

# SAMPLING FROM POTTS ON RANDOM GRAPHS OF UNBOUNDED DEGREE VIA RANDOM-CLUSTER DYNAMICS

ANTONIO BLANCA AND REZA GHEISSARI

**ABSTRACT.** We consider the problem of sampling from the ferromagnetic Potts and random-cluster models on a general family of random graphs via the Glauber dynamics for the random-cluster model. The random-cluster model is parametrized by an edge probability  $p \in (0, 1)$  and a cluster weight  $q > 0$ . We establish that for every  $q \geq 1$ , the random-cluster Glauber dynamics mixes in optimal  $\Theta(n \log n)$  steps on  $n$ -vertex random graphs having a prescribed degree sequence with bounded average branching  $\gamma$  throughout the full high-temperature uniqueness regime  $p < p_u(q, \gamma)$ .

The family of random graph models we consider includes the Erdős–Rényi random graph  $G(n, \gamma/n)$ , and so we provide the first polynomial-time sampling algorithm for the ferromagnetic Potts model on Erdős–Rényi random graphs for the full tree uniqueness regime. We accompany our results with mixing time lower bounds (exponential in the largest degree) for the Potts Glauber dynamics, in the same settings where our  $\Theta(n \log n)$  bounds for the random-cluster Glauber dynamics apply. This reveals a novel and significant computational advantage of random-cluster based algorithms for sampling from the Potts model at high temperatures.

## 1. INTRODUCTION

The ferromagnetic Potