - 2 & \text{if two alive half-edges are merged.} \end{cases} \tag{8.15}$$

Recall the notion of size-biased ordering a vertex set from Section 6.2.1. By construction $(v(1), v(2), \dots v(n))$ is a size-biased reordering of the vertices using the (percolated) degree sequence $\{\tilde{d}_i : i \in [n]\}$ as the vertex weights. Further note that if H denotes the first time that the walk above hits zero then the size of the component $\mathscr{C}_{v(1)} \leqslant H$. Consider the walk that ignores the transitions where we might merge two already alive half-edges (second case in (8.15)) namely $\tilde{S}_n(i) = 2 + \sum_{j=1}^{i} (\tilde{d}_{v(j)} - 2)$ and write $\tilde{H}$ for the corresponding hitting time of this process. By the description of the transitions in (8.15), we have $H \leqslant \tilde{H}$. Thus it is enough to show

$$\mathbb{P}\left(\tilde{H} > \frac{\beta(\log n)}{a^2}\right) \leqslant n^{-\kappa}. \tag{8.16}$$

For ease of notation let $m_n = \beta \log n / a^2$. To analyze this hitting time, we will use a reformulation of the problem using an artificial time parameter, see [5]. Let $\{\xi_i : i \in [n]\}$ be a collection of independent exponential random variables with ξ_i having rate $\tilde{d}_i / \sum_j \tilde{d}_j$. For $t \geqslant 0$, write

$$N_n(t) = \sum_{i=1}^{n} \mathbb{1}\{\xi_i \leqslant t\}.$$

As described in Section 6.2.1, for any $t > 0$, conditional on $N_n(t) = k$, the vertices with $\xi_{v(1)} < \xi_{v(2)} < \cdots \xi_{v(k)} < t$ are the first k terms in the size biased random re-ordering. For $m \geqslant 1$, define the stopping time $T_m = \inf\{t : N_n(t) = m\}$. Abusing notation write

$$\tilde{S}_n(t) = 2 + \sum_{i=1}^{n} (\tilde{d}_i - 2) \mathbb{1}\{\xi_i \leqslant t\}, \qquad t \geqslant 0.$$

We will ignore the term 2 for simplicity, it will not play a role in the analysis. Let $t_n = m_n/2$. Then

$$\begin{aligned} \mathbb{P}(\tilde{H} > m_n) &\leqslant \mathbb{P}(\tilde{S}_n(t) > 0 \text{ for all } 0 < t < t_n, T_{m_n} > t_n) + \mathbb{P}(T_{m_n} < t_n), \\ &\leqslant \mathbb{P}(\tilde{S}_n(t_n) > 0) + \mathbb{P}(T_{m_n} < t_n). \end{aligned} \tag{8.17}$$

Let us analyze the first term. First note that the expectation satisfies

$$\begin{aligned} \mathbb{E}(\tilde{S}_n(t_n)) &= \sum_{i=1}^{n} (\tilde{d}_i - 2)(1 - \exp(-t_n \tilde{d}_i / \sum_j \tilde{d}_j)) \\ &= t_n \frac{\sum_i (\tilde{d}_i)(\tilde{d}_i - 2)}{\sum_i \tilde{d}_i} + O\left(\frac{n^{4\delta} (\log n)^2}{n}\right) \\ &= -t_n a \nu + O\left(\frac{A \log^q n}{\sqrt{n}}\right) \end{aligned} \tag{8.18}$$

where the last line follows using (8.12) and the condition $\delta < 1/4$. Also note that $S_n(t_n) = \sum_i Y_i$ where $Y_i = \tilde{d}_i \mathbb{1}\{\xi_i \leqslant t_n\}$ are a collection of independent random variables where using Assumptions 8.1, $|Y_i| \leqslant \lambda \log n$ and using (8.13),

$$C_1 t_n < \sum_i \mathbb{E}(Y_i^2) \leqslant C_2 t_n$$

In particular, using Bennet's inequality [23] we get

$$\begin{aligned}\mathbb{P}(\tilde{S}_n(t_n) > 0) &= \mathbb{P}(\tilde{S}_n(t_n) - \mathbb{E}(\tilde{S}_n(t_n)) > -\mathbb{E}(\tilde{S}_n(t_n))) \\ &\leqslant \exp\left(-\frac{C_1 t_n}{(\lambda \log n)^2} h\left(\frac{\lambda \log n |\mathbb{E}(\tilde{S}_n(t_n))|}{C_2 t_n}\right)\right),\end{aligned}$$

where $h(u) = (1+u)\log(1+u) - u$. Now using (8.18) for $\mathbb{E}(\tilde{S}_n(t_n))$ we have

$$h\left(\frac{\lambda \log n |\mathbb{E}(\tilde{S}_n(t_n))|}{C_2 t_n}\right) = h\left(\frac{\lambda a \nu \log n}{C_2}\right)$$

By the choice of the range of a in (8.5), we have $\frac{\lambda a \nu \log n}{C_2} \leqslant 4$. Further for $u \leqslant 4$, $h(u) \geqslant u^2/4$. Thus we get

$$\mathbb{P}(\tilde{S}_n(t_n) > 0) \leqslant \exp\left(-\frac{\nu^2 \beta C_1}{8 C_2^2} \log n\right) \leqslant n^{-\kappa},$$

by our choice of β in (8.14).

Let us now analyze the second term in (8.17). Note that $T_{m_n} < t_n$ implies that $N_n(m_n/2) > m_n$. Note that $N_n(m_n/2)$ is a sum of independent indicators and in particular is a self-bounding function. Further

$$\mathbb{E}(N_n(m_n/2)) = \frac{m_n}{2} + O\left(\frac{n^{4\delta} \log^2 n}{n}\right)$$

Using concentration inequalities for self-bounding functions [23, Theorem 6.12] we get

$$\mathbb{P}(N_n(m_n/2) > m_n) \leqslant \exp\left(-\frac{3 m_n}{16}\right) < n^{-\kappa},$$

again by the choice of β in (8.14) and the bounds on a. This completes the proof for the maximal component.

Analysis of the diameter: Arguing as for the maximal component, it is enough to work with the component of a vertex chosen according to the size biased distribution and show that one can choose β' (independent of a) such that

$$\mathbb{P}(\mathrm{diam}(\mathscr{C}_{v(1)}) \geqslant \beta' \log n / a) \leqslant n^{-\kappa}.$$

We will see that $\beta' = 2(\kappa + 3\delta)$ where $\delta < 1/4$ is as in (8.5) works. Fix $\kappa' > \kappa + 2\delta$ and using the previous analysis on the size of the component, choose $\beta = \beta(\kappa)$ so that

$$\mathbb{P}(\mathscr{C}_{v(1)} > \beta \log n / a^2) \leqslant n^{-(\kappa'+1)}. \tag{8.19}$$

Now consider the previous construction of the component but in this case we consider the breadth first construction where at each stage we look at all stubs in generation r and sequentially (in an arbitrary order) match these to available half-stubs until all the half stubs in generation r are matched. If they are matched to a new vertex v then they create a vertex with $\tilde{d}_v - 1$ children. Write $\{\mathscr{F}_r : r \geqslant 0\}$ for the natural filtration of the process and

G_r for the number of half-stubs in generation r with G_r obviously adapted to $\mathscr{F}_r$. For some constant $C > 0$, suppose we can show that for each $r \leqslant (\kappa + 3\delta)\log n/a$, there exists a set $A_r \in \mathscr{F}_r$

(a) On the set A_r we have,

$$\mathbb{E}(G_{r+1}|\mathscr{F}_r) \leqslant (1-a)\left(1 + \frac{C(\log n)^2}{na^2}\right).$$

(b) We have $\mathbb{P}(A_r^c) \leqslant n^{-(\kappa'+1)}$. On this set we use the trivial bound $\mathbb{E}(G_r|\mathscr{F}_r) \leqslant n$.

Then we get the recursion

$$\mathbb{E}(G_{r+1}) \leqslant (1-a)\left(1 + \frac{C(\log n)^2}{na^2}\right)\mathbb{E}(G_r) + \frac{1}{n^{\kappa'}}$$

Iterating this starting at $r_n = (\kappa + 3\delta)\log n/a$ gives

$$\begin{aligned}\mathbb{E}(G_{r_n}) &\leqslant \left[(1-a)\left(1 + \frac{C(\log n)^2}{na^2}\right)\right]^{r_n} + \frac{1}{n^{\kappa'}}\frac{1}{1-(1-a)(\frac{C(\log n)^2}{na^2})} \\ &\leqslant \left(1 + \frac{C\log^3 n}{n^{1-3\delta}}\right)e^{-r_n a} + O(\frac{1}{n^{\kappa'-\delta}}) \leqslant \frac{1}{n^{\kappa}},\end{aligned} \tag{8.20}$$

by our choices of κ' and r_n and the assumption that $\delta < 1/4$. Note that

$$\mathbb{P}(\operatorname{diam}(\mathscr{C}_{v(1)}) > 2r_n) \leqslant \mathbb{E}(G_{r_n}) \leqslant n^{-\kappa},$$

by (8.20) and this completes the proof. So we need to show the two assertions. The definition of the set A_r is almost obvious,

$$A_r = \left\{\sum_{j=1}^{r} G_j \leqslant \beta\lambda(\log n)^2/a^2\right\}.$$

Since by Assumption 8.1, the maximum number of stubs of any vertex is $\lambda\log n$, by (8.19) we get $\mathbb{P}(A_r^c) \leqslant n^{-\kappa'+1}$. Now let us prove the assertion about conditional expectation. Let $\mathscr{A}_r$ denote the set of all vertices that the exploration process has reached in $\mathscr{F}_r$ and let $\mathscr{D}_r = [n]\setminus\mathscr{A}_r$ be the remaining vertices. Further note that each half-stub in G_r either connects to one of the other half-stubs in G_r or to one of the vertices in $\mathscr{D}_r$. As we sequentially make the connections of the half-stubs in G_r, every new vertex added is added through the size-biased distribution. We start with the following elementary Lemma which we give without proof.

Lemma 8.9. *Given a finite set of elements $\mathscr{D}$ and an associated set of positive weights $\{w_v : v \in \mathscr{D}\}$, consider the size-biased ordering of the elements as $(w_{v(1)}, w_{v(2)}, \ldots, w_{v(|\mathscr{D}|)})$. Then for any $k \geqslant 1$ we have $w_{v(k)} \leqslant_{st} w_{v(1)}$ where $\leqslant_{st}$ denotes stochastic domination. In particular $\mathbb{E}(w_{v(1)}) \geqslant \mathbb{E}(w_{v(k)})$.*

Now proceeding with the proof, using the above Lemma conditional on $\mathscr{F}_r$ immediately gives

$$\mathbb{E}(G_{r+1}|\mathscr{F}_r) \leqslant \alpha_r G_r,$$

where $\alpha_r = \mathbb{E}(d_{v_r(1)} - 1|\mathscr{F}_r)$, where $v_r(1)$ is a vertex selected via size biased sampling from $\mathscr{D}_r$.

$$\mathbb{E}(d_{v_r(1)} - 1|\mathscr{F}_r) = \frac{\sum_{i=1}^n \tilde{d}_i(\tilde{d}_i - 1) - \sum_{v\in\mathscr{F}_r} \tilde{d}_v(\tilde{d}_v - 1)}{\sum_{i=1}^n \tilde{d}_i - \sum_{v\in\mathscr{F}_r} \tilde{d}_v} \leqslant \frac{\sum_{i=1}^n \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i=1}^n \tilde{d}_i - \sum_{v\in\mathscr{F}_r} \tilde{d}_v}$$

On the set G_r, using (8.12) and the first assertion in (8.13) implies that

$$\alpha_r \leqslant (1 - a\nu)(1 + C(\log n)^2/na^2).$$

for an appropriate constant C. This completes the proof. ■

8.3. **Properties at the entrance boundary.** The aim of this section is to prove Theorem 4.10 on the susceptibility functions at the entrance boundary. Note that this result describes the scaling of these functions at the fixed time t_n. The idea behind the proof is as follows: we will in fact study the evolution of these functions for all time $t \leqslant t_n$. We will study the cumulative changes in these functions as the process $\mathrm{CM}_n(\cdot)$ evolves. Using semi-martingale approximation techniques we will show that owing to the maximal component size and diameter bound established in the previous Section, these random functions stay close to deterministic trajectories all the way till time t_n. The deterministic limits satisfy the assertions of Theorem 4.10 and the approximation result completes the proof.

We will first need some notation. Let $\mathscr{F} := \{\mathscr{F}_t : t \geqslant 0\}$ denote the natural filtration of $\{\mathrm{CM}_n(t) : t \geqslant 0\}$. For a $\mathscr{F}$- adapted semimartingale $J(t)$ of the form

$$dJ(t) = \alpha(t)dt + dM(t), \qquad \langle M, M\rangle(t) = \int_0^t \gamma(s)ds. \tag{8.21}$$

Write $\mathbf{d}(J)(t) := \alpha(t)$, $\mathbf{v}(J)(t) := \gamma(t)$ and $\mathbf{M}(J)(t) = M(t)$. We also write

$$\mathbb{E}[\Delta J \mid \mathscr{F}_t] = \mathbf{d}(J)(t)\Delta t, \qquad \mathbb{E}[(\Delta J)^2 \mid \mathscr{F}_t] = \mathbf{v}(J)(t)\Delta t,$$

where $\Delta J(t) = J(t+\Delta t) - J(t)$. For fixed $n \geqslant 1$, time $T > 0$, a non-negative stochastic process $\{\xi(t)\}_{0\leqslant t\leqslant T}$, and a deterministic constant $\alpha(n)$, a term of the form $O_T(\xi(t)\alpha(n))$ represents a stochastic process $\{\varepsilon(t)\}_{0\leqslant t\leqslant T}$ such that there is a constant $d_1 > 0$, independent of n such that for all $0 \leqslant t \leqslant T$, $\varepsilon(t) \leqslant d_1\xi(t)\alpha(n)$.

Now recall that $f_i(t)$ denoted the number of free half-edges in $\mathscr{C}_i(t)$. Define

$$I(t) := \lambda \log n|\mathscr{C}_1(t)|, \qquad \mathscr{S}_k(t) = \sum_i [f_i(t)]^k, \tag{8.22}$$

where λ is as in Assumption 8.1 with $\max\deg(i) \leqslant \lambda \log n$. Now recall the functions $\bar{s}_l(t), \bar{g}(t)$ and $\bar{\mathscr{D}}(t)$ from (4.11). The final function we will need later in the proof is the actual component susceptibility

$$\bar{s}_2^\star(t) := \frac{1}{n}|\mathscr{C}_i(t)|^2, \qquad t \geqslant 0. \tag{8.23}$$

Lemma 8.10. *The processes $\bar{s}_2, \bar{s}_3, \bar{g}, \bar{\mathscr{D}}$ and $\bar{s}_2^\star$ are $\mathscr{F}$-semi-martingales as in* (8.21) *with the following decompositions*

(a) $\mathbf{d}(\bar{s}_2)(t) = F_2^{\mathbf{s}}(\bar{s}_1(t), \bar{s}_2(t)) + O_{t_c}(I^2(t)\bar{s}_2(t)/n\bar{s}_1(t))$ *where*

$$F_2^{\mathbf{s}}(s_1, s_2) := \frac{1}{s_1}\left[2s_2^2 + 4s_1^2 - 8s_2 s_1\right]. \tag{8.24}$$

(b) $\mathbf{d}(\bar{s}_3)(t) = F_3^{\mathbf{s}}(\bar{s}_1(t), \bar{s}_2(t), \bar{s}_3(t)) + O_{t_c}(I^3(t)\bar{s}_2/n\bar{s}_1(t))$ *where*

$$F_3^{\mathbf{s}}(s_1, s_2, s_3) := \frac{s_2}{s_1}(6s_3 - 12s_2) + (24s_2 - 12s_3 - 8s_1). \tag{8.25}$$

(c) $\mathbf{d}(\bar{g})(t) = F^{\mathbf{g}}(\bar{s}_1(t), \bar{s}_2(t), \bar{g}(t)) + O_{t_c}(I^2(t)\bar{s}_2(t)/n\bar{s}_1(t))$ *where*

$$F^{\mathbf{g}}(s_1, s_2, g) := \frac{1}{s_1}\left[2gs_2 - 4gs_1\right]. \tag{8.26}$$

(d) $\mathbf{d}(\bar{\mathscr{D}})(t) = F^{\mathbf{d}}(\bar{s}_1(t), \bar{s}_2(t), \bar{\mathscr{D}}(t)) + O_{t_c}(\mathrm{diam}_{\max}(t)I^2(t)\bar{s}_2(t)/n\bar{s}_1(t))$ *where*

$$F^{\mathbf{d}}(s_1, s_2, d) := \frac{1}{s_1}\left[4ds_2 + 2s_2^2 - 4ds_1 - 4s_2s_1 + 2s_1^2 - 4ds_1\right] \tag{8.27}$$

(e) $\mathbf{v}(\bar{s}_2)(t) = O_{t_c}(I^2(t)\bar{s}_2^2(t)/n\bar{s}_1(t))$.

(f) $\mathbf{v}(\bar{\mathscr{D}})(t) = O_{t_c}(\bar{s}_2^2 I^2(\mathrm{diam}_{\max}(t))^2/n\bar{s}_1(t))$.

(g) $\mathbf{d}(\bar{s}_2^{\star})(t) = F_2^{\star}(\bar{s}_1(t), \bar{g}(t)) + O_{t_c}(I^2(t)\bar{s}_2(t)/n\bar{s}_1(t))$ *where*

$$F_2^{\star}(s_1, g) := \frac{2g^2}{s_1}. \tag{8.28}$$

Proof: We will prove assertions (a), (d) and (f) above. All the remaining results follow in an identical fashion. We start with (a). First note that for $i \neq j$, a clock rings in component $\mathscr{C}_i(t)$ at rate $f_i(t)$ and then this half edge decides to connect to a vertex in component j with probability $f_j(t)/n\bar{s}_1(t)$. Thus the change

$$\Delta\bar{s}_2(t) = \frac{1}{n}[(f_i(t) + f_j(t) - 2)^2 - f_i^2(t) - f_j^2(t)] \qquad \text{at rate } f_i(t)f_j(t)/ns_1(t).$$

For $j = i$, the rate of an alive half-edge in $\mathscr{C}_i(t)$ ringing and then connecting to a half-edge in the same component occurs at rate $f_i(t)(f_i(t) - 1)/n\bar{s}_1(t)$. In this case the change

$$\Delta\bar{s}_2(t) := \frac{(f_i(t) - 2)^2 - f_i^2(t)}{n} = \frac{4(1 - f_i(t))}{n}.$$

Summing over i, j and collecting terms we get

$$\begin{aligned}
\mathbf{d}(\bar{s}_2)(t) &= F_2^{\mathbf{s}}(\bar{s}_1(t), \bar{s}_2(t)) - \frac{1}{n^2\bar{s}_1(t)}\left[2\sum_i f_i^4(t) + 4\sum_i f_i^2(t) + 8\sum_i f_i^3(t) + 4\sum_i f_i(t)(f_i(t) - 1)^2\right] \\
&= F_2^{\mathbf{s}}(\bar{s}_1(t), \bar{s}_2(t)) - \varepsilon_2(t),
\end{aligned}$$

where

$$\varepsilon_2(t) \leqslant \frac{28\mathscr{S}_4(t)}{\bar{s}_1(t)n^2} \leqslant \frac{28I^2(t)\bar{s}_2(t)}{n\bar{s}_1(t)}.$$

This completes the proof of (a).

To prove (d), for simplicity for the rest of the proof write $\mathscr{C}_i = \mathscr{C}_i(t)$. Fix $t > 0$ and two components $\mathscr{C}_i(t) \neq \mathscr{C}_j(t)$ and two alive half edges $e_0 \in \mathscr{C}_i(t)$ and $f_0 \in \mathscr{C}_j(t)$. The rate at which half-edge e_0 rings and forms a full edge by connecting to half edge f_0 is $1/n\bar{s}_1(t)$. Recall from Section 4.2 that for an alive half-edge e we used $\mathscr{D}(e)$ to denote the sum of all distances of alive half-edges in the same component as e. Then the change in $\bar{\mathscr{D}}_1(t)$ under this event is

$$\begin{aligned}
n(\Delta\bar{\mathscr{D}}(t)) &= 2\sum_{\substack{e\in\mathscr{C}_i,\\ e\neq e_0}}\sum_{\substack{f\in\mathscr{C}_j,\\ f\neq f_0}}(d(e,e_0)+d(f,f_0)+1)-2\sum_{e\in\mathscr{C}_i}d(e_0,e)-2\sum_{f\in\mathscr{C}_j}d(f_0,f)\\
&=2\left[\sum_{e\in\mathscr{C}_i}\sum_{f\in\mathscr{C}_j}(d(e_0,e)+d(f,f_0)+1)-\sum_{e\in\mathscr{C}_i}(d(e,e_0)+1)-\sum_{f\in\mathscr{C}_j}(d(f,f_0)+1)+1\right]\\
&\quad -2\mathscr{D}(u)-2\mathscr{D}(v)\\
&=2\left[\mathscr{D}(u)f_j+f_i\mathscr{D}(v)+f_if_j-\mathscr{D}(u)-f_i-\mathscr{D}(v)-f_j+1\right]-2\mathscr{D}(u)-2\mathscr{D}(v).
\end{aligned}$$

where $f_i=f_i(t)$ denotes the number of free stubs in component. When the clock corresponding to an alive half-edge rings and it decides to connect to another alive half-edge in the same component then the change can be bounded by

$$n\Delta\bar{\mathscr{D}}(t)\leqslant 2f_i(t)(f_i(t)-1)d_{\max}(t).$$

Summing over all pairs of alive half edges and collecting terms shows that

$$\begin{aligned}
\mathbf{d}(\bar{\mathscr{D}})(t) &= F^{\mathbf{d}}(\bar{s}_1(t),\bar{s}_2(t),\bar{\mathscr{D}}(t))+O_{t_c}\left(d_{\max}(t)\frac{\sum_i f_i^4(t)}{n^2}\right)\\
&=F^{\mathbf{d}}(\bar{s}_1(t),\bar{s}_2(t),\bar{\mathscr{D}}(t))+O_{t_c}\left(\frac{I^4(t)\bar{s}_2(t)}{n}\right),
\end{aligned}$$

where in the last line we have used the fact that $d_{\max}(t)\leqslant I^2(t)$. This completes the proof. Finally to prove (f) note that if a half-stub in component i is merged with component j then the change in $\bar{\mathscr{D}}$ can be bounded by

$$\Delta\bar{\mathscr{D}}(t)\leqslant\frac{8f_i(t)f_j(t)\,\mathrm{diam}_{\max}(t)}{n}.$$

Thus

$$\mathbf{v}(\bar{\mathscr{D}})(t)\leqslant\frac{1}{n\bar{s}_1}\sum_{i,j}f_if_j\frac{(8f_i(t)f_j(t)d_{\max}(t))^2}{n^2}\leqslant\frac{64}{n\bar{s}_1}\bar{s}_3^2(\mathrm{diam}_{\max}(t))^2.$$

Using $\bar{s}_3\leqslant\bar{s}_2 I$ completes the proof. ■

Remark 6. Note that $\bar{s}_1(t)$ is a decreasing jump process which further by Lemma 8.2, whp for all $t<t_c$, $\mu^{-1}\leqslant 1/\bar{s}_1(t)\leqslant\mu^{-1}e^{2t_c}$. For the rest of the proof we will drop the term $\bar{s}_1$ in the error terms in the Lemma.

It turns out that one can explicitly solve the ODE's postulated in Lemma 8.10 and which we will prove below are in fact the limits of the susceptibility functions as $n\to\infty$. First note that by Lemma 8.2 for any $T>0$, $\sup_{0\leqslant t\leqslant T}|\bar{s}_1(t)-\mu\exp(-2t)|=O_P(n^{-\gamma})$ for any $\gamma<1/2$. Further note that $\bar{\mathscr{D}}(0)=0$ since at time zero we start with n disconnected vertices and by convention, the distance between any two half-stubs attached to the same vertex is zero. Further by Assumptions 8.1 on the degree sequence,

$$|\bar{s}_2(0)-(\nu+1)\mu|\leqslant\frac{\log^q n}{\sqrt{n}},\qquad|\bar{s}_3(0)-(\beta+\mu(3\nu+1))|\leqslant\frac{\log^q n}{\sqrt{n}}.\tag{8.29}$$

On the interval $[0,t_c)$ define the following functions:

$$s_2(t) = \frac{\mu e^{-2t}\left(-2\nu+(\nu-1)e^{2t}\right)}{-\nu+e^{2t}(\nu-1)}. \tag{8.30}$$

Let

$$s_3(t) = \frac{-\beta+e_3(t)}{[-\nu+(\nu-1)\exp(2t)]^3}, \tag{8.31}$$

where

$$\begin{aligned}e_3(t) = & -4\nu^3\mu - 9\nu^2\mu e^{2t} + 9\nu^3\mu e^{2t} - 6\nu\mu e^{4t} + 12\nu^2\mu e^{4t}\\ & -6\nu^3\mu e^{4t} - \mu e^{6t} + 3\mu\nu e^{6t} - 3\nu^2\mu e^{6t} + \nu^3\mu e^{6t}.\end{aligned}$$

Note that $e_3(t_c)=0$. Finally let

$$g(t) = \frac{\mu}{\nu-(\nu-1)e^{2t}}. \tag{8.32}$$

and

$$\mathscr{D}(t) := \frac{\nu^2\mu(1-e^{-2t})}{(\nu-(\nu-1)e^{2t})^2}. \tag{8.33}$$

Note the singularity of each of the above functions at t_c. The following lemma can be easily checked.

Lemma 8.11. *Writing $s_1(t)=\mu\exp(-2t)$, the functions s_2, s_3, g and $\mathscr{D}$ are the unique solutions on $[0,t_c)$ of the ODE's $s_2' = F_2^{\mathbf{s}}(s_1,s_2)$, $s_3' = F_3^{\mathbf{s}}(s_1,s_2,s_3)$, $g' = F^{\mathbf{g}}(s_1,s_2,g)$ and $\mathscr{D}' = F^{\mathbf{d}}(s_1,s_2,\mathscr{D})$ with boundary values $s_2(0)=(\nu+1)\mu$, $s_3(0)=\beta+\mu(3\nu+1)$, $g(0)=\mu$ and $\mathscr{D}(0)=0$. Further replacing $\bar{s}_2, \bar{s}_3, \bar{g}, \bar{\mathscr{D}}$ in Theorem* 4.10 *with $s_2, s_3, g, \bar{\mathscr{D}}$, the assertions of the theorem namely* (4.12), (4.13) *and* (4.14) *are satisfied for all $\delta < 1/3$.*

Thus to complete the proof of Theorem 4.10 it is enough to show that the susceptibility functions are close to their asserted limits. More precisely define the processes

$$Y(t) := \frac{1}{\bar{s}_2(t)}, \qquad Z(t) = \frac{\bar{s}_3(t)}{(\bar{s}_2(t))^3}, \qquad U(t) = \frac{\bar{g}(t)}{\bar{s}_2(t)}, \qquad V(t) := \frac{\bar{\mathscr{D}}(t)}{(\bar{s}_2(t))^2}.$$

Let $y(\cdot), z(\cdot), u(\cdot)$ and $v(\cdot)$ be the corresponding deterministic functions obtained using the asserted limiting functions namely $y(t) = 1/s_2(t)$, $z(t) = s_3(t)/(s_2(t))^3$ etc. Using the explicit forms of the functions above it is easy to check that

$$y(t) = \frac{2\nu}{\mu(\nu-1)}(t_c-t)(1+O(t_c-t)), \qquad z(t) = \frac{\beta}{\mu^3(\nu-1)^3}(1+O(t_c-t)), \text{ as } t\uparrow t_c, \tag{8.34}$$

and similarly

$$u(t) = \frac{1}{\nu-1}(1+O(t_c-t)), \qquad v(t) = \frac{\nu}{\mu(\nu-1)^2}(1+O(t_c-t)). \tag{8.35}$$

The following completes the proof of Theorem 4.10.

Proposition 8.12. *Fix $\delta \in (1/6, 1/5)$ and let t_n be as in* (4.10)*. Then*

$$\sup_{0\le t\le t_n} n^{1/3}\left|Y(t)-y(t)\right| \xrightarrow{\mathrm{P}} 0, \qquad \sup_{0\le t\le t_n} |V(t)-v(t)| \xrightarrow{\mathrm{P}} 0, \tag{8.36}$$

and

$$\sup_{0\le t\le t_n} \max(|Z(t)-z(t)|, |U(t)-u(t)|) \xrightarrow{\mathrm{P}} 0 \tag{8.37}$$

Proof: We will prove (8.36). Equation (8.37) follows in a similar manner. The main tool is the following result from [10].

Lemma 8.13 ([10, Lemma 6.10])**.** *Let $\{t_n\}$ be a sequence of positive reals such that $t_n \in [0, t_c)$ for all n. Suppose that $U^{(n)}$ is a semi-martingale of the form* (8.21) *with values in $\mathbb{D} \subset \mathbb{R}$. Let $g : [0, t_c) \times \mathbb{D} \to \mathbb{R}$ be such that, for some $C_4(g) \in (0,\infty)$,*

$$\sup_{t\in[0,t_c)} |g(t,u_1)-g(t,u_2)| \le C_4(g)|u_1-u_2|,\ u_1,u_2 \in \mathbb{D}. \tag{8.38}$$

Let $\{u(t)\}_{t\in[0,t_c)}$ be the unique solution of the differential equation

$$u'(t) = g(t, u(t)), \quad u(0) = u_0.$$

Further suppose that there exist positive sequences:

(i) $\{\theta_1(n)\}$ *such that, whp,* $|U^{(n)}(0) - u_0| \le \theta_1(n)$.

(ii) $\{\theta_2(n)\}$ *such that, whp,*

$$\int_0^{t_n} \left|\mathbf{d}(U^{(n)})(t) - g(t, U^{(n)}(t))\right| dt \le \theta_2(n).$$

(iii) $\{\theta_3(n)\}$ *such that, whp,* $\langle \mathbf{M}(U^{(n)}), \mathbf{M}(U^{(n)})\rangle_{t_n} \le \theta_3(n)$.

Then, whp,

$$\sup_{0\le t\le t_n} |U^{(n)}(t) - u(t)| \le e^{C_4(g)t_c}(\theta_1(n)+\theta_2(n)+\theta_4(n)),$$

where $\theta_4 = \theta_4(n)$ is any sequence satisfying $\sqrt{\theta_3(n)} = o(\theta_4(n))$.

Let us now proceed with the proof. We start with the semi-martingale decomposition of the processes $Y(\cdot)$ and $V(\cdot)$. Define the functions

$$F^{\mathbf{y}}(s_1, y) := -\frac{1}{s_1}\left[2 + 4s_1^2 y^2 - 8s_1 y\right] \tag{8.39}$$

$$F^{\mathbf{v}}(s_1, y, v) := \frac{1}{s_1}\left[2 - 8vs_1 - 4s_1 + 4s_1 y + 2y^2 - 4s_1^2 vy - 8v\right]. \tag{8.40}$$

Finally define

$$\varepsilon_n(t) = \frac{6I(t)}{n\bar{s}_2(t)}. \tag{8.41}$$

Lemma 8.14. (a) *For the process $Y(\cdot)$ with $F^{\mathbf{y}}$ as in* (8.39) *we have,*

$$\mathbf{d}(Y)(t) = F^{\mathbf{y}}(s_1(t), Y(t)) + O_{t_c}\left(\frac{I^2(t)Y(t)}{n(1-\varepsilon_n(t))}\right), \qquad \mathbf{v}(Y)(t) := O_{t_c}\left(\frac{I^2(t)Y^2(t)}{n}\right). \tag{8.42}$$

(b) *For the process $V(\cdot)$ with $F^{\mathbf{v}}$ as in* (8.40) *we have*

$$\mathbf{d}(V)(t) = F^{\mathbf{v}}(s_1(t), Y(t), V(t)) + O_{t_c}\left(\frac{\mathrm{diam}_{\max}(t)I^2(t)Y(t)}{n(1-\varepsilon_n(t))}\right), \tag{8.43}$$

$$\mathbf{v}(V)(t) = O_{t_c}\left(\frac{(\mathrm{diam}_{\max}(t))^2 I^2(t)Y^2(t)}{n}\right). \tag{8.44}$$

Proof: The proof follows along the lines of [10, Lemma 6.7]. We start with part (a). For the jumps of the process $Y(\cdot)$ we have

$$\Delta Y(t) = \frac{1}{\bar{s}_2 + \Delta \bar{s}_2} - \frac{1}{\bar{s}_2} = -\frac{\Delta \bar{s}_2}{\bar{s}_2^2} + \frac{(\Delta \bar{s}_2)^2}{\bar{s}_2^2(\bar{s}_2 + \Delta \bar{s}_2)}$$

Now note that if change happens owing to the merger of a half-stub in $\mathscr{C}_i(t)$ with a half stub in $\mathscr{C}_j(t)$ then it is easy to check that $\Delta \bar{s}_2(t) \geqslant 0$. If the change occurs owing to the merger in the same component say component i^* then

$$|\Delta \bar{s}_2(t)| = \left| \frac{(f_{i^*}(t) - 2)^2 - f_{i^*}^2(t)}{n} \right| \leqslant \frac{6I(t)}{n}.$$

Thus we have

$$\bar{s}_2 + \Delta \bar{s}_2(t) \geqslant \bar{s}_2(t)(1 - \varepsilon_n(t)).$$

Thus we get

$$\Delta Y(t) = -\frac{\Delta \bar{s}_2}{\bar{s}_2^2} + O_{t_c}\left(\frac{(\Delta \bar{s}_2)^2}{\bar{s}_2^3(1 - \varepsilon_n(t))} \right).$$

Using Lemma 8.10(a) and (e) yields the desired form of the infinitesimal mean $\mathbf{d}(Y)$. For the variance note that $(\Delta Y(t))^2 \leqslant (\Delta \bar{s}_2)^2 / \bar{s}_2^4$. Thus

$$\mathbf{v}(Y)(t) \leqslant \frac{\mathbf{v}(\bar{s}_2)(t)}{\bar{s}_2^4} = O_{t_c}\left(\frac{I^2(t) Y^2(t)}{n} \right).$$

This completes part (a). Let us now prove (b). Suppose at time t a full edge if formed by connecting components $\mathscr{C}_i(t)$ and $\mathscr{C}_j(t)$. Then

$$\Delta V = \Delta(\bar{\mathscr{D}} Y^2) = Y^2 \Delta \bar{\mathscr{D}} + 2\bar{\mathscr{D}} Y \Delta Y + \left[\bar{\mathscr{D}}(\Delta Y)^2 + 2Y \Delta Y \Delta \bar{\mathscr{D}} + \Delta \bar{\mathscr{D}}(\Delta Y)^2 \right]. \tag{8.45}$$

Since $|\Delta Y| \leqslant 2Y^2 f_i(t)|f_j(t)|/n$, $|\Delta \bar{\mathscr{D}}| \leqslant 6D|f_i||f_j|/n$ and $\bar{\mathscr{D}} \leqslant \mathrm{diam}_{\max}(t)\bar{s}_2$, the term in the squared brackets can be bounded by

$$|\bar{\mathscr{D}}(\Delta Y)^2 + 2Y \Delta Y \Delta \bar{\mathscr{D}} + \Delta \bar{\mathscr{D}}(\Delta Y)^2| = O(DY^3 (f_i(t))^2 (f_j(t))^2 / n^2). \tag{8.46}$$

Using the formulae for $\mathbf{d}(\bar{\mathscr{D}})(t)$ and $\mathbf{d}(Y)(t)$ from Lemma 8.10 and part(a) of this lemma implies

$$\mathbf{d}(V)(t) = F^{\mathbf{v}}(\bar{s}_1(t), Y(t), V(t)) + O_{t_c}\left(\frac{Y^2 \bar{s}_2 I^2 \,\mathrm{diam}_{\max}}{n} \right) + O_{t_c}\left(\frac{\bar{\mathscr{D}} I^2 Y^2}{n(1 - \varepsilon_n)} \right) + O\left(\mathrm{diam}_{\max} \frac{Y^3 \bar{s}_3^2}{n} \right)$$

where the three big-O terms comes from $\mathbf{d}(Y)(t)$, $\mathbf{d}(\bar{\mathscr{D}})(t)$ and (8.46) respectively. Using $Y\bar{s}_2 = 1$ and $\bar{\mathscr{D}} \leqslant \mathrm{diam}_{\max} \bar{s}_2$ and collecting the error terms gives (8.43).

To prove (8.44) using the expression of the jump in (8.45), some algebra gives that if the change occurs owing to an alive half-edge from $\mathscr{C}_i$ merging with an alive half-edge from component $\mathscr{C}_j$

$$|Y^2 \Delta \bar{\mathscr{D}} + 2\bar{\mathscr{D}} Y \Delta Y| \leqslant O_{t_c}\left(\frac{\mathrm{diam}_{\max} Y^2 f_i f_j}{n} \right), \tag{8.47}$$

while the remaining terms are of lower order. This gives

$$(\Delta V)^2 = O\left(\frac{\mathrm{diam}^2_{\max} Y^4 f_i^2 f_j^2}{n^2}\right).$$

Thus

$$\mathbf{v}(V)(t) = O(\mathrm{diam}^2_{\max} Y^4 \bar{s}_3^2/n) = O(\mathrm{diam}^2_{\max} I^2 Y^2/n).$$

This completes the proof of the Lemma. ■

Let us now proceed with the proof of Proposition 8.12. We start by proving the weaker result:

$$\sup_{0\leqslant t\leqslant t_n} |Y(t) - y(t)| = O(n^{-1/5}). \tag{8.48}$$

Note that by the bound on the maximal component, Theorem 4.11, Lemma 8.2 and the fact that $\bar{s}_2 \geqslant \bar{s}_1$ for all t, there exists $\beta > 0$ such that we have with high probability for all $t < t_n$

$$I(t) \leqslant \frac{\beta \log^2 n}{(t_c - t)^2}, \qquad |\bar{s}_1(t) - \mu \exp(-2t)| \leqslant \frac{\log^q n}{\sqrt{n}}, \qquad Y(t) \leqslant \frac{\exp(2t_c)}{\mu}. \tag{8.49}$$

In particular

$$\sup_{t\leqslant t_n} |F^{\mathbf{y}}(\bar{s}_1, Y(t)) - F^{\mathbf{y}}(s_1, Y(t))| = O\left(\frac{\log^q n}{\sqrt{n}}\right) \qquad , \sup_{t\leqslant t_n} \varepsilon_n(t) \to 0, \tag{8.50}$$

where ε_n is as in (8.41). Now we will use Lemma 8.13. It is easy to check that under (8.49), the function $g(t, y(t)) = F^{\mathbf{y}}(s_1(t), y(t))$ satisfies the Lipschitz condition in (8.38). By Assumptions 8.1,

$$|Y(0) - y(0)| \leqslant \frac{\log^q n}{\sqrt{n}}. \tag{8.51}$$

Using Lemma 8.14(a), (8.49) and (8.50) we have

$$\int_0^{t_n} |\mathbf{d}(Y)(t) - F^{\mathbf{y}}(s_1(t), Y(t))| dt = O\left(\frac{\log^q n}{\sqrt{n}} + \int_0^{t_n} \frac{\log^4 n}{n(t_c - t)^4}\right) = O\left(\frac{1}{n^{1-3\delta}}\right) = o(n^{-2/5}), \tag{8.52}$$

for $\delta < 1/5$. Finally using (8.42) we get

$$< \mathbf{M}(Y), \mathbf{M}(Y) >_{t_n} = O\left(\int_0^{t_n} \frac{I^2(t) Y^2(t)}{n} dt\right) = O\left(\int_0^{t_n} \frac{I^2(t)}{n} dt\right) = o(n^{-2/5}). \tag{8.53}$$

Now using the Semi-martingale approximation Lemma 8.13

$$\theta_1 = \frac{\log^q n}{\sqrt{n}}, \qquad \theta_2 = \theta_3(n) = \frac{1}{n^{1-3\delta}},$$

completes the proof of (8.48). We will now strengthen this estimate. For $\delta < 1/5$, using (8.34) we have $y(t_n) = \Theta(n^{-\delta})$, in particular for $\delta < 1/5$, using (8.48) we have whp for all $t \leqslant t_n$, $Y(t) \leqslant 2y(t_n)$. Redoing the above error bounds now allows us to replace (8.52) with

$$\int_0^{t_n} |\mathbf{d}(Y)(t) - F^{\mathbf{y}}(s_1(t), Y(t))| dt = O\left(\int_0^{t_n} \frac{I^2(t) y^2(t)}{n(t_c - t)^4}\right) = O\left(\frac{\log^4 n}{n^{\delta-1}}\right) = o(n^{-2/5}), \tag{8.54}$$

Similarly $< \mathbf{M}(Y), \mathbf{M}(Y) >_{t_n} = o(n^{-2/5})$. Since $s o(n^{-2/5}) = o(n^{-1/3})$, this completes the proof of the first part of (8.36). The second part namely showing

$$\sup_{0 \le t \le t_n} |V(t) - v(t)| \xrightarrow{\mathrm{P}} 0, \tag{8.55}$$

follows in an identical fashion now using the semi-martingale decomposition of V in Lemma 8.14 (b), the semi-martingale approximation Lemma 8.13 and the above concentration results of the process Y about the limit y. This completes the proof of Proposition 8.12 and thus Theorem 4.10.

■

We end this section with the following result on the *component susceptibility* $\bar{s}_2^\star$ as defined in (8.23). The proof is identical to that of Proposition 8.12 and we omit the proof. Recall the function $F_2^\star$ from Lemma 8.10 (g) in the semimartingale decomposition of $\bar{s}_2$. Consider the deterministic analogue, the function $s_2^\star$ that satisfies the differential equation

$$(s_2^\star)'(t) = F_2^\star(s_1(t), g(t)), \qquad s_2^\star(0) = 0,$$

where $s_1(\cdot)$ and $g(\cdot)$ are the limit functions in (8.1) and (8.32). This ODE can be explicitly solved on $[0, t_c)$ as

$$s_2^\star(t) := \left(1 - \frac{\mu}{\nu - 1}\right) + \frac{\mu}{(\nu-1)(\nu - e^{2t}(\nu - 1))}, \qquad 0 \le t < t_c.$$

Note that at $t_n = t_c - \nu/[2n^\delta(\nu - 1)]$

$$\left| \frac{s_2^\star(t_n)}{n^\delta} - \frac{\mu}{\nu^2} \right| \xrightarrow{\mathrm{P}} 0, \qquad \text{as } n \to \infty. \tag{8.56}$$

Lemma 8.15. *We have* $n^{-\delta}|\bar{s}_2^\star(t_n) - s_2^\star(t_n)| \xrightarrow{\mathrm{P}} 0$ *as* $n \to \infty$. *Thus using* (8.56)

$$\left| \frac{\bar{s}_2^\star(t_n)}{n^\delta} - \frac{\mu}{\nu^2} \right| \xrightarrow{\mathrm{P}} 0, \qquad \textit{as } n \to \infty.$$

8.4. **Modified process.** In this section, we will describe a modification of the original process which evolves like the multiplicative coalescent and study properties of this modification in the next section. Note that the previous section describes the evolution of various susceptibility functions till time t_n as defined in (4.10). Thus to get $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$ starting from $\mathrm{CM}_n(t_n)$, we need to run the dynamic construction for an additional

$$r_n(\lambda) := \frac{\nu}{2(\nu - 1)} \frac{1}{n^\delta} + \frac{\lambda}{n^{1/3}}, \tag{8.57}$$

units of time. Further for any $t \in (t_n, t_n + r_n(\lambda)]$ two components $\mathscr{C}_i(t)$ and $\mathscr{C}_j(t)$ form an edge between each other via a half edge from $\mathscr{C}_i(t)$ ringing to connect with a half-edge from j or vice-versa at rate

$$f_i(t)\frac{f_j(t)}{n\bar{s}_1(t) - 1} + f_j(t)\frac{f_i(t)}{n\bar{s}_1(t) - 1} \approx \frac{2\nu f_i(t) f_j(t)}{n\mu(\nu - 1)}, \tag{8.58}$$

where as before $f_i(t), f_j(t)$ denote the number of still free (alive) edges in the respective components and we have used Lemma 8.2 to approximate $\bar{s}_1$ by $s_1(t_c)$ uniformly in the interval $(t_n, t_n + r_n(\lambda)]$. Thus:

(a) Components merge at rate proportional not to the product of component sizes but rather the product of the number of still *free edges*.
(b) When two components merge, the weight of the new component is **not** $f_i(t)+f_j(t)$ as in the multiplicative coalescent but $f_i(t)+f_j(t)-2$ since two half-edges were used up to complete the edge and are now considered *dead* half-edges.

Using $s_1(t_c)=\mu(\nu-1)/\nu$, this suggests the following modification of the original process. Fix all the free edges $\mathrm{FR}(t_n)$ at time t_n. Define the modified process $\mathscr{G}_n^{\mathrm{modi}}$ as follows: For every ordered pair of free half edges $\mathbf{e}=(u,v)\in \mathrm{FR}(t_n)\times\mathrm{FR}(t_n)$, let $\mathscr{P}_{\mathbf{e}}$ be a Poisson process with rate $1/(n\mu(\nu-1)/\nu)$, independent across (ordered) pairs. Every time one of these ring, complete the edge corresponding to these 2 half edges but continue to consider them as "alive". Run this process for time $r_n(\lambda)$ and write $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ for the process observed after time $r_n(\lambda)$. Recall that connected components at time $\mathrm{CM}_n(t_n)$ are called"blobs". The rate of creation of edges between two blobs $\mathscr{C}_i(t_n)$ and $\mathscr{C}_j(t_n)$ in the modified process is given by

$$\frac{2}{n\mu}\frac{\nu}{\nu-1}f_i(t_n)f_j(t_n). \tag{8.59}$$

Compare this with (8.58). For simplicity, we will refer to the components $\mathscr{C}_i(t_n)$ and $\mathscr{C}_j(t_n)$ as blobs i and j and the corresponding number of free stubs as $f_i:=f_i(t_n)$. Then note that conditional on $\mathrm{CM}_n(t_n)$, to get the connectivity structure between blobs in $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ we connect blobs i and j with connection probability

$$p_{ij}=1-\exp\left(-f_if_j\left[\frac{1}{n^{1+\delta}}\frac{\nu^2}{\mu(\nu-1)^2}+\frac{1}{n^{4/3}}\frac{2\nu}{\mu(\nu-1)}\lambda\right]\right) \tag{8.60}$$

Remark 7. To simplify notation and reduce gargantuan expressions in the statement of the results, for the rest of the proof we will use $\mathscr{C}(\lambda)$ and $\mathscr{C}^{\mathrm{modi}}(\lambda)$ instead of $\mathscr{C}(t_c+\lambda/n^{1/3})$ and $\mathscr{C}^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ for connected components in $\mathrm{CM}_n(t_c+\lambda/n^{1/3})$ and $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ respectively.

8.5. **Properties of the modified process.** Using (8.60), the blob-level superstructure for $\mathscr{G}_n^{\mathrm{modi}}(\lambda)$, namely as in Section 3.2, the connectivity pattern when each component $\mathscr{C}_i(t_n)$ is viewed as a single vertex i and we use connection probabilities (8.60), has the same distribution as the random graph $\mathscr{G}(\mathbf{x},q)$ as considered by Aldous (see Section 2.4) where

$$x_i=\frac{\beta^{1/3}}{\mu(\nu-1)}\frac{f_i}{n^{2/3}},\qquad q=n^{1/3-\delta}\frac{\mu\nu^2}{\beta^{2/3}}+\frac{2\mu(\nu-1)\nu}{\beta^{2/3}}\lambda. \tag{8.61}$$

Note that at the vertex level, this corresponds to a rescaling of the following mass measure, assigning each *vertex* mass

$$\mu_{\mathrm{free}}(\{v\}):=\#\text{ of still free edges attached to } v \text{ at time } t_n,\quad v\in[n]. \tag{8.62}$$

8.5.1. *Component sizes counted according to number of free edges in $\mathscr{G}_n^{\mathrm{modi}}$.* The aim of this section is to use Theorem 2.2 to understand the maximal weighted components in $\mathscr{G}_n(\lambda)$. Here the **weight** of a component $\mathscr{C}\subset\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ is made of the number of free edges at time t_n in the blobs that make up $\mathscr{C}$ namely

$$\mathscr{W}(\mathscr{C})=\sum_{\mathrm{blob}\in\mathscr{C}}f_{\mathrm{blob}}(t_n) \tag{8.63}$$

Write $\mathfrak{C}_i^{\mathrm{modi}}(\lambda)$ for the i-th largest component in $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ where the size is counted as in (8.63). The number of vertices in a component $\mathscr{C} \subseteq \mathscr{G}_n^{\mathrm{modi}}(\lambda)$ is given by $|\mathscr{C}| = \sum_{i\in\mathscr{C}} |\mathscr{C}_i(t_n)|$. In the next section we will make rigorous the idea that for each fixed i, $\mathscr{C}_i(\lambda) \approx \mathfrak{C}_i(\lambda)$, thus reading off properties about the original model CM_n using the modified process $\mathscr{G}_n$modi. For the rest of this section we work with the modified process. To use Theorem 2.2, we will use Theorems 4.10 and 4.11 to verify Assumption 2.1. First note that

$$\frac{\sum_i x_i^3}{(\sum_i x_i^2)^3} = \frac{[\mu(\nu-1)]^3}{\beta}\frac{s_3(t_n)}{[s_2(t_n)]^3} \xrightarrow{\mathrm{P}} 1, \qquad \text{by (4.13)} . \tag{8.64}$$

Next using (4.12) and the definition of q from (8.61) we get

$$q - \frac{1}{\sum_i x_i^2} = q - \frac{(\mu(\nu-1))^2}{\beta^{2/3}}\frac{n^{1/3}}{s_2(t_n)} \xrightarrow{\mathrm{P}} \frac{2\mu\nu(\nu-1)}{\beta^{2/3}}\lambda \tag{8.65}$$

Finally Theorem 4.11, (4.12) and the assumption that $\delta \in (1/6, 1/5)$ gives

$$\frac{x_{\max}}{\sum_i x_i^2} = O_P\left(\frac{\log^2 n}{n^{1/3-\delta}}\right) \xrightarrow{\mathrm{P}} 0 \tag{8.66}$$

Using Theorem 2.2 now gives the following.

Proposition 8.16 (Mass of free weight maximal components)**.** *The maximal components in* $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ *where components are counted according to total number of free edges in constituent blobs at time* t_n *as in* (8.63) *satisfy*

$$\left(\frac{\beta^{1/3}}{\mu(\nu-1)}\frac{\mathscr{W}(\mathfrak{C}_i^{\mathrm{modi}}(\lambda))}{n^{2/3}} : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \boldsymbol{\xi}\left(\frac{2\mu\nu(\nu-1)}{\beta^{2/3}}\lambda\right), \tag{8.67}$$

where $\boldsymbol{\xi}$ *are the excursions away from zero of the reflected inhomogeneous Brownian motion as in Theorem* 2.2.

8.5.2. *Component sizes and continuum scaling limits of maximal components in* $\mathscr{G}_n^{\mathrm{modi}}$. The main aim of this section is to understand the number of vertices and scaling limits of the metric structure of the free-weight maximal components $\mathfrak{C}_i^{\mathrm{modi}}$ defined in the previous Section. We start with the following Proposition.

Proposition 8.17 (Number of vertices in free weight maximal components)**.** *When the components* $\mathfrak{C}_i^{\mathrm{modi}}(\lambda)$ *are counted according to number of vertices then we have (in terms of finite-dimensional convergence),*

$$\left(\frac{\beta^{1/3}}{\mu}\frac{|\mathfrak{C}_i^{\mathrm{modi}}(\lambda)|}{n^{2/3}} : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \boldsymbol{\xi}\left(\frac{2\mu\nu(\nu-1)}{\beta^{2/3}}\lambda\right), \tag{8.68}$$

where $\boldsymbol{\xi}(\cdot)$ *as in Theorem* 2.2 *denote the excursions of the reflected inhomogeneous Brownian motion from zero as in* (2.6).

Proof: Recall the proof of Proposition 6.7, in particular the second assertion about average distances within the maximal components. Here, we related the average distances $\sum_{v\in\mathscr{C}_1} x_v u_v$ within the maximal components with the *global distance* $\sum_{i\in[n]} x_i^2 u_i$, where the extra square arises due to the size-biased construction of $\mathscr{G}(\mathbf{x}, q)$ using the weight sequence $\mathbf{x}$. Here we use the exact same argument, where we construct the graph using a weight sequence $\mathbf{x}$ which is proportional to $\{f_i(t_n), i \in \mathrm{CM}_n(t_n)\}$ but unlike the distances, here we are interested in the true component size namely we replace average distances in blobs u_i with sizes of blobs $|\mathscr{C}_i(t_n)|$. The role of c_n^{-1} (see Lemma 6.8) is played by

$$\frac{\sum_i f_i^2(t_n)}{\sum_i f_i(t_n)|\mathscr{C}_i(t_n)|} = \frac{s_2(t_n)}{g(t_n)} \xrightarrow{\mathrm{P}} (\nu-1), \tag{8.69}$$

where the last assertion follows from (4.14). Arguing as in Proposition 6.7 now shows that for any fixed i

$$\frac{\mathscr{W}(\mathfrak{C}_i^{\mathrm{modi}}(\lambda))}{|\mathfrak{C}_i^{\mathrm{modi}}(\lambda)|} \cdot \frac{\sum_i f_i(t_n)|\mathscr{C}_i(t_n)|}{\sum_i f_i^2(t_n)} \xrightarrow{\mathrm{P}} 1. \tag{8.70}$$

Combining (8.69) and (8.70) completes the proof. ■

Now using Theorem 4.10 in particular the distance scaling result in (4.14) and Theorem 4.11 for bounds on the maximal component and diameter at time t_n and arguing as above, we check that Assumption 3.3 are met. Using Theorem 3.4 gives us the following result for $\mathscr{G}_n^{\mathrm{modi}}$.

Theorem 8.18. *The free weight maximal components* $(\mathfrak{C}_i^{\mathrm{modi}}(\lambda) : i \geqslant 1)$ *of* $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ *viewed as connected metric spaces with vertex set* $[n]$ *where we incorporate both the Blob-level superstructure and inter-blob distances as in Section* 3.2 *and equipped with mass measure* μ_{free} *as in* (8.62) *have the same scaling limits as those asserted for the maximal components of* $\mathrm{CM}_n(\lambda)$ *in Theorem* 4.9 *namely*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu}\frac{1}{n^{1/3}}, \frac{\beta^{1/3}}{\mu(\nu-1)n^{2/3}}\mu_{\mathrm{free}}\right)\mathfrak{C}_i^{\mathrm{modi}}(\lambda) : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty\left(\frac{2\nu(\nu-1)\mu}{\beta^{2/3}}\lambda\right), \qquad \text{as } n \to \infty.$$

Remark 8. Equipping the metric spaces above with the measure μ_{free} is a little unnatural in the context of random graphs. While this is enough to prove the scaling limits of just the metric structure of components at criticality, at the end of the next section (Theorem 8.26) we will show that μ_{free} can be replaced by an appropriately rescaled version of the counting measure. The proof relies on a construction closely related to one used for the configuration model in the next section, so we delay statement of the proof and the result.

Let us now summarize one of the repercussions of the above two results. To simplify notation define the constant arising in the limit on the right in both Proposition 8.17 and Theorem 8.18 as

$$\alpha = \alpha(\lambda) := \frac{2\nu(\nu-1)\mu}{\beta^{2/3}}. \tag{8.71}$$

As in (2.6) define

$$\tilde{W}_\alpha(t) := W_\alpha(t) - \inf_{s\in[0,t]} W_\alpha(s), \;\; W_\alpha(t) := B(t) + \alpha t - \frac{1}{2}t^2, \qquad t \geqslant 0,$$

Recall that $\boldsymbol{\xi}(\alpha)$ denotes the lengths of the excursions from zero of $\tilde{W}_\alpha$. Conditional on $\tilde{W}_\alpha$, for each $i \geqslant 1$, let $\mathrm{Pois}_i(\alpha)$ be a Poisson random variable with mean equal to the area underneath the i-th excursion of $\tilde{W}_\alpha$, independent across $i \geqslant 1$ (conditional on W_α). Write $N_i^{(n),\mathrm{modi}}(\lambda)$ for the number of surplus edges created in $\mathscr{G}_n^{\mathrm{modi}}(t_c + \lambda/n^{1/3})$ in the interval

$[t_n, t_n + r_n(\lambda)]$. Using the definition of the limiting metric spaces $\mathbf{Crit}_\infty(\alpha)$ in Theorem 8.18 (see Section 2) and the statement of Proposition 8.17 now gives the following Corollary.

Corollary 8.19. *The sizes and surplus of maximal components in $\mathscr{G}_n^{\mathrm{modi}}(\lambda)$ satisfy*

$$\left(\left[\frac{\beta^{1/3}}{\mu}\frac{|\mathfrak{C}_i^{\mathrm{modi}}(\lambda)|}{n^{2/3}}, N_i^{(n),\mathrm{modi}}(\lambda)\right] : i \geqslant 1\right) \xrightarrow{\mathrm{w}} ([\xi_i(\alpha(\lambda)), \mathrm{Pois}_i(\alpha(\lambda))] : i \geqslant 1) \tag{8.72}$$

as $n \to \infty$.

8.5.3. *Blob-level functionals of* CM_n. The final set of ingredients required to relate the original model to the modified process are various "gross" features of $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$, including the number of surplus edges created in the interval $[t_n, t_c + \lambda/n^{1/3}]$ as we move from the entrance boundary to the critical scaling window. We start by recalling known results about $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$. Let $N_i^{(n)}(\lambda)$ denote the number of surplus edges in $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$. As before let $|\mathscr{C}_i(\lambda)|$ denote the number of vertices in component $\mathscr{C}_i(\lambda)$.

Theorem 8.20 ([44, 54])**.** *The component sizes and number of surplus edges in* $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$ *satisfy the same result as that of the corresponding objects in the modified process (Corollary* 8.19*) namely*

$$\left(\left[\frac{\beta^{1/3}}{\mu}\frac{|\mathscr{C}_i(\lambda)|}{n^{2/3}}, N_i^{(n)}(\lambda)\right] : i \geqslant 1\right) \xrightarrow{\mathrm{w}} ([\xi_i(\alpha(\lambda)), \mathrm{Pois}_i(\alpha(\lambda))] : i \geqslant 1)\,.$$

Remark 9. The result for component sizes in the special case $\lambda = 0$ assuming only finite third moments for the degree distribution is shown in [44], while the result above for both component sizes and surplus for general λ, but where the all the degrees are assumed to be bounded by some $d_{\max} < \infty$ was shown in [54, Theorem 1.3]. We will in fact use the same construction as in [54] below but let us briefly comment on this assumption of bounded degree. A high level description of the proof in [54] is as follows. First a bound for the largest component in the **barely subcritical** configuration model, analagous to Theorem 8.5 (for the maximal component size, not diameter) is proved, see [54, Theorem 1.2]. Then in the critical regime, the graph is constructed via a breadth first exploration walk $\{Z_n(i) : i \geqslant 0\}$ which keeps track of components explored via times to reach beyond past minima. It is shown that this walk with time and space rescaled as $\{n^{-1/3} Z_n(sn^{2/3}) : s \geqslant 0\}$ converges to Brownian motion with parabolic drift as in (2.6). This part of the proof does not require finiteness of $d_{\max}$, just finite third moments suffice as it essentially following via the same arguments as in [5]. Then [54] uses the result for the barely subcritical regime to show for any fixed k, the k largest components $\{\mathscr{C}_i(\lambda) : 1 \leqslant i \leqslant \lambda\}$ are found by time $O_P(n^{2/3})$ whp and thus excursions of the $\tilde{W}_\alpha$ from zero **do** encode maximal component sizes of $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$ (renormalized by $n^{2/3}$). With our assumptions using exponential tails on the degree distribution as opposed to a uniform bound on the maximal degree, using Theorem 8.5 in place [54, Theorem 1.2] extends the results [54] with the same proof to our context.

Recall that we refer to the components at time t_n as "blobs". A connected component $\mathscr{C} \subseteq \mathrm{CM}_n(t_c + \lambda/n^{1/3})$ is made up a collection blobs connected via edges formed in the interval $[t_n, t_c + \lambda/n^{1/3}]$. We will consider two other ways to count such components, one which has already arisen for the modified process, see (8.63). Let

$$\mathscr{W}(\mathscr{C}) := \sum_{\text{blob}\in\mathscr{C}} f_{\text{blob}}(t_n), \qquad \mathscr{B}(\mathscr{C}) = \#\text{ of blobs in }\mathscr{C}. \tag{8.73}$$

Theorem 8.21. *Fix K and consider the K maximal components $\mathscr{C}_i(\lambda) \subseteq \mathrm{CM}_n(t_c + \lambda/n^{1/3})$. Then for each fixed $1 \leqslant i \leqslant K$, we have*

(a) *With high probability all the surplus edges in $\mathscr{C}_i(\lambda)$ are created in the interval $[t_n, t_c + \lambda/n^{1/3}]$.*

(b) *The free edge weight counts satisfy $\mathscr{W}(\mathscr{C}_i(\lambda))/|\mathscr{C}_i(\lambda)| \xrightarrow{\mathrm{P}} \nu - 1$. In particular, these counts have the same distributional limits as the modified process (Proposition* 8.16*) namely*

$$\left(\frac{\beta^{1/3}}{\mu(\nu-1)}\frac{\mathscr{W}(\mathscr{C}_i(\lambda))}{n^{2/3}} : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \boldsymbol{\xi}\left(\frac{2\mu\nu(\nu-1)}{\beta^{2/3}}\lambda\right),$$

Proof: The construction and proof are identical to [54] so we will only give the main ideas of the proof. To simplify notation, for the rest of the proof we will assume $\lambda = 0$, the same proof works for general λ. Let $N_i^{(n)}(t_n, t_c)$ denote the number of surplus edges created in the interval $[t_n, t_c]$. Obviously

$$N_i^{(n)}(t_n, t_c) \leqslant N_i^{(n)}(0)$$

Then to prove (a) it is enough to show that the random variable on the left has the same distributional limit as $N_i^{(n)}(0)$ (Theorem 8.20) namely

$$N_i^{(n)}(t_n, t_c) \xrightarrow{\mathrm{w}} \mathrm{Pois}_i(\alpha(0)). \tag{8.74}$$

First note that at time t_n, the number of *alive* free half-edges is

$$n\bar{s}_1(t_n) \approx n\mu\frac{\nu-1}{\nu} + \frac{n\mu}{n^{\delta}},$$

while at time t_c the number of alive free half-edges is approximately $n\mu(\nu-1)/\nu$. Thus $n\mu/n^{\delta}$ edges are used up in the interval $[t_n, t_c]$. Conditional on $\mathrm{CM}_n(t_n)$ consider the process $\mathrm{CM}_n(t_n, p)$ constructed in the following two steps

(i) Let

$$p = \frac{\nu}{\nu-1}\frac{1}{n^{\delta}}. \tag{8.75}$$

Retain each alive edge in $\mathrm{CM}_n(t_n)$ with probability p and discard otherwise. Let a_i denote the number of retained alive edges in blob i and note that $a_i = \mathrm{Bin}(f_i, p)$.

(ii) Now create a uniform matching amongst all retained half-edges.

Then using [34] (see Lemma 8.3 and Proposition 8.4) it can be shown that the process at time t_c namely $\mathrm{CM}_n(t_c)$ is asymptotically equivalent (for all the functionals of interest in Theorem 8.21) to $\mathrm{CM}_n(t_n, p)$. More precisely there exists a positive sequence $\varepsilon_n = o(n^{-\delta})$ such that we can couple $\mathrm{CM}_n(t_c)$ with two constructions

$$\mathrm{CM}_n(t_n, p - \varepsilon_n) \subseteq \mathrm{CM}_n(t_c) \subseteq \mathrm{CM}_n(t_n, p + \varepsilon_n),$$

such that for $\mathrm{CM}_n(t_n, p - \varepsilon_n)$ and $\mathrm{CM}_n(t_n, p + \varepsilon_n)$, the assertions of Theorem 8.21(a) and (8.74) (see, for example, Section 8.7 for a variant of this argument). Using the coupling above it is thus enough to work with this "percolation" variant.

The rest of this section describes how to analyze $\mathrm{CM}_n(t_n, p)$. To reduce additional notation we continue to refer to this process as $\mathrm{CM}_n(t_c)$. We will be leaning heavily on the behavior of the susceptibility functions in Theorem 4.10 as well as the analysis of the tail behavior of component sizes at time t_n that was used to derive bounds on the maximal component size in Section 8.2. Let a_i denote the number of free edges retained in blob i out of f_i possible free edges after the percolation step in the above construction. Note that conditional on the blob (free weight) sizes $\{f_i : i \in [m]\}$ $a_i = \mathrm{Bin}(f_i, p)$. For the rest of the proof we will use

$$m = \text{ \# of blobs at time } t_n.$$

The same proof technique as in Section 8.2, standard concentration inequalities [23] and Theorem 4.10 implies that when $\delta < 1/4$, the sequence $\mathbf{a} = \{a_i : i \in [m]\}$ satisfies $a_{\max} = O_P(n^\delta \log^3 n)$, $\sum_i a_i \sim p \sum_i f_i$ and $\sum_i a_i(a_i - 1) \sim p^2 \sum_i f_i(f_i - 1)$. For example to see why the first assertion is true, note that the self-bounding concentration inequality [23, Theorem 6.12] implies that for a $X = \mathrm{Binomial}(n, p)$,

$$\mathbb{P}(X \geqslant \mathbb{E}(X) + t) \leqslant \exp\left(-\frac{t^2}{2\mathbb{E}(X) + 2t/3}\right) \tag{8.76}$$

Using Theorem 4.11 resulting in the very crude bound $f_i(t_n) \leqslant \beta n^{2\delta} \log^2 n$ whp for an appropriate constant β and all $i \in [m]$ then proves the assertion about $a_{\max}$. Similar arguments give the second two assertions. Using (4.12) gives

$$\sum_i a_i \sim \mu n^{1-\delta}, \qquad \sum_i a_i(a_i - 1) \sim \mu n^{1-\delta}. \tag{8.77}$$

Using (4.13) gives

$$\sum_i a_i^3 \sim \frac{\beta}{\nu^3} n. \tag{8.78}$$

Note that the third moment is an order of magnitude larger than the first two moments. This plays a major role in the non-standard scaling of the exploration walk below. Similarly using (4.14) gives

$$\sum_i a_i f_i \sim \frac{\mu(\nu - 1)}{\nu} n, \qquad \sum_i a_i |\mathscr{C}_i(t_n)| \sim \frac{\mu}{\nu} n. \tag{8.79}$$

Now conditional on the percolated degree sequence $\mathbf{a}$ let us describe Riordan's exploration construction of $\mathrm{CM}_n(t_n, p)$. This is closely related to the walk used in Section 8.2 to analyze the maximal component size and diameter. By stage i in the exploration, i-blobs have been 'reached' and $m - i$ vertices unreached, let $\mathscr{A}_i^v$ and $\mathscr{U}_i^v$ denote the respective set of reached and unreached blobs (here the superscript v is used to denote that these correspond to the ones for blobs, we will need similar objects for edges). A certain (random) number of stubs have been paired to form full edges and each unpaired half-edge is either active (belongs to a reached blob) or belongs to an unreached blob and is designated unreached. Write $\mathscr{A}_i^e$ and $\mathscr{U}_i^e$ for the respective sets and let $A(i) = |\mathscr{A}_i^e|$ and $U(i) = |\mathscr{U}_i^e|$ denote the number of these half-edges.

(a) At time $i = 0$, let $A(0) = 0$. Initialize by selecting a blob $i \in [m]$ with probability proportional to the percolated free half-edge degree $\mathbf{a}$.

(b) Having constructed the process till i, at time $i+1$ if $A(i)>0$ pick an active half-edge s^a_{i+1} in some predetermined order. Reveal the partner of this half-edge s^u_{i+1} which will necessarily be in $\mathscr{U}^e_i$ (see (c) below for why this is true). Let $v(i+1)$ denote the blob corresponding to s^u_{i+1}. Writing $\mathscr{F}_i$ for the natural sigma-field generated by the process till time i, by construction, conditional on $\mathscr{F}_i$, $v(i+1)$ is selected with probability proportional to $\{a_u : u\in\mathscr{U}^v_i\}$. Let $\eta_{i+1}=a_{v(i+1)}$ denote the (percolated) degree of the blob selected at time i. This blob now has $\eta_{i+1}-1$ half-edges since one of the η_{i+1} half-edges was used to form the full edge at time $i+1$. Declare the rest of the half-edges active.

(c) Before moving on, we inspect each of the remaining half-edges of $v(i+1)$ and see if any of these are either paired with some other active half-edge in $\mathscr{A}^e_i$ or if they are paired with each other. Any such full edge creates a surplus edge in the presently explored component and in [54] is referred to as a "back edge". Let θ_{i+1} denote the number of back-edges found during step $i+1$.

(d) If $A(i)=0$ this implies we have finished exploring a component and we selected the next blob to start exploring the component of with probability proportional to $\{a_u : u\in\mathscr{U}^v_i\}$.

Let $\mathrm{Comp}(i)$ denote the number of components that we have started exploring within the first i steps and

$$Z_n(i)=A(i)-2\,\mathrm{Comp}(i),\qquad Y_n(i)=\sum_{j\leqslant i}\theta_j. \tag{8.80}$$

By [54, Equation 2.2]

$$Z_n(i+1)-Z_n(i)=\eta_{i+1}-2-2\theta_{i+1}. \tag{8.81}$$

Further for fixed $k\geqslant 1$, writing $t_k=\min\{i: Z_n(i)=2k\}$, the size of the k-th component (not necessarily k-th largest) explored by the walk then $|\mathscr{C}_k|=t_k-t_{k-1}$. Now define the process

$$W_\star(s)=\sqrt{\frac{\beta}{\nu^3\mu}}B(s)-\frac{\beta}{\mu^2\nu^3}\frac{s^2}{2},\qquad s\geqslant 0, \tag{8.82}$$

where as before $B(\cdot)$ is standard Brownian motion. Let $\tilde W_\star$ denote the reflection of the above process at zero. Following Aldous in [5] adjoin the process $\tilde W_\star$ with a point process of marks $\Xi(\cdot)$ informally described as

$$\mathbb{P}(\Xi(ds)=1\,|\,\{W_\star(u),u\leqslant s\})=\tilde W_\star(s)ds,$$

and precisely described as the counting process such that

$$\left(\Xi[0,s]-\int_0^s \tilde W_\star(u)du\right)_{s\geqslant 0}\ \text{is a martingale.} \tag{8.83}$$

Proposition 8.22. *Consider the processes*

$$\bar Z_n(s)=\frac{1}{n^{1/3}}Z_n\left(sn^{2/3-\delta}\right),\qquad \bar Y(s)=Y(sn^{2/3-\delta}),\qquad s\geqslant 0.$$

Then $(\bar Z_n,\bar Y)\xrightarrow{\mathrm{w}}(W_\star,\Xi)$ *as* $n\to\infty$.

Remark 10. Note the non-standard time and space scaling in the above convergence, a direct consequence of (8.77) and (8.78). At this point we would like to remind the reader that since in the above exploration, the "vertices" actually correspond to blobs, the above result implies results about the *number* of blobs $\mathscr{B}(\mathscr{C})$ for connected components $\mathscr{C} \subseteq \mathrm{CM}_n(t_c)$.

Before diving into the proof, let us read a consequence of the above Proposition. Let $W = W_0$ denote the Brownian motion with parabolic drift as in (2.6) with $\lambda = 0$ and let $\tilde{W}$ denote the corresponding reflected process at zero. Brownian scaling implies the distributional equivalence

$$(W_\star(s) : s \geqslant 0) \stackrel{d}{=} \left(\frac{\beta^{1/3}}{\nu} W\left(\frac{\beta^{1/3}}{\nu\mu} s\right) : s \geqslant 0\right) \tag{8.84}$$

where $W_\star(\cdot)$ is the limit process in Proposition 8.22. Also recall that $\boldsymbol{\xi}(0) = (\xi_1(0), \xi_2(0), \ldots)$ denoted the sizes of excursions from zero of $\tilde{W}$.

Lemma 8.23. *Fix any $k \geqslant 1$. Then there exist components $\mathscr{I}_1, \mathscr{I}_2, \ldots, \mathscr{I}_k \subseteq \mathrm{CM}_n(t_c)$ such that the number of blobs $\mathscr{B}(\mathscr{I}_j)$ and number of surplus edges $N_{\star,j}^{(n)}$ in these components satisfy*

$$\left(\left[\frac{\beta^{1/3}}{\nu\mu}\frac{\mathscr{B}(\mathscr{I}_j)}{n^{2/3-\delta}}, N_{\star,j}^{(n)}\right] : 1 \leqslant j \leqslant k\right) \xrightarrow{\mathrm{w}} ([\xi_i(\alpha(0)), \mathrm{Pois}_i(\alpha(0))] : 1 \leqslant i \leqslant k).$$

Note that at this stage, we are not asserting that $\mathscr{I}_i = \mathscr{C}_i$ for $1 \leqslant i \leqslant k$, however if we were able to prove this assertion, this would immediately prove (8.74) and thus (a) of Theorem 8.21. This assertion follows later from the proof.

Proof of Proposition 8.22: The proof follows via a standard application of the Martingale functional central limit Theorem (see e.g. [32, Chapter 7]) to the walk Z_n as in [5, 54]. We sketch the proof. We start with the walk. We will later show that for any fixed $T > 0$, the number of surplus edges created by time $Tn^{2/3-\delta}$, $Y(Tn^{2/3-\delta}) = O_P(1)$. Thus as as in [54], defining the walk

$$S_n(i+1) = S_n(i) + (\eta_{i+1} - 2), \qquad i \geqslant 0,$$

we have

$$\sup_{0 \leqslant s \leqslant Tn^{2/3-\delta}} |S_n(i) - Z_n(i)| = O_P(1). \tag{8.85}$$

Thus it is enough to prove Proposition 8.22 for S_n as opposed to Z_n. The following is obvious from the construction of the walk.

Lemma 8.24. *The order of the blobs reached by the exploration process namely $(v(1), v(2), \ldots, v(m))$ is in the size-biased random order using the weight sequence $\mathbf{a}$ of the percolated free-edge weights.*

Now let us calculate the infinitesimal mean of the process S_n. We have

$$\mathbb{E}(\eta_{i+1} - 1 | \mathscr{F}_i) = \sum_{u \notin \mathscr{U}_i^v} \frac{a_u}{\sum_{u' \notin \mathscr{U}_i^v} a_{u'}} (a_u - 1) = \frac{\sum_{j=1}^m a_j(a_j - 1) - \sum_{j=1}^i a_{v(j)}(a_{v(j)} - 1)}{\sum_{j=1}^m a_j - \sum_{j=1}^i a_{v(j)}} \tag{8.86}$$

Standard calculations with size-biased random re-ordering (Lemma 6.8) imply that uniformly in $i \leqslant Tn^{2/3-\delta}$,

$$\sum_{j=1}^{i} a_{v(j)} \sim i\frac{\sum_j^m a_j^2}{\sum_j^m a_j}, \qquad \sum_{j=1}^{i} a_{v(j)}(a_{v(j)}-1) \sim i\frac{\sum_j^m a_j^2(a_j-1)}{\sum_j^m a_j}.$$

Using these in (8.86) along with (8.77) and (8.78) gives

$$\mathbb{E}(\eta_{i+1}-1|\mathscr{F}_i) = 1 - \frac{\beta}{\mu^2\nu^3}\frac{i}{n^{1-2\delta}} + O\left(\frac{1}{n^{1-\delta}}\right) \tag{8.87}$$

Thus the drift accumulated by time $sn^{2/3-\delta}$ for the process $S_n(\cdot)$ is

$$-\frac{s^2}{2}\frac{\beta}{\mu^2\nu^3}\frac{n^{4/3-2\delta}}{n^{1-2\delta}} + o(n^{1/3}) = -\frac{s^2}{2}\frac{\beta}{\mu^2\nu^3}n^{1/3} + o(n^{1/3}).$$

This explains the parabolic drift in (8.82). Similarly calculating the infinitesimal variance gives

$$\mathrm{Var}(\eta_{v(i+1)}|\mathscr{F}_i) \approx \frac{\sum_{i=1}^m a_i^3}{\sum_i a_i} = \frac{\beta}{\nu^3\mu}n^\delta + o(n^\delta). \tag{8.88}$$

Thus the accumulated variance by time $sn^{2/3-\delta}$ is

$$\frac{\beta}{\nu^3\mu}sn^{2/3} + o(n^{2/3}).$$

This explains the scaling of the Brownian motion in (8.82).

Finally let us study the surplus edge process. Recall that $A(i), U(i)$ denote the total number of active and unreached edges respectively and note that as in [54, Equation 2.9], for the surplus edge process $Y(\cdot)$ we have uniformly for $i \leqslant Tn^{2/3-\delta}$,

$$\mathbb{E}(\theta(i+1)|\mathscr{F}_i) = \mathbb{E}(\eta_{i+1}-1|\mathscr{F}_i)\frac{A(i)}{U(i)} = (1+o(1))\frac{A(i)}{\sum_{j=1}^m a_j} \tag{8.89}$$

$$= (1+o(1))\frac{A(i)}{\mu n^{1-\delta}}. \tag{8.90}$$

By construction $A(i) = Z_n(i) - \min_{0\leqslant j\leqslant i} Z_n(j)$, coupled with the above result for S_n implies that $Y_n(Tn^{2/3-\delta}) = O_P(1)$. Using this with (8.85) and (8.89) completes the analysis of the asymptotics for the surplus edge process and thus the Proposition. ∎

Proof of Theorem 8.21: We start by showing that for any fixed $k \geqslant 1$,

$$\mathbb{P}(\mathscr{I}_i = \mathscr{C}_i(0)\ \forall\ 1 \leqslant i \leqslant k) \to 1, \qquad \text{as } n \to \infty. \tag{8.91}$$

Using the asymptotics for the number of surplus edges in Proposition 8.22 and (8.74) would then prove (a) of the Theorem. For fixed $i \geqslant 1$ the size (number of vertices) in $\mathscr{I}_i$ is given by $|\mathscr{I}_i| = \sum_{\text{blob}\in\mathscr{I}_i} |\mathscr{C}_{\text{blob}}(t_n)|$. Thus it is enough to show

$$\frac{\nu|\mathscr{I}_i|}{\mathscr{B}(\mathscr{I}_i)n^\delta} \xrightarrow{\mathrm{P}} 1. \tag{8.92}$$

Then using Lemma 8.24 and comparing this with Theorem 8.20 shows that the sizes $|\mathscr{I}_i|$ scaled by $n^{-2/3}$ have the same distributional asymptotics as $\mathscr{C}_i(0)$ which proves (8.91) by the maximality of the components $\mathscr{C}_i$. Now recall that in the construction above, we explore vertices in the size-biased random re-ordering using the weight sequence **a**. Thus (8.92) follows directly from Lemma 6.8 (taking the weight sequence $x_i = a_i$ and $u_i = |\mathscr{C}_i(t_n)|$ in the statement of the Lemma) and using (8.79). Thus henceforth we will refer to $\mathscr{I}_i = \mathscr{C}_i(0)$. Following the same steps but now using the weight sequence f_i and again using (8.79) and Lemma 6.8 then shows that

$$\frac{\mathscr{W}(\mathscr{C}_i(0))}{\frac{\nu-1}{\nu} n^{\delta} \mathscr{B}(\mathscr{C}_i(0))} \xrightarrow{\mathrm{P}} 1. \tag{8.93}$$

This coupled with Proposition 8.22 for $\mathscr{B}(\mathscr{C}_i(0))$ proves (b) of the Theorem. This completes the proof.

■

An indirect consequence of (8.91) is the following lemma.

Lemma 8.25. *Consider the construction of the walk Z_n as in* (8.80)*. Fix any $k \geq 1$ and $\varepsilon > 0$. For fixed T write $\mathscr{E}_n(T,k)$ for the event that the walk has finished exploring all k maximal components $\{\mathscr{C}_i(0) : 1 \leq i \leq k\}$ by time $T n^{2/3-\delta}$. Then there exists a constant $T = T(k,\varepsilon) < \infty$ such that*

$$\limsup_{n\to\infty} \mathbb{P}([\mathscr{E}_n(T,k)]^c) \leq \varepsilon$$

A similar result is true in the last part of this section which once again deals with the modified process $\mathscr{G}_n^{\mathrm{modi}}(t_c + \lambda/n^{1/3})$. Similar walk constructions as to the one used in this section are used in the proof and explain why the statement and proof of this result has been deferred to this section. Recall that Theorem 8.18 described scaling limits of maximal components in $\mathscr{G}_n^{\mathrm{modi}}(t_c + \lambda/n^{1/3})$ but where the measure used was μ_{free}. We end this section by showing that this can be replaced by the counting measure.

Theorem 8.26. *The free weight maximal components $(\mathfrak{C}_i^{\mathrm{modi}}(\lambda) : i \geq 1)$ of $\mathscr{G}_n^{\mathrm{modi}}(t_c+\lambda/n^{1/3})$ viewed as connected metric spaces with vertex set $[n]$ where we incorporate both the Blob-level superstructure and inter-blob distances as in Section* 3.2 *but now equipped with the counting measure μ_{ct} where each vertex is assigned mass one has the same scaling limits as those asserted for the maximal components of $\mathrm{CM}_n(\lambda)$ in Theorem* 4.9 *namely*

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu}\frac{1}{n^{1/3}}, \frac{\beta^{1/3}}{\mu n^{2/3}}\mu_{\mathrm{ct}}\right)\mathfrak{C}_i^{\mathrm{modi}}(\lambda) : i \geq 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty\left(\frac{2\nu(\nu-1)\mu}{\beta^{2/3}}\lambda\right), \qquad \text{as } n \to \infty.$$

Remark 11. Comparing the scaling in μ_{free} and μ_{ct} in Theorem 8.18 and 8.26, the above result says that in a certain uniform sense, $\mu_{\mathrm{free}} \approx (\nu-1)\mu_{\mathrm{ct}}$ when restricted to the maximal components. The proof below makes this notion rigorous.

Proof: We work with the maximal component $\mathfrak{C}_1^{\mathrm{modi}}$ with $\lambda = 0$. The same proof works for general $k \geq 1$ and $\lambda \in \mathbb{R}$. First note that Theorem 8.18 already gives scaling limits for the metric structure. Comparing μ_{free} and μ_{ct} and the assertion of Theorem 8.18 where the mass of a vertex is the number of still free edges at time t_n, to prove Theorem 8.26 it is enough to show that

$$\frac{\sum_{\mathrm{blob}\ni\mathfrak{C}_1^{\mathrm{modi}}(0)} \left| |\mathscr{C}_{\mathrm{blob}}(t_n)| - \frac{f_{\mathrm{blob}}(t_n)}{(\nu-1)} \right|}{n^{2/3}} \xrightarrow{\mathrm{P}} 0, \qquad \text{as } n \to \infty. \tag{8.94}$$

Now recall the construction of CM_n and in particular the walk $Z_n(\cdot)$ in (8.80) that enabled us to prove Lemma 8.23 namely that the **number** of blobs in the maximal component

$$\mathscr{B}(\mathscr{C}_1(0)) \sim \frac{\nu\mu}{\beta^{1/3}} n^{2/3-\delta}\xi_1(\alpha(0)).$$

Switching perspective to the random graph $\mathscr{G}_n^{\mathrm{modi}}(t_c)$, at the blob-level this object belongs to the random graph family $\mathscr{G}(\mathbf{x},q)$ as defined in Section 2.4 with the special choices $\mathbf{x}$ and q as in (8.61) (with $\lambda=0$). Recall Aldous's construction of $\mathscr{G}(\mathbf{x},q)$ in Section 6.2.1 through the size biased construction using the weight sequence $\mathbf{x}$ resulting in the sequence of blobs explored as

$$(v(1), v(2), \ldots, v(m)).$$

Note that since $\mathbf{x}$ is a constant multiple of the number of free edges in blobs $\mathbf{f} = \{f_i(t_n) : i \in [m]\}$, the size-biased order is equivalent to size-biasing with respect to weight sequence $\mathbf{f}$. Recall that this construction initialized with $v(1)$ selected using the weight measure $\mathbf{f}$ and for each $i \geqslant 1$, we find a number of "children" $c(i)$ of $v(i)$ (unexplored blobs connected to $v(i)$). Similar to the walk construction (8.80), define the walk

$$Z_n^{\mathrm{modi}}(i) = \sum_{j=1}^{i}(c(j)-1), \qquad i \geqslant 1.$$

Using arguments identical to CM_n we now get the following analog of Lemmas 8.23 and Lemma 8.25. We omit the proof.

Lemma 8.27. *Fix $k \geqslant 1$. The modified process $\mathscr{G}_n^{\mathrm{modi}}$ satisfies the following asymptotics.*

(a) *The number of k maximal free weight components satisfy*

$$\left(\frac{\beta^{1/3}}{\nu\mu}\frac{\mathscr{B}(\mathfrak{C}_i^{\mathrm{modi}}(0))}{n^{2/3-\delta}} : 1 \leqslant i \leqslant k\right) \xrightarrow{\mathrm{w}} (\xi_i(\alpha(0)) : 1 \leqslant i \leqslant k).$$

(b) *Fix $\varepsilon > 0$ and for fixed T write $\mathscr{E}_n^{\mathrm{modi}}(T,k)$ for the event that the walk $Z_n^{\mathrm{modi}}(\cdot)$ has finished exploring all k maximal components $\{\mathfrak{C}_i^{\mathrm{modi}}(0) : 1 \leqslant i \leqslant k\}$ by time $Tn^{2/3-\delta}$. Then there exists a constant $T = T(k,\varepsilon) < \infty$ such that*

$$\limsup_{n\to\infty} \mathbb{P}([\mathscr{E}_n^{\mathrm{modi}}(T,k)]^c) \leqslant \varepsilon$$

Now using (b) of the above Lemma, we see that for fixed $\varepsilon > 0$, there exists constant $T < \infty$ such that with probability $\geqslant 1-\varepsilon$ as $n \to \infty$

$$\frac{\sum_{\mathrm{blob}\in\mathfrak{C}_1^{\mathrm{modi}}(0)}\left||\mathscr{C}_{\mathrm{blob}}(t_n)| - \frac{f_{\mathrm{blob}}(t_n)}{(\nu-1)}\right|}{n^{2/3}} \leqslant \frac{\sum_{i=1}^{Tn^{2/3-\delta}}\left||\mathscr{C}_{v(i)}(t_n)| - \frac{f_{v(i)}(t_n)}{(\nu-1)}\right|}{n^{2/3}},$$

where as before $(v(i) : i \in [m])$ is the size-biased re-ordering of the blobs using the weight sequence $\mathbf{f}$. The following lemma completes the proof of (8.94) and thus the proof of Theorem 8.26.

Lemma 8.28. *For any fixed $T < \infty$,*

$$\frac{\sum_{i=1}^{Tn^{2/3-\delta}}\left||\mathscr{C}_{v(i)}(t_n)| - \frac{f_{v(i)}(t_n)}{(\nu-1)}\right|}{n^{2/3}} \xrightarrow{\mathrm{P}} 0, \qquad \text{as } n \to \infty.$$

Proof: Using standard properties of size-biased reordering (Lemma 6.8) implies that for fixed T we can relate the above sum through the size-biased re-ordering to the the size-biased average as

$$\sum_{i=1}^{Tn^{2/3-\delta}} \left| |\mathscr{C}_{v(i)}(t_n)| - \frac{f_{v(i)}(t_n)}{(\nu-1)} \right| \sim Tn^{2/3-\delta} \frac{\sum_{j\in[m]} f_j(t_n) \left| |\mathscr{C}_j(t_n)| - \frac{f_j(t_n)}{(\nu-1)} \right|}{\sum_{j\in[m]} f_j(t_n)}. \tag{8.95}$$

Note that the denominator on the left hand side is just the total number of free edges at time t_n is $\geqslant ns_1(t_c)$ whp. Dividing both sides of (8.95) by $n^{2/3}$, using Cauchy-Schwartz and simplifying implies that

$$\frac{\sum_{i=1}^{Tn^{2/3-\delta}} \left| |\mathscr{C}_{v(i)}(t_n)| - \frac{f_{v(i)}(t_n)}{(\nu-1)} \right|}{n^{2/3}} = \Theta_P\left(\frac{\sqrt{\sum_i f_i^2(t_n)}\sqrt{\sum_i (\mathscr{C}_i(t_n) - f_i/(\nu-1))^2}}{n^{1+\delta}} \right). \tag{8.96}$$

Expanding the bound on the right, note that it can be expressed in terms of the susceptibility functions that were analyzed in great detail in Section 8.3 as

$$\frac{\sqrt{n\bar{s}_2(t_n)}\sqrt{n\left(\bar{s}_2^\star(t_n) + \frac{\bar{s}_2(t_n)}{(\nu-1)^2} - 2\frac{\bar{g}(t_n)}{(\nu-1)}\right)}}{n^{1+\delta}}$$

Using Proposition 8.12 for the asymptotics of $\bar{s}_2$ and $\bar{g}$ and Lemma 8.15 for $\bar{s}_2^\star$ completes the proof. ■

8.6. **Coupling the modified process and** CM_n**.** Let us briefly summarize the developments of the last few sections. Starting at time t_n with $\mathrm{CM}_n(t_n)$ we now have two processes, the original process CM_n and the modified process $\mathscr{G}_n^{\mathrm{modi}}$. We know a wide array of properties of $\mathscr{G}_n^{\mathrm{modi}}$ including the scaling limit of the metric structure (Theorems 8.18 and 8.26) whilst for the original process CM_n we know a few macroscopic properties including component sizes (Theorem 8.20) and weight of maximal edges counted via free half-edge weight (Theorem 8.21). The aim of this section is to couple the two processes and see how results for $\mathscr{G}_n^{\mathrm{modi}}$ imply the same for CM_n. To simplify notation we assume $\lambda = 0$, the same proof works for general λ.

First note that both CM_n and $\mathscr{G}_n^{\mathrm{modi}}$ are continuous time graph-valued Markov processes. Decomposing each into the embedded discrete chain (represented respectively as CM_n^* and $\mathscr{G}_n^{*,\mathrm{modi}}$) and the time of jumps results in the following simple descriptions:

For the process $\mathrm{CM}_n(t)$ for $t > t_n$:

(a) Starting from $\mathrm{CM}_n^*(0) = \mathrm{CM}_n(t_n)$, the process $\{\mathrm{CM}_n^*(k) : k \geqslant 1\}$ is obtained by sequentially selecting pairs of half-edges uniformly at random **without replacement** and forming full edges.

(b) The rate of formation of edges is given by $n\bar{s}_1(t)$. Recall that Lemma 8.2 derives asymptotics for $\bar{s}_1$. In particular for $t > 0$, letting $\mathfrak{R}_n[t_n, t]$ denote number of *full* edges formed in the interval $[t_n, t]$, Lemma 8.2 implies for any $\gamma < 1/2$, whp

$$\left| \mathfrak{R}_n[t_n, t_c] - \frac{n^{1-\delta}\mu}{2} \right| \leqslant n^{1-\gamma}. \tag{8.97}$$

For the process $\mathscr{G}_n^{\mathrm{modi}}$ for $t > t_n$:

(a′) Starting from $\mathscr{G}_n^{*,\mathrm{modi}}(0) = \mathrm{CM}_n(t_n)$, the process $\left\{\mathscr{G}_n^{*,\mathrm{modi}}(k) : k \geqslant 0\right\}$ is obtained by sequentially selecting half-edges at random **with replacement** and forming full edges.

(b′) Using (8.59), the rate at which a half-edge rings to form a full edge is constant and is given by

$$\alpha_n := \frac{n^2 \bar{s}_1^2(t_n)}{n s_1(t_c)} = n s_1(t_c) \frac{\bar{s}_1^2(t_n)}{s_1^2(t_c)}. \tag{8.98}$$

There is an obvious coupling between (a) and (a′) on a common probability space such that

$$\mathrm{CM}_n^*(k) \subseteq \mathscr{G}_n^{*,\mathrm{modi}}(k), \qquad \text{for all } k \geqslant 0. \tag{8.99}$$

Here every time a half-edge is selected in CM_n that was sampled before, this is not used in the original process CM_n and the corresponding edge formed is called a "bad" edge. Else the full edge formed registers both in CM_n and $\mathscr{G}_n^{\mathrm{modi}}$ and is recorded as a "good" edge. The process CM_n has all the good edges while $\mathscr{G}_n^{\mathrm{modi}}$ has all the good edges and a number of bad edges.

We start by understanding asymptotics for α_n in (8.98). Since we work at the entrance boundary, we need to be rather precise with our estimates. The following follows from Lemma 8.2.

Lemma 8.29. *With high probability as $n \to \infty$*

$$1 - \frac{4\nu}{\nu-1}\frac{1}{n^\delta} \quad \leqslant \frac{\bar{s}_1^2(t_n)}{s_1^2(t_c)} \quad \leqslant 1$$

Now note that conditional on α_n, the rate at which half-edges ring in $\mathscr{G}_n^{\mathrm{modi}}$ is just a Poisson process with rate α_n and in particular for any $t > t_n$, the number of rings (alternatively full edges formed) in $\mathscr{G}_n^{\mathrm{modi}}$ by time t,

$$\mathfrak{R}_n^{\mathrm{modi}}[t_n, t] \stackrel{d}{=} \text{ Poisson } (\alpha_n(t - t_c)) \tag{8.100}$$

By Lemma 8.29 and (8.98) whp

$$n s_1(t_c)\left(1 - \frac{4\nu}{\nu-1}\frac{1}{n^\delta}\right) \leqslant \quad \alpha_n \quad \leqslant n s_1(t_c) \tag{8.101}$$

Now define

$$\varepsilon_n := \frac{\nu}{2(\nu-1)} \cdot \frac{A}{n^{2\delta}}, \qquad A > 4\left(\frac{\nu}{\nu-1}\right)^2, \tag{8.102}$$

where A above is an appropriately chosen constant independent of n. Below we will have one more constraint on A. Using the distribution of $\mathfrak{R}_n^{\mathrm{modi}}$ from (8.100), the bounds on α_n from (8.101) and standard tail estimates for the Poisson distribution, we get whp

$$\frac{n^{1-\delta}\mu}{2} + \frac{n^{1-2\delta}\mu}{8}\left(A - 4\left(\frac{\nu}{\nu-1}\right)^2\right) \quad \leqslant \quad \mathfrak{R}_n^{\mathrm{modi}}[t_n, t_c + \varepsilon_n] \quad \leqslant \quad \frac{n^{1-\delta}\mu}{2} + n^{1-2\delta}\mu A \tag{8.103}$$

This explains the bound on A in (8.102). Also note that since by assumption $\delta > 1/6$, thus $\varepsilon_n = O(n^{-2\delta}) = o(n^{-1/3})$. Thus by the results in Section 8.4 the following Proposition easily follows.

Proposition 8.30. *The asymptotics in Propositions* 8.16*,* 8.17 *and Theorems* 8.18 *and* 8.26 *hold with* $\lambda = 0$ *for* $\mathscr{G}_n^{\mathrm{modi}}(\varepsilon_n)$.

The final ingredient is bounding

$$\mathfrak{B}_n[t_n, t] := \#\text{ of bad edges in } [t_n, t].$$

As we sequentially construct the coupled discrete time processes $(\mathrm{CM}_n^*(k), \mathscr{G}_n^*(k) : k \geqslant 0)$ both started at $\mathrm{CM}_n(t_n)$ at time $k = 0$, let $\mathscr{F}_k$ denote the natural σ-field at time k. Then for any $k \geqslant 1$, conditional on $\mathscr{F}_k$, the number of bad edges created at time $k+1$ is stochastically dominated by $Y_i = \mathrm{Bin}(2, 2i/n\bar{s}_1(t_n))$. In particular, using the right side of (8.103) and $n\bar{s}_1(t_n) \geqslant n\mu(\nu-1)/\nu$, we get whp

$$\mathfrak{B}_n[t_n, t_c + \varepsilon_n] \leqslant \frac{4\nu}{\nu - 1} \leqslant 4\frac{\nu\mu}{\nu-1} n^{1-2\delta}. \tag{8.104}$$

Now using (8.97), the left side of (8.103) and assuming

$$A > 4\left(\frac{\nu}{\nu-1}\right)^2 + 4\frac{\nu\mu}{\nu-1}, \tag{8.105}$$

we get the following important result.

Proposition 8.31. *With high probability as $n \to \infty$ we have* $\mathrm{CM}_n(t_c) \subseteq \mathscr{G}_n^{\mathrm{modi}}(\varepsilon_n)$.

Now recall that for fixed $i \geqslant 1$, we used $\mathfrak{C}_i^{\mathrm{modi}}(\varepsilon_n)$ for the maximal component in $\mathscr{G}_n^{\mathrm{modi}}$ where the size of a component is counted according to the number of free edges at time t_n of the constituent blobs, namely using the weight function $\mathscr{W}$ as in (8.73). Comparing Proposition 8.16 for $\mathscr{G}_n^{\mathrm{modi}}$ with Theorem 8.21 for the original process $\mathrm{CM}_n(t_c)$ showing that $\mathscr{W}(\mathscr{C}_i(0))$ and $\mathscr{W}(\mathfrak{C}_i^{\mathrm{modi}}(\varepsilon_n))$ have the same distributional limits now yields the following result.

Corollary 8.32. *Fix* $k \geqslant 1$*. Then whp for all* $1 \leqslant i \leqslant k$ *we have* $\mathscr{C}_i(0) \subseteq \mathfrak{C}_i^{\mathrm{modi}}(\varepsilon_n)$.

By Theorem 8.18 we know the scaling limit of the metric structure of $\mathfrak{C}_i^{\mathrm{modi}}(\varepsilon_n)$ in the modified process. We want to show that $\mathscr{C}_i(0)$ has the exact same limit. Without loss of generality we just work with $i = 1$, the same argument works for general i.

By definition, every edge in $\mathscr{C}_1(0)$ is a good edge in the above coupling and further $\mathscr{C}_1(t_c)$ forms a connected subset of $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$. Thus between any two vertices $u, v \in \mathscr{C}_1(t_c)$, there exist paths completely composed of good edges and deletion of any "bad" edges connecting two vertices $u, v \in \mathscr{C}_1(0)$ cannot disconnect $\mathscr{C}_1(t_c)$. Thus any possible bad edge connecting two vertices in the connected subset $\mathscr{C}_1(0) \subseteq \mathfrak{C}_i^{\mathrm{modi}}(t_c + \varepsilon_n)$ is necessarily a surplus edge. Recall from Section 8.4 that $N_1^{(n)}(t_n, t_c)$ denoted the number of surplus edges born into $\mathscr{C}_i(0)$ in the interval $[t_n, t_c]$ while $N_1^{(n),\mathrm{modi}}(\varepsilon_n)$ denoted the corresponding number in $\mathscr{G}_n^{\mathrm{modi}}(t_c + \varepsilon_n)$. Obviously by the above coupling, whp $N_1^{(n)}(t_n, t_c) \leqslant N_1^{(n),\mathrm{modi}}(\varepsilon_n)$ since every surplus edge in $\mathscr{C}_1(0)$ will also be a surplus edge in $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$. However by Corollary 8.19 and Theorem 8.21(a) both random variables have the same distributional limits. This implies the following result.

Lemma 8.33. *With high probability as $n \to \infty$, every surplus edge in $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ is a good edge and is also a surplus edge in $\mathscr{C}_1(0)$.*

This then implies that for **every** pair of vertices in $\mathscr{C}_1(0)$, all paths between these two vertices in $\mathscr{C}_1(0)$ and $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ are the same whp. More precisely, the metric of $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ restricted to $\mathscr{C}_1(0)$ coincides exactly with the metric on $\mathscr{C}_1(0)$. Figure 8.1 gives a graphical description of the regime. Now equip both $\mathscr{C}_1(0)$ and $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ with the counting measure.

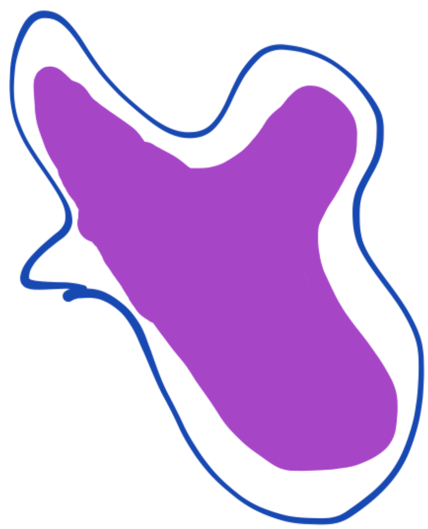

FIGURE 8.1. The purple shaded region represents $\mathscr{C}_1(0)$, a connected subset of the blue region which corresponds to $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$.

Consider the measured metric spaces

$$\bar{\mathscr{C}}_1(0) := \mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu n^{1/3}}, \frac{\beta^{1/3}}{\mu n^{2/3}}\right)\mathscr{C}_1(0), \qquad \bar{\mathfrak{C}}_1^{\mathrm{modi}}(\varepsilon_n) := \mathrm{scl}\left(\frac{\beta^{2/3}}{\mu\nu n^{1/3}}, \frac{\beta^{1/3}}{\mu n^{2/3}}\right)\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n),$$

namely (ignoring constants) we rescale distances by $n^{-1/3}$ and the counting measure by $n^{-2/3}$. The following proposition coupled with Theorem 8.26 that describes the scaling limits of $\bar{\mathfrak{C}}_1^{\mathrm{modi}}(\varepsilon_n)$ completes the proof of Theorem 4.9.

Proposition 8.34. *Under the Gromov-Hausdorff-Prokhorov distance d_{GHP} we have*

$$d_{\mathrm{GHP}}\left(\bar{\mathscr{C}}_1(0)\,,\, \bar{\mathfrak{C}}_1^{\mathrm{modi}}(\varepsilon_n)\right) \xrightarrow{\mathrm{P}} 0, \qquad \text{as } n \to \infty.$$

Proof: The intuitive idea behind the proof is simple. Ignoring surplus edges for the time being, note that by Theorem 8.26, $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ converges to a tilted version of the continuum random tree. Further the **connected** subgraph $\mathscr{C}_1(0) \subseteq \mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ using Theorem 8.20 and

Proposition 8.17 has asymptotically full measure which implies the Proposition via properties of the limit metric space. In pictures, in Figure 8.1 the white region once distances have been rescaled by $n^{-1/3}$ vanishes in the limit.

Let us now make this precise. For the rest of the proof just for notational convenience will ignore the constants $\beta^{2/3}/\mu\nu$ and $\beta^{1/3}/\mu$ that arise in the scaling of the distance and the counting measure respectively. We will first need to recall a few ideas of why Theorem 8.26 is true, which followed via using Theorem 3.4 to the special case of $\mathscr{G}_n^{\mathrm{modi}}(t_c+\varepsilon_n)$. Now this result relied on the connection between tilted versions of $\mathbf{p}$-trees and connected components in the graph $\mathscr{G}(\mathbf{x},q)$, see Section 6.2.3 and in particular Proposition 6.3 which followed from [15]. We elaborate now on some of the details in the proof of Proposition 6.3 as they will be needed here.

Recall from Section 6.4.1 that $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ can be viewed as being composed of three ingredients,

(i) the Blob-level superstructure describing the connectivity pattern when viewing each blob as a single vertex,
(ii) the inter-blob structure when we bring in the internal structure of the blobs,
(iii) and Blob to blob junction points which describe from which points within the blobs, edges are created to vertices in other blobs.

Let $m = m(n)$ denote the number of blobs in $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ and let $\mathbf{M} := \left\{f_{\mathrm{blob}} : \mathrm{blob} \in \mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)\right\}$ denote the corresponding blobs weights and let $\mathbf{M} = \left\{M_{\mathrm{blob}} : \mathrm{blob} \in \mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)\right\}$ be the corresponding blobs. For the time being let us only consider the blob-level superstructure of $\mathfrak{C}_1^{\mathrm{modi}}$, to be consistent with Section 6.4.1 and (8.61) we will write this graph as $\mathscr{G}_1^{\mathbf{p},\mathrm{modi}}$ where the probability measure $\mathbf{p} = (p_1, p_2, \dots, p_m)$ is given by

$$p_i = \frac{x_i}{\sum_{j\in\mathfrak{C}_1^{\mathrm{modi}}} x_j}, \qquad i \text{ is a blob in } \mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n), \tag{8.106}$$

Here by Theorem 3.2 and the scaling properties of the sequence $\mathbf{x}$ as defined in (8.61)

$$\frac{\mu(\nu-1)^2}{\nu^2 n^{1/3-\delta}}\mathscr{G}_1^{\mathbf{p},\mathrm{modi}} \xrightarrow{\mathrm{w}} \mathrm{Crit}_1(0). \tag{8.107}$$

Why this is true: Following [15], conditional on $\mathfrak{C}_1^{\mathrm{modi}}(\varepsilon_n)$ we will define an exploration of $\mathscr{G}_1^{\mathbf{p},\mathrm{modi}}$ called randomized Depth first search (rDFS) in [15, Section 7] which outputs a random planar tree using the superstructure $\mathscr{G}_1^{\mathbf{p},\mathrm{modi}}$. Initialize the process by selecting a blob $v(1)$ with probability $\mathbf{p}$. The exploration proceeds as follows. At each step $1 \leqslant i \leqslant m$ we track three types of blobs:

(a) The set of already explored blobs $\mathscr{O}(i)$.
(b) The set of active blobs $\mathscr{A}(i)$. We think of $\mathscr{A}(i)$ as a vertical stack of blobs.
(c) The set of unexplored blobs $\mathscr{U}(i) = [m] \setminus (\mathscr{A}(i) \cup \mathscr{O}(i))$.

Initialize the above with $\mathscr{A}(0) = \{v(1)\}$, $\mathscr{O}(0) = \emptyset$. At step $i \geqslant 1$ let $v(i)$ denote the blob on top of the stack $\mathscr{A}(i-1)$ and let $\mathscr{D}(i) = \left\{u(j) : 1 \leqslant j \leqslant d_{v(i)}\right\}$ denote the set of *yet to be explored* neighbors of $v(i)$ by the process. Here note that $d_{v(i)}$ does not represent the true degree of $v(i)$, rather just the number of blobs connected to $v(i)$ that have not yet been explored by the time the exploration process gets to $v(i)$. Now update the stack $\mathscr{A}(i)$ as follows:

(i) Delete $v(i)$ from $\mathscr{A}(i-1)$.

(ii) Generate a uniform random permutation $\boldsymbol{\pi}(i)$ on $[d(i)]$. Arrange the blobs in $\mathscr{D}(i)$ on top of the stack $\mathscr{A}(i-1)$ (after deleting $v(i)$) in the order prescribed by $\boldsymbol{\pi}(i)$.

Write $\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ for the random tree generated via this procedure and consider the collected of permitted edges

$$\mathscr{P}(\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}) := \left\{(v(i), j) : i \in [m], j \in \mathscr{A}(i-1) \setminus \{v(i)\}\right\}.$$

Then by [15] the edges in $\mathscr{G}_1^{\mathbf{p},\text{modi}} \setminus \mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ belong to $\mathscr{P}(\mathscr{T}_{\text{tilt,modi}}^{\mathbf{p}})$. Proposition 6.3 is proven in [15] by showing that conditional on the blobs in $\mathscr{G}^{\mathbf{p}}$, the distribution of $\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ has the tilted $\mathbf{p}$-tree distribution (6.8) and conditional on $\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ additional edges are added from the collection of permitted edges $\mathscr{P}(\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}})$ independently with prerequisite probabilities (Proposition 6.3(b)).

Now recall that the limit metric space $\text{Crit}_1(0)$ as in (8.107) is obtained via first sampling a tilted continuum random tree (conditional on the length $\gamma_1(0)$ obtained from Theorem 2.2). The reason for this as proved in [15] is that under technical assumptions (6.5),

$$\frac{\mu(\nu-1)^2}{\nu^2 n^{1/3-\delta}} \mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}} \xrightarrow{\text{w}} 2\tilde{\mathbf{e}}_{\gamma_1(0)} \tag{8.108}$$

Here we think of the right hand side as the random real tree encoded by $2\tilde{\mathbf{e}}_{\gamma_i(0)}$. We will denote this random compact metric space by CRT_{tilt}. Without loss of generality we will assume we work on a probability space where this convergence happens almost surely. Now note that $\mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ was a tree on the blob level picture $\mathscr{G}_1^{\mathbf{p},\text{modi}}$. As in Theorem 6.4 let $\bar{\mathscr{T}}_{\text{tilt}}^{\mathbf{p},\text{modi}}$ denote the corresponding metric space on $\mathfrak{C}_1^{\text{modi}}(\varepsilon_n)$. Using Proposition 6.5 and (8.108) now shows that

$$d_{\text{GHP}}\left(\frac{\beta^{2/3}}{\mu\nu n^{1/3}} \bar{\mathscr{T}}_{\text{tilt}}^{\mathbf{p},\text{modi}}, \text{CRT}_{\text{tilt}}\right) \xrightarrow{\text{a.e.}} 0. \tag{8.109}$$

Let

$$\mathscr{T}^{\text{CM}} = \mathscr{T}_{\text{tilt}}^{\mathbf{p},\text{modi}} \cap \mathscr{C}_1(0), \qquad \bar{\mathscr{T}}^{\text{CM}} = \bar{\mathscr{T}}_{\text{tilt}}^{\mathbf{p},\text{modi}} \cap \mathscr{C}_1(0), \tag{8.110}$$

where viewing $\mathscr{T}^{\text{CM}}$ as a graph at the blob-level (ignoring internal structure), we have that $\mathscr{T}^{\text{CM}}$ is a connected tree. Finally using Lemma 8.33 on the relationship between the surplus edges in $\mathscr{C}_1(0)$ and $\mathfrak{C}_1^{\text{modi}}(\varepsilon_n)$, it is now enough to show that

$$d_{\text{GHP}}(n^{-1/3} \bar{\mathscr{T}}_{\text{tilt}}^{\mathbf{p},\text{modi}}, n^{-1/3} \bar{\mathscr{T}}^{\text{CM}}) \xrightarrow{\text{P}} 0. \tag{8.111}$$

Now note that $\mathfrak{C}_1^{\text{modi}}(\varepsilon_n)$ is obtained from $\mathscr{C}_1(0)$ by attaching some blobs to $\mathscr{C}_1(0)$. More precisely, there exist metric spaces $\mathscr{S}_1^{(n)}, \mathscr{S}_2^{(n)}, \ldots, \mathscr{S}_{r_n}^{(n)}$, each of which has a tree superstructure whose vertices are blobs; vertices $t_1^{(n)}, \ldots, t_{r_n}^{(n)}$ in $\mathscr{C}_1(0)$ and vertices $s_1^{(n)}, \ldots, s_{r_n}^{(n)}$ with $s_i^{(n)} \in \mathscr{S}_i^{(n)}$ such that $\mathscr{C}_1^{\text{modi}}$ is obtained by placing an edge between between $s_i^{(n)}$ and $t_i^{(n)}$ for all $1 \leqslant i \leqslant r_n$. Assume that

$$\text{diam}(\mathscr{S}_1^{(n)}) \geqslant \text{diam}(\mathscr{S}_2^{(n)}) \geqslant \cdots \geqslant \text{diam}(\mathscr{S}_{r_n}^{(n)}).$$

By [1, Theorem 2.9], (8.109) implies that $n^{-1/3}\bar{\mathscr{T}}^{\text{CM}}$ is pre-compact under the metric d_{GHP}. Since the limit metric space CRT_{tilt} is compact, via a diagonalization argument for every subsequence we can find a further subsequence $\{n_k : k \geq 1\}$ along which the following (possibly subsequence dependent) assertions hold:

(i) A connected compact metric space $\mathscr{X}$ such that $n^{-1/3}\mathscr{T}^{\mathrm{CM}}$ converges to $\mathscr{X}$ in d_{GHP} along that subsequence.

(ii) A collection of points $\{t_i^\infty : i \geq 1\} \in \mathscr{X}$ such that $t_i^{(n_k)} \to t_i^{(\infty)}$ in the associated correspondence. Again this can be done since $\mathscr{X}$ is compact.

(iii) There exist a sequence of compact metric spaces $\{S_i^\infty : i \geq 1\}$ with distinguished points $\{s_i^\infty : i \geq 1\}$ such that for all $i \geq 1$ in the rooted GHP topology we have

$$d_{\mathrm{GHP,rooted}}((n^{-1/3} S_i^{n_k}, s_i^{(n_k)}), (S_i^\infty, s_i^\infty)) \to 0. \tag{8.112}$$

(iv) Finally the limit metric space $\mathrm{CRT}_{\mathrm{tilt}}$ can be obtained by adjoining $\mathscr{S}_i^\infty$ with $\mathscr{X}$ for all $i \geq 1$ by identifying the points s_i^∞ with t_i^∞.

Thus, there exists a measure-preserving isometry between $\mathrm{CRT}_{\mathrm{tilt}}$ and $\mathscr{X}$ which carries $\mathscr{X}$ to a connected compact subset of $\mathrm{CRT}_{\mathrm{tilt}}$. Further, by Corollary 8.19 and Theorem 8.20, $\mathscr{X}$ has full measure in $\mathrm{CRT}_{\mathrm{tilt}}$. Thus the image of $\mathscr{X}$ is the whole $\mathrm{CRT}_{\mathrm{tilt}}$. This implies that (8.111) is true along the subsequence $\{n_k : k \geq 1\}$. However this implies that for any subsequence $\{n_k : k \geq 0\}$, there exists a further subsequence along which (8.111) holds. This implies (8.111) and completes the proof of Proposition 8.34.

■

8.7. **Percolation on the configuration model.** The aim of this section is to complete the proof of Theorem 4.7 regarding percolation on the configuration model. Recall the definition of the edge retention probability $p(\lambda)$ from (4.8). The basic idea is to just match the number of edges in the continuous time dynamic construction and the percolation model and then use the equivalence established in [34,36]. To simplify notation write $\mathrm{Perc}_n(p(\lambda))$ for the random graph obtained through percolation of $\mathrm{CM}_n(\infty)$ with edge retention probability $p(\lambda)$. Recall that $\mathrm{CM}_n(t)$ denotes the state of the continuous time construction at time t and we for this process the critical scaling window corresponds to time of the form, $t = t_c + \lambda'/n^{1/3}$. Here we use λ' to distinguish this parameter from the one used for $p(\lambda)$. Further note that Theorem 4.9 which has been proven in the previous sections establishes continuum scaling limits for $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$. Now recall the equivalence between the dynamic version of the configuration model and percolation as expounded in Lemma 8.3 and Proposition 8.4.

For the rest of the proof it will be convenient to parametrize each model by the number of half edges, denoted by $\mathscr{H}(\mathrm{Perc}_n(p(\lambda)))$ and $\mathscr{H}_n(\mathrm{CM}_n(t))$ respectively. Fix $\gamma < 1/2$. Standard tail bounds for the Binomial distribution imply that the number of half-edges used in $\mathrm{Perc}_n(p(\lambda))$ whp satisfies

$$\left|\mathscr{H}(\mathrm{Perc}_n(p(\lambda)) - \left(\frac{n\mu}{\nu} + n^{2/3}\mu\lambda\right)\right| \leq n^{1-\gamma}. \tag{8.113}$$

Using Lemma 8.2, in the dynamic construction, in $\mathrm{CM}_n(t_c + \lambda/n^{1/3})$ the number of used half-edges whp satisfies

$$\left|\mathscr{H}(\mathrm{CM}_n(t_c + \lambda'/n^{1/3})) - \left(\frac{n\mu}{\nu} + n^{2/3}\mu\frac{2(\nu-1)}{\nu}\lambda'\right)\right| \leq n^{1-\gamma}. \tag{8.114}$$

Comparing (8.113) and (8.114) and using Lemma 8.3 and Proposition 8.4, for fixed λ, taking $\lambda' = \lambda\nu/2(\nu-1)$ in (8.114), we find that for any fixed $1/3 < \gamma < 1/2$ we can couple $\mathrm{Perc}_n(p(\lambda))$ with the dynamic construction CM_n such that whp as graphs we have

$$\mathrm{CM}_n\left(t_c + \frac{\nu}{2(\nu-1)}\frac{\lambda}{n^{1/3}} - \frac{1}{n^{\gamma}}\right) \subseteq \mathrm{Perc}_n(p(\lambda)) \subseteq \mathrm{CM}_n\left(t_c + \frac{\nu}{2(\nu-1)}\frac{\lambda}{n^{1/3}} + \frac{1}{n^{\gamma}}\right) \qquad (8.115)$$

Using Theorem 4.9 for the two process sandwiching $\mathrm{Perc}_n(p(\lambda))$ completes the proof of Theorem 4.7. ■

9. PROOFS: SCALING LIMITS OF THE BOHMAN-FRIEZE PROCESS

The proof for this model is easier than the previous models since many of the hard technical estimates have already been proven in [11] where the sizes of the components in the critical scaling window were studied and shown to converge after appropriate normalization to the standard multiplicative coalescent. Further, the proof of the only additional approximation result we need, Proposition 9.3 on average inter-blob distances, follows almost identically to the corresponding result for the configuration model (Proposition 8.12).

In the next section we start by recalling various estimates from [13] and then prove Proposition 9.3. We then complete the proof in Section 9.2.

9.1. **Preliminaries for the BF model.** Recall the susceptibility functions $\bar{s}_2, \bar{s}_3$ and the density of singletons $x_n(\cdot)$ as well as their deterministic limits in (4.15), (4.16) and (4.17). As before write $I(t) = \mathscr{C}_i(t)$ and let $J(t)$ be the maximum diameter of a connected component at time t. Fix $\delta \in (1/6, 1/5)$ and let $t_n = t_c - n^{-\delta}$.

I. Bounds on the Maximal component: Using an associated inhomogeneous random graph model [21] with infinite dimensional type space $\chi = \mathbb{R}_+ \times \mathbb{D}([0,\infty) : \mathbb{N}_0)$ where $\mathbb{D}([0,\infty) : \mathbb{N})$ is the Skorohod D-space of rcll on the set of integers $\mathbb{N}$ and analyzing the size of a multitype branching process approximation of this random graphs [11, Proposition 1.2] and analyzing the size of a multitype branching process approximation of this random graphs shows that there exists a constant $B = B(\delta)$ such that

$$\mathbb{P}\left(I(t) \leqslant \frac{B\log^4 n}{(t_c - t)^2}, \text{ for all } t \in [0, t_n]\right) \to 1, \qquad (9.1)$$

as $n \to \infty$. In [55] it was shown that the technique in [13] can be strengthened much further and the term $\log^4 n$ term in the above bound can be replaced by $\log n$ which is optimal. Now instead of analyzing the total *size* of the approximating branching process, analyzing the number of individuals in generations as in the proof of the configuration model (Theorem 8.5 and (8.20)), the exact same proof in [13] (strengthened using [55]) shows the following.

Lemma 9.1. *There exist absolute constant $B_1, B_2 > 0$ such that*

$$\mathbb{P}\left(I(t) \leqslant \frac{B_1 \log n}{(t_c - t)^2},\ J(t) \leqslant \frac{B_2 \log n}{t_c - t}, \textit{ for all } t \in [0, t_n]\right) \to 1, \qquad (9.2)$$

as $n \to \infty$.

We omit the proof.

II. Concentration of susceptibility functions: Recall from Section 8.3 that a major ingredient in the analysis of the barely subcritical regime of the configuration model was showing concentration of associated susceptibility functions around their deterministic limits using the Semi-martingale approximation Lemma 8.13. In particular defining $Y(t) = 1/\bar{s}_2(t)$ and $Z(t) = \bar{s}_3(t)/\bar{s}_2^3(t)$ and similarly the deterministic analogously $y(t) = 1/s_2(t)$ and $z(t) = s_3(t)/[s_2(t)]^3$, [11, Lemma 6.4, Proposition 7.1 and 7.6] show that for any $\gamma < 1/2$

$$\sup_{0\leqslant t\leqslant t_n} \max\left(n^{\gamma}|\bar{x}(t)-x(t)|,\ \ n^{1/3}|Y(t)-y(t)|,\ \ |Z(t)-z(t)|\right) \xrightarrow{\mathrm{P}} 0, \tag{9.3}$$

as $n\to\infty$. A key ingredient was the following on the semi-martingale decomposition of these two processes.

Lemma 9.2 ([10, Section 6.4]). *We have*

$$\mathbf{d}(\bar{s}_2)(t) = \bar{x}^2 + (1-\bar{x}^2)\bar{s}_2^2 + O_{t_c}\left(\frac{I^2\bar{s}_2}{n}\right), \qquad \mathbf{v}(\bar{s}_2)(t) = O_{t_c}(I^2\bar{s}_2^2/n),$$

$$\mathbf{d}(Y)(t) = \bar{x}^2Y^2 + (1-\bar{x}^2) + O_{t_c}\left(\frac{I^2 Y}{n}\right), \qquad \mathbf{v}(Y)(t) = O_{t_c}\left(\frac{I^2Y^2}{n}\right).$$

Let us now setup notation for the remaining approximation ingredient required. For a graph $\mathscr{G}$ with vertex set $[n]$ and vertex v write $\mathscr{C}(v;\mathscr{G})$ for the connected component of v in $\mathscr{G}$. For two vertices u,v in the same component let $d(u,v)$ denote the graph distance between these two vertices. Let $\mathscr{D}(v) = \sum_{u\in\mathscr{C}(v;\mathscr{G})} d(u,v)$ denote the sum of all distances between v and all vertices in the same component as v. For a connected component $\mathscr{C}$, write $\mathscr{D}(\mathscr{C}) := \sum_{u\in\mathscr{C}}\mathscr{D}(u) = \sum_{u,v\in\mathscr{C}} d(u,v)$. Finally write

$$\mathscr{D}(\mathscr{G}) := \sum_{\mathscr{C}\subset\mathscr{G}} \mathscr{D}_k(\mathscr{C}), \qquad \mathscr{D}(t) := \mathscr{D}(\mathrm{BF}_n(t)), \qquad \bar{\mathscr{D}}(t) := \frac{\mathscr{D}(t)}{n}. \tag{9.4}$$

Now from (4.19), recall the function $v(\cdot)$ on $[0,t_c)$ obtained as the unique solution of the equation $v'(t) = F(x(t),y(t),v(t))$ with $v(0)=0$ where

$$F(x,y,v) := -2x^2yv + x^2y^2/2 + 1 - x^2. \tag{9.5}$$

Let $V(t) = \bar{\mathscr{D}}(t)/\bar{s}_2^2(t)$. The following is the analogue of (8.34) proved for the configuration model.

Proposition 9.3. *For $\delta\in(1/6,1/5)$, we have*

$$\sup_{t\in[0,t_n]} |V(t)-v(t)| \xrightarrow{\mathrm{P}} 0,$$

as $n\to\infty$.

Proof: The plan is to study the semi-martingale decomposition of $V = \bar{\mathscr{D}}/\bar{s}_2^2$ and use Lemma 8.13 coupled with bounds on the diameter and maximal components, Lemma 9.1 to prove convergence to the deterministic limit v. We start with the form of the semi-martingale decomposition of $\bar{\mathscr{D}}$.

Lemma 9.4. *The following hold.*

$$\mathbf{d}(\bar{\mathscr{D}})(t) = 2(1-\bar{x}^2)\bar{s}_2\bar{\mathscr{D}} + \frac{1}{2}\bar{x}^2 + (1-\bar{x}^2)\bar{s}_2^2 + O(\bar{s}_2 I^2 J/n), \tag{9.6}$$

$$\mathbf{v}(\bar{\mathscr{D}})(t) = O(\bar{s}_2^2 I^2 J^2/n). \tag{9.7}$$

Proof: The proof of (9.7) is identical to the corresponding result for the configuration model namely Lemma 8.10. Let us prove (9.6). Given a graph $\mathscr{G}$ and two vertices $i, j \in \mathscr{G}$, write $\mathscr{G}_{ij}$ for the graph obtained by adding an edge (i, j) to $\mathscr{G}$. When $\mathscr{C}(i) \neq \mathscr{C}(j)$, for any two vertices i', j' not in either of these two components Thus we have

$$\begin{aligned}\mathscr{D}(\mathscr{G}_{ij}) - \mathscr{D}(\mathscr{G}) =& 2\sum_{i'\in\mathscr{C}(i)}\sum_{j'\in\mathscr{C}(j)} [d(i', i) + d(j', j) + 1] \\ =& 2\mathscr{D}(i)|\mathscr{C}(j)| + 2\mathscr{D}(j)|\mathscr{C}(i)| + 2|\mathscr{C}(i)||\mathscr{C}(j)|,\end{aligned}$$

where the factor of two arises since every distance $d(i', j')$ is counted twice in $\mathscr{D}(\mathscr{G}_{ij})$. On the other hand, when the edge is place between vertices in the same component namely $\mathscr{C}(i) = \mathscr{C}(j) = \mathscr{C}$, then only distances within $\mathscr{C}$ changes, therefore we have

$$0 \geqslant \mathscr{D}(\mathscr{G}_{ij}) - \mathscr{D}(\mathscr{G}) \geqslant -\mathscr{D}(\mathscr{C}).$$

Let us now bring in the dynamics of the BF process which determines the rate of various edges forming. For any time t, depending on if the first edge connects two singletons or not, we can classify events into two classes:

(a) The first edge connects two singleton: There are n^2X^2 such quadruples, and this increase $\mathscr{D}(t)$ by 1.
(b) The first edge does not connect two singletons: There $(n^2 - X^2)$ different choice for the first two vertices in the quadruple, then the second edge can connect any two vertices.

Write $\Delta_{ij}(t) := \mathscr{D}(\mathscr{G}_{ij}) - \mathscr{D}(\mathscr{G})$, where $\mathscr{G}$ is taken to be $\mathrm{BF}_n(t)$. We have

$$\begin{aligned}\mathbf{d}(\mathscr{D})(t) =& \frac{1}{2n^3}\cdot n^2X^2\cdot 1 + \frac{1}{2n^3}(n^2 - X^2)\sum_{i,j\in[n]}\Delta_{ij}(t) \\ =& \frac{n\bar{x}^2}{2} + \frac{1-\bar{x}^2}{2n}\sum_{i,j\in[n]}\Delta_{ij}(t)\mathbb{1}_{\{\mathscr{C}(i)\neq\mathscr{C}(j)\}} + \frac{1-\bar{x}^2}{2n}\sum_{i,j\in[n]}\Delta_{ij}(t)\mathbb{1}_{\{\mathscr{C}(i)=\mathscr{C}(j)\}}.\end{aligned} \tag{9.8}$$

Expanding the second term above and recalling that $\mathscr{S}_2(t) = \sum_i |\mathscr{C}_i(t)|^2$ we have

$$\begin{aligned}\sum_{i,j\in[n]}\Delta_{ij}(t)\mathbb{1}_{\{\mathscr{C}(i)\neq\mathscr{C}(j)\}} = & 2\sum_{i,j\in[n]}(\mathscr{D}(i)|\mathscr{C}(j)| + \mathscr{D}(j)|\mathscr{C}(i)| + |\mathscr{C}(i)||\mathscr{C}(j)|)\mathbb{1}_{\{\mathscr{C}(i)\neq\mathscr{C}(j)\}} \\ =& 2\left[2\mathscr{D}(t)\mathscr{S}_2(t) + (\mathscr{S}_2(t))^2\right] - 2\sum_{\mathscr{C}\subset\mathrm{BF}_n(t)}\sum_{i,j\in\mathscr{C}}(\mathscr{D}(i)|\mathscr{C}| + \mathscr{D}(j)|\mathscr{C}| + |\mathscr{C}|^2) \\ =& 2\left[2\mathscr{D}(t)\mathscr{S}_2(t) + (\mathscr{S}_2(t))^2\right] - 2\sum_{\mathscr{C}\subset\mathrm{BF}_n(t)}\left(2\mathscr{D}(\mathscr{C})|\mathscr{C}|^2 + |\mathscr{C}|^4\right),\end{aligned} \tag{9.9}$$

where the second equation uses the fact $\sum_{i\in\mathscr{G}}\mathscr{D}(i) = \mathscr{D}(\mathscr{G})$ and $\sum_{i\in\mathscr{G}}|\mathscr{C}(i)| = \mathscr{S}_2(\mathscr{G})$ and we sum over all connected components $\mathscr{C}$ in $\mathrm{BF}_n(t)$. Combining (9.8) and (9.9) in (9.6), we collect the error terms as follows:

$$\begin{aligned}
&\left|\mathbf{d}(\bar{\mathscr{D}})(t) - \left[\frac{1}{2}\bar{x}^2 + (1-\bar{x}^2)(2\bar{s}_2\bar{\mathscr{D}} + \bar{s}_2^2)\right]\right| \\
=&\left|\frac{1-\bar{x}^2}{2n^2}\sum_{i,j\in\mathscr{G}(t)}\Delta_{ij}(t)\mathbb{1}_{\{\mathscr{C}(i)=\mathscr{C}(j)\}} - \frac{1-\bar{x}^2}{2n^2}\cdot 2\sum_{\mathscr{C}\subset\mathscr{G}(t)}\left(2\mathscr{D}(\mathscr{C})|\mathscr{C}|^2 + |\mathscr{C}|^4\right)\right| \\
\leq&\frac{1}{2n^2}\left[\sum_{\mathscr{C}\subset\mathscr{G}(t)}\sum_{i,j\in\mathscr{C}}\mathscr{D}(\mathscr{C}) + 2\sum_{\mathscr{C}\subset\mathscr{G}(t)}\left(2\mathscr{D}(\mathscr{C})|\mathscr{C}|^2 + |\mathscr{C}|^4\right)\right] \\
=&\frac{1}{2n^2}\left[5\sum_{\mathscr{C}\subset\mathscr{G}(t)}|\mathscr{C}|^2\mathscr{D}(\mathscr{C}) + 2\sum_{\mathscr{C}\subset\mathscr{G}(t)}|\mathscr{C}|^4\right],
\end{aligned}$$

where the above inequality uses the fact that $|\Delta_{ij}(t)\mathbb{1}_{\{i,j\in\mathscr{C}\}}| \leq \mathscr{G}(\mathscr{C})$. Then using $|\mathscr{C}| \leq I$ and $\mathscr{D}(\mathscr{C}) \leq I^2 J$ in the above bound proves (9.6).

■

Recall in the study of the configuration model, once we had the semi-martingale decomposition of $\bar{\mathscr{D}}$ and Y from Lemma 8.10, this lead to the corresponding result for $V = \bar{\mathscr{D}}Y^2$. The same proof but now using Lemma 9.2 and Lemma 9.4 proves the following.

Lemma 9.5. *Recall the function F from* (9.5). *Then, for the process $V(\cdot)$, we have*

$$\mathbf{d}(V)(t) = F(\bar{x}, Y, V) + O(JI^2Y/n), \tag{9.10}$$

$$\mathbf{v}(V)(t) = O(J^2I^2Y^2/n). \tag{9.11}$$

Completing the proof of Proposition 9.3: Now we are ready to prove Proposition 9.3 using the semimartingale approximation Lemma 8.13 as in Section 8.3. First note that with F as in (9.5), $g(t,u) := F(x(t), y(t), u)$ satisfies the assumption in Lemma 8.13. To check conditions (ii) and (iii) of Lemma 8.13 first note that

$$\begin{aligned}
|\mathbf{d}(V)(t) - g(t,V(t))| \leq& (1+J)\max\left\{\sup_{t\leq t_n}|\bar{x}(t)-x(t)|, \sup_{t\leq t_n}|Y(t)-y(t)|\right\} + O\left(\frac{JI^2Y}{n}\right) \\
=& O\left(\frac{1}{n^{1/3-\delta}}\right) + O\left(\frac{JI^2Y}{n}\right), \qquad \text{by (9.3)}.
\end{aligned}$$

Taking $\theta_2(n) = \theta_3(n) = n^{1-3\delta}$ and using Lemma 9.1, (9.3) for the approximation of Y by the deterministic limit y and the fact that $y(t) = O(t_c - t)$ as $t \uparrow t_c$ to get

$$\begin{aligned}
\int_0^{t_n}|\mathbf{d}(V)(t) - g(t,V(t))|dt =& O\left(\frac{1}{n}\int_0^{t_n} J(t)I^2(t)Y(t)dt\right) \\
=& O\left(\frac{1}{n}\int_0^{t_n}\frac{1}{(t_c-t)^4}dt\right) = O\left(\frac{1}{n^{1-3\delta}}\right).
\end{aligned}$$

Similarly we get

$$\int_0^{t_n}\mathbf{v}(V)(t)dt = O\left(\frac{1}{n}\int_0^{t_n}J^2(t)I^2(t)Y^2(t)dt\right) = O\left(\frac{1}{n}\int_0^{t_n}\frac{1}{(t_c-t)^4}dt\right) = O\left(\frac{1}{n^{1-3\delta}}\right).$$

Lemma 8.13 now completes the proof of Proposition 9.3

■

Combining the results in (4.18), (4.20), (9.3) and Proposition 9.3, we have the following asymptotics about $\mathrm{BF}_n(t_n)$.

Proposition 9.6. *There exist constants $\alpha \approx 1.063$, $\beta \approx .764$, $\varrho \approx .811$ such that as $n \to \infty$,*

$$n^{1/3}\left|\frac{1}{\bar{s}_2(t_n)} - \frac{1}{\alpha n^{\delta}}\right| \xrightarrow{\mathrm{P}} 0, \quad \frac{\bar{s}_3(t_n)}{\beta\alpha^3 n^{3\delta}} \xrightarrow{\mathrm{P}} 1, \quad \frac{\bar{\mathscr{D}}(t_n)}{\varrho\alpha^2 n^{2\delta}} \xrightarrow{\mathrm{P}} 1.$$

These asymptotics will be used in verifying Assumption 3.3 in Section 9.2 so that we can apply Theorem 3.4.

9.2. **Completing the proof of Theorem 4.12:** In this section, we will prove Theorem 4.12 via a sandwiching argument used in [11, 13] in the analysis of the sizes of components in the critical regime of the Bohman-Frieze process and general bounded size rules respectively. We only give a sketch of the proof. Recall that $t_n = t_c - n^{-\delta}$ for fixed $\delta \in (1/6, 1/5)$. Let $t_n^+ := t_c + n^{-\delta}$. For any fixed $t \geqslant 0$ let $\mathrm{BF}^*(t)$, be the graph obtained from $\mathrm{BF}(t)$ by deleting all the singletons. The goal is to construct two Erdős-Rényi type random graph processes $\mathbf{G}^-(t)$ and $\mathbf{G}^+(t)$ for $t \in [t_n, t_n^+]$ such that whp $\mathbf{G}^-(t) \subset \mathrm{BF}^*(t) \subset \mathbf{G}^+(t)$ for all $t \in [t_n, t_n^+]$ and to show that both $\mathbf{G}^-(t)$ and $\mathbf{G}^+(t)$ have the same scaling limit as using Theorem 4.12. Since $t_n^+ = t_c + \lambda_n/n^{1/3}$ where $\lambda_n = n^{1/3-\delta} \to \infty$, thus this completes the proof for the scaling limit of the maximal components for any time ${}^{\lambda} := t_c + \alpha\beta^{2/3}\lambda/n^{1/3}$ for any fixed $\lambda \in \mathbb{R}$.

We start with some notation required to define this sandwich argument used in [11, 13]. Given an initial graph G_0 with vertex set $\mathcal{V}(G_0)$ a subset of $[n]$, let $\{\mathrm{ER}_n(t; G_0) : t \geqslant 0\}$, be the Erdős-Rényi random graph process with initial graph G_0. More precisely

(i) Initialize the process at time $t = 0$ with $\mathrm{ER}_n(0; G_0) = G_0$.
(ii) Each edge $\{i, j\}$ with $i \neq j \in \mathcal{V}(G_0)$ is added at rate $1/n$.

Note that multi-edges are allowed in this construction. We will use this construction where the initial graph G_0 is also random. Now the two sandwiching processes $\mathbf{G}^-(t)$ and $\mathbf{G}^+(t)$, $t \in [t_n, t_n^+]$ are defined as follows:

$$\mathbf{G}^-(t) := \mathrm{ER}_n((t - t_n)(\alpha^{-1} - n^{-1/6}); G_0^-),$$
$$\mathbf{G}^+(t) := \mathrm{ER}_n((t - t_n)(\alpha^{-1} + n^{-1/6}); G_0^+),$$

where $G_0^- := \mathrm{BF}^*(t_n)$ and G_0^+ is defined as follow. For every edge in $\mathrm{BF}^*(t_n^+)$, we say the edge is “good” if it satisfies the following two conditions:

(a) It was added at some time $t \in [t_n, t_n^+]$.
(b) It connected two non-singleton components when it was added.

Then G_0^+ is the graph obtained from $\mathrm{BF}^*(t_n^+)$ by deleting all the “good” edges. This completes the construction of $\mathbf{G}^-(t)$ and $\mathbf{G}^+(t)$. Let $\mathscr{C}_i^*(t)$ [resp. $\mathscr{C}_i^-(t)$ and $\mathscr{C}_i^+(t)$] denote the i-th largest component of $\mathrm{BF}^*(t)$ [resp. $\mathbf{G}^-(t)$ and $\mathbf{G}^+(t)$] and as before let $t^\lambda := t_c + \alpha\beta^{2/3}\lambda/n^{1/3}$ for fixed $\lambda \in \mathbb{R}$.

Lemma 9.7 ([13], Lemma 7.2, Proposition 7.5)**.** *There is a coupling of the three processes $\{\mathbf{G}^-(t), \mathrm{BF}^*(t), \mathbf{G}^+(t) : t \in [t_n, t_n^+]\}$ on a common probability space such that*

$$\mathbb{P}\left(\mathbf{G}^-(t) \subset \mathrm{BF}^*(t) \subset \mathbf{G}^+(t) \textit{ for all } t \in [t_n, t_n^+]\right) \to 1 \textit{ as } n \to \infty.$$

Further, for any fixed $K > 0$ and $\lambda \in \mathbb{R}$, the maximal components at time t_λ satisfy

$$\mathbb{P}\left(\mathscr{C}_i^-(t^\lambda) \subset \mathscr{C}_i^*(t^\lambda) \subset \mathscr{C}_i^+(t^\lambda) \text{ for all } 1 \leqslant i \leqslant K\right) \to 1 \text{ as } n \to \infty.$$

Proof of Theorem 4.12: By Lemma 9.7 and properties of the Gromov-Hausdorff-Prokhorov distance, if suffice to show that both $\mathbf{G}^-(t^\lambda)$ and $\mathbf{G}^+(t^\lambda)$ have the same scaling limit as in (4.21). Arguing as in Proposition 8.34 this completes the proof.

First, we deal with $\mathbf{G}^-(t^\lambda)$. Let $\mathscr{C}_i^-(t)$, $t \in [t_n, t_n^+]$, be the i-th largest component of $\mathbf{G}^-(t^\lambda)$. We write $\mathscr{C}_i^- = \mathscr{C}_i^-(t_n)$. Define the blobs $\mathbf{M}^- = \{(M_i, d_i, \mu_i)\}_i$ where $M_i := \mathscr{V}(\mathscr{C}_i^-)$, d_i is the graph distance on $\mathscr{C}_i^-$, and μ_i is the uniform measure on M_i. Define $\mathbf{x}^- = (x_i)_i$ where $x_i := \beta^{1/3}|\mathscr{C}_i^-|/n^{2/3}$. Since edges are added between vertices at rate $1/n$, thus in $\mathbf{G}^-(t^\lambda)$, the number of edges between $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$ is a Poisson random variable with mean

$$\frac{1}{n}|\mathscr{C}_i^-||\mathscr{C}_j^-| \cdot (t^\lambda - t_n)\left(\frac{1}{\alpha} - \frac{1}{n^{1/6}}\right) = \frac{1}{n}|\mathscr{C}_i^-||\mathscr{C}_j^-| \cdot \left(\frac{\alpha\beta^{2/3}\lambda}{n^{1/3}} + \frac{1}{n^\delta}\right)\left(\frac{1}{\alpha} - \frac{1}{n^{1/6}}\right).$$

Defining

$$q^- := \frac{n^{1/3}}{\beta^{2/3}}\left(\frac{\alpha\beta^{2/3}\lambda}{n^{1/3}} + \frac{1}{n^\delta}\right)\left(\frac{1}{\alpha} - \frac{1}{n^{1/6}}\right) = \frac{n^{1/3-\delta}}{\alpha\beta^{2/3}} + \lambda + o(1), \tag{9.12}$$

we have

$$\mathbf{G}^-(t^\lambda) \approx_d \bar{\mathscr{G}}(\mathbf{x}^-, q^-, \mathbf{M}^-), \tag{9.13}$$

where the error in the above approximation is because in $\mathbf{G}^-(t^\lambda)$ we may have (1) multiple edges between $\mathscr{C}_i^-$ and $\mathscr{C}_j^-$, and (2) additional edges within $\mathscr{C}_i^-$. One can show that the total number of these extra edges in $\mathbf{G}^-(t^\lambda)$ is $O_P(1)$ as $n \to \infty$. By the bounds on $J(t_n)$ in Lemma 9.1, the effect of these edges after scaling is $O(n^{\delta-1/3}\log n) = o(1)$. The details are omitted.

For $k \geqslant 1$, let $\bar{s}_k^- = n^{-1}\sum_i |\mathscr{C}_i^-|^k$ be the susceptibility and $\bar{\mathscr{D}}^-$, I^- and J^- be the average distances, maximal component size and maximal diameter respectively for G_0^-. Since we only ignore singleton vertices in forming G_0^- is is easy to check that $\bar{s}_k^- - \bar{s}_k(t_n^-) = O(1)$ for $k = 2, 3$, $\bar{\mathscr{D}}^- = \bar{\mathscr{D}}$, $I^- = I$ and $J^- = J$. Therefore the asymptotic behavior of these constructs are the same as in Lemma 9.1 and Proposition 9.6. The key quantities in Theorem 3.4 are

$$\sigma_2 = \frac{\beta^{2/3}}{n^{1/3}}\bar{s}_2^- \sim \frac{\alpha\beta^{2/3}}{n^{1/3-\delta}}, \; \sigma_3 = \frac{\beta}{n}\bar{s}_3^- \sim \frac{\beta^2\alpha^3}{n^{1-3\delta}}, \; \sum_{i=1}^\infty x_i^2 u_i = \frac{\beta^{2/3}}{n^{1/3}}\bar{\mathscr{D}}^- \sim \frac{\varrho\alpha^2\beta^{2/3}}{n^{1/3-2\delta}},$$

$$x_{\max} = O(n^{2\delta-2/3}\log n), \; x_{\min} \geqslant \frac{\beta^{1/3}}{n^{2/3}}, \; d_{\max} = O(n^\delta \log n).$$

Since $\delta \in (1/6, 1/5)$, Assumption 3.3 is verified with any $\eta_0 \in (0, 1/2)$ and $r_0 \in [5, \infty)$. Using Theorem 3.4 on $\mathscr{G}(\mathbf{x}^-, q^-, \mathbf{M}^-)$ and noting that the weight of each vertex has been scaled by $\beta^{1/3}/n^{2/3}$ thus resulting in

$$\frac{\sigma_2^2}{\sigma_2 + \sum_{i=1}^\infty x_i^2 u_i} \sim \frac{\sigma_2^2}{\sum_{i=1}^\infty x_i^2 u_i} \sim \frac{\beta^{2/3}}{\varrho n^{1/3}}.$$

Combine this with (9.13) and using Theorem 3.4 gives

$$\left(\operatorname{scl}\left(\frac{\beta^{2/3}}{\varrho n^{1/3}}, \frac{\beta^{1/3}}{n^{2/3}}\right)\mathscr{C}_i^-(t^\lambda) : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty(\lambda), \qquad \text{as } n \to \infty. \tag{9.14}$$

Next, we treat $\mathbf{G}^+(t^\lambda)$. Let $\bar{s}_2^+$, $\bar{s}_3^+$, $\bar{\mathscr{D}}^+$, I^+ and J^+ be the corresponding quantities for G_0^+. Once we show that these random variables also have the same asymptotic behavior as in Lemma 9.1 and Proposition 9.6, the rest of the proof is identical to the above analysis for $\mathbf{G}^-(t^\lambda)$. The asymptotic behavior of $\bar{s}_2^+$, $\bar{s}_3^+$ and I^+ are analyzed in [13, Proposition 7.4]. We only need to show, as $n \to \infty$,

$$\frac{\bar{\mathscr{D}}^+}{\bar{\mathscr{D}}^-} \xrightarrow{\mathrm{P}} 1, \; J^+ - J = O(n^\delta \log n). \tag{9.15}$$

The argument follows as in [13, Proposition 7.4]. Here we sketch the details. Note that $\mathbf{G}^+(t_n) = G_0^+$ and there are two sources that contribute to the difference $\bar{\mathscr{D}}^+ - \bar{\mathscr{D}}$ and $J^+ - J$:

(1) The components in G_0^- is stochastically smaller than those in $\mathbf{G}_0^+$ by construction with each component $\mathscr{C}^- \subset G_0^-$ contained within a component $\mathscr{C}^+ \subset G_0^+$ owing to the attachment of singleton vertices to $\mathscr{C}^-$ in the time interval $[t_n, t_n^+]$.

(2) A new component is formed by connecting two singletons during the time interval $[t_n, t_n^+]$. This component may also grow in size after it was created.

We only bound the effect of the first case, the second case can be treated similarly. Note that there are always $O(n)$ number of singletons, thus the size of $\mathscr{C}^-$ grows at rate $O(n|\mathscr{C}^-|\cdot \frac{1}{n})$. Since $t_n^+ - t_n^- = O(n^{-\delta})$ we get $|\mathscr{C}^+| - |\mathscr{C}^-| = O(|\mathscr{C}^-|n^{-\delta})$. The increase in $\bar{\mathscr{D}}$ caused by adding one singleton to the component $\mathscr{C}^-$ can be bounded by $|\mathscr{C}^-|J/n$. Therefore

$$\bar{\mathscr{D}}^+ - \bar{\mathscr{D}}^- = O\left(\sum_{i\geqslant 0} |\mathscr{C}_i^-|n^{-\delta}\cdot|\mathscr{C}_i^-|J/n\right) = O\left(\frac{J}{n^\delta}\bar{s}_2^-\right) = O(n^\delta \log n)$$

Since $\bar{\mathscr{D}}^- \sim \varrho\alpha^2 n^{2\delta}$, then the above bound implies $\bar{\mathscr{D}}^+/\bar{\mathscr{D}}^- \xrightarrow{\mathrm{P}} 1$ as $n \to \infty$. The second asymptotics in (9.15) follows from $|\mathscr{C}^+| - |\mathscr{C}^-| = O(|\mathscr{C}^-|n^{-\delta}) = O(n^\delta \log n)$ using the bound on the maximal component in Lemma 9.1. Thus we have (9.15). By approximating $\mathbf{G}^+(t^\lambda)$ by $\bar{\mathscr{G}}(\mathbf{x}^+, q^+, \mathbf{M}^+)$ defined analogously to $\bar{\mathscr{G}}(\mathbf{x}^-, q^-, \mathbf{M}^-)$ and applying Theorem 3.4 we have

$$\left(\mathrm{scl}\left(\frac{\beta^{2/3}}{\varrho n^{1/3}}, \frac{\beta^{1/3}}{n^{2/3}}\right)\mathscr{C}_i^+(t^\lambda) : i \geqslant 1\right) \xrightarrow{\mathrm{w}} \mathbf{Crit}_\infty(\lambda), \qquad \text{as } n \to \infty. \tag{9.16}$$

Combining (9.15), (9.16) and Lemma 9.7, completes the proof of Theorem 4.12. ■

Acknowledgements

We would like to thank Amarjit Budhiraja and Steve Evans for many stimulating conversations. SB has been partially supported by NSF-DMS grants 1105581 and 1310002 and SES grant 1357622. SS has been supported in part by NSF grant DMS-1007524 and Netherlands Organisation for Scientific Research (NWO). XW has been supported in part by the National Science Foundation (DMS-1004418, DMS-1016441), the Army Research Office (W911NF-0-1-0080, W911NF-10-1-0158) and the US-Israel Binational Science Foundation (2008466).

References

[1] R. Abraham, J.-F. Delmas, and P. Hoscheit, *A note on the gromov-hausdorff-prokhorov distance between (locally) compact metric measure spaces*, Electron. J. Probab. **18** (2013), no. 14, 1–21.

[2] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *Critical random graphs: limiting constructions and distributional properties*, Electron. J. Probab **15** (2010), no. 25, 741–775.

[3] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *The continuum limit of critical random graphs*, Probability Theory and Related Fields **152** (2012), no. 3-4, 367–406.

[4] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, arXiv preprint arXiv:1301.1664 (2013).

[5] D. Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, The Annals of Probability (1997), 812–854.

[6] D. Aldous, G. Miermont, and J. Pitman, *The exploration process of inhomogeneous continuum random trees, and an extension of jeulin's local time identity*, Probability theory and related fields **129** (2004), no. 2, 182–218.

[7] D. J Aldous and B. Pittel, *On a random graph with immigrating vertices: Emergence of the giant component*, Random Structures and Algorithms **17** (2000), no. 2, 79–102.

[8] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York-Heidelberg, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. MR0373040 (51 #9242)

[9] E. A. Bender and E. R. Canfield, *The asymptotic number of labeled graphs with given degree sequences*, J. Combinatorial Theory Ser. A **24** (1978), no. 3, 296–307. MR0505796 (58 #21793)

[10] S. Bhamidi, A. Budhiraja, and X. Wang, *The augmented multiplicative coalescent and critical dynamic random graph models*, Accepted in Probability Theory and Related Fields (2012).

[11] S. Bhamidi, A. Budhiraja, and X. Wang, *Aggregation models with limited choice and the multiplicative coalescent*, Random Structures & Algorithms (2013), n/a–n/a.

[12] S. Bhamidi, A. Budhiraja, and X. Wang, *Aggregation models with limited choice and the multiplicative coalescent*, Random Structures & Algorithms (2013).

[13] S. Bhamidi, A. Budhiraja, and X. Wang, *The augmented multiplicative coalescent, bounded size rules and critical dynamics of random graphs*, Probability Theory and Related Fields (2013), 1–64 (English).

[14] S. Bhamidi, A. Budhiraja, and X. Wang, *Bounded-size rules: The barely subcritical regime*, Combinatorics, Probability and Computing **FirstView** (20146), 1–34.

[15] S. Bhamidi, S. Sen, and X. Wang, *Continuum limit of critical inhomogeneous random graphs*, arXiv preprint arXiv:1404.4118 (2014).

[16] S. Bhamidi, R. van der Hofstad, and J. van Leeuwaarden, *Novel scaling limits for critical inhomogeneous random graphs*, The Annals of Probability **40** (2012), no. 6, 2299–2361.

[17] S. Bhamidi, R. van der Hofstad, and J. S. H. van Leeuwaarden, *Scaling limits for critical inhomogeneous random graphs with finite third moments*, Electron. J. Probab. **15** (2010), no. 54, 1682–1703. MR2735378 (2011m:60021)

[18] T. Bohman and A. Frieze, *Avoiding a giant component*, Random Structures & Algorithms **19** (2001), no. 1, 75–85.

[19] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), no. 4, 311–316. MR595929 (82i:05045)

[20] B. Bollobás, *Random graphs*, Second, Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. MR1864966 (2002j:05132)

[21] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures & Algorithms **31** (2007), no. 1, 3–122.

[22] B. Bollobás and O. Riordan, *The phase transition in the Erdős-Rényi random graph process*, Erdös centennial, 2013, pp. 59–110. MR3203594

[23] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, 2013.

[24] L. A Braunstein, S. V Buldyrev, R. Cohen, S. Havlin, and H E. Stanley, *Optimal paths in disordered complex networks*, Physical review letters **91** (2003), no. 16, 168701.

[25] T. Britton, M. Deijfen, and A. Martin-Löf, *Generating simple random graphs with prescribed degree distribution*, Journal of Statistical Physics **124** (2006), no. 6, 1377–1397.

[26] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Vol. 33, American Mathematical Society Providence, 2001.

[27] M. Camarri and J. Pitman, *Limit distributions and random trees derived from the birthday problem with unequal probabilities*, Electron. J. Probab **5** (2000), no. 2, 1–18.

[28] F. Chung and L. Lu, *Connected components in random graphs with given expected degree sequences*, Annals of combinatorics **6** (2002), no. 2, 125–145.

[29] F. R. K Chung and L. Lu, *Complex graphs and networks*, AMS Bookstore, 2006.

[30] R. Durrett, *Random graph dynamics*, Vol. 20, Cambridge university press, 2007.

[31] P. Erdős and A. Rényi, *On random graphs. I*, Publ. Math. Debrecen **6** (1959), 290–297. MR0120167 (22 #10924)

[32] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986. Characterization and convergence. MR838085 (88a:60130)

[33] S. N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR2351587 (2009d:60014)

[34] N. Fountoulakis, *Percolation on sparse random graphs with given degree sequence*, Internet Mathematics **4** (2007), no. 4, 329–356.

[35] S. Janson and J. Spencer, *Phase Transitions for Modified Erdos–Rényi Processes*, Arxiv preprint arXiv:1005.4494 (2010).

[36] S. Janson, *On percolation in random graphs with given vertex degrees*, Electron. J. Probab. **14** (2009), no. 5, 87–118. MR2471661 (2010b:60023)

[37] S. Janson, *Asymptotic equivalence and contiguity of some random graphs*, Random Structures & Algorithms **36** (2010), no. 1, 26–45.

[38] S. Janson, *Susceptibility of random graphs with given vertex degrees*, J. Comb. **1** (2010), no. 3-4, 357–387. MR2799217 (2012c:05278)

[39] S. Janson, D. E Knuth, T. Łuczak, and B. Pittel, *The birth of the giant component*, Random Structures & Algorithms **4** (1993), no. 3, 233–358.

[40] S. Janson and M. J. Luczak, *Susceptibility in subcritical random graphs*, J. Math. Phys. **49** (2008), no. 12, 125207, 23. MR2484338 (2010b:05152)

[41] S. Janson and M. J Luczak, *A new approach to the giant component problem*, Random Structures & Algorithms **34** (2009), no. 2, 197–216.

[42] S. Janson, T. Luczak, and A. Rucinski, *Random graphs*, Vol. 45, John Wiley & Sons, 2011.

[43] S. Janson and O. Riordan, *Susceptibility in inhomogeneous random graphs*, Electron. J. Combin. **19** (2012), no. 1, Paper 31, 59. MR2880662

[44] A. Joseph, *The component sizes of a critical random graph with given degree sequence*, arXiv preprint arXiv:1012.2352 (2010).

[45] T. G. Kurtz, *Strong approximation theorems for density dependent Markov chains*, Stochastic Processes Appl. **6** (1977/78), no. 3, 223–240. MR0464414 (57 #4344)

[46] K. Lange, M. Boehnke, and R. Carson, *Moment computations for subcritical branching processes*, Journal of Applied Probability (1981), 52–64.

[47] J.-F. Le Gall, *Random trees and applications*, Probab. Surv. **2** (2005), 245–311. MR2203728 (2007h:60078)

[48] T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures & Algorithms **1** (1990), no. 3, 287–310.

[49] M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput. **7** (1998), no. 3, 295–305. MR1664335 (2000c:05130)

[50] A. Nachmias and Y. Peres, *Critical random graphs: diameter and mixing time*, The Annals of Probability (2008), 1267–1286.

[51] A. Nachmias and Y. Peres, *Critical percolation on random regular graphs*, Random Structures & Algorithms **36** (2010), no. 2, 111–148.

[52] I. Norros and H. Reittu, *On a conditionally poissonian graph process*, Advances in Applied Probability **38** (2006), no. 1, 59–75.

[53] J. Pitman, *Random mappings, forests, and subsets associated with abel-cayley-hurwitz multinomial expansions*, Sém. Lothar. Combin **46** (2001), 02.

[54] O. Riordan, *The phase transition in the configuration model*, Combinatorics, Probability and Computing **21** (2012), no. 1-2, 265–299.
[55] S. Sen, *On the largest component in the subcritical regime of the bohman-frieze process*, arXiv preprint arXiv:1307.2041 (2013).
[56] J. Spencer and N. Wormald, *Birth control for giants*, Combinatorica **27** (2007), no. 5, 587–628.
[57] T. S Turova, *Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1*, arXiv preprint arXiv:0907.0897 (2009).
[58] R. Van Der Hofstad, *Random graphs and complex networks*, Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf (2009).
[59] R. van der Hofstad, *Critical behavior in inhomogeneous random graphs*, Random Structures & Algorithms **42** (2013), no. 4, 480–508.
[60] N. C Wormald, *Differential equations for random processes and random graphs*, The annals of applied probability (1995), 1217–1235.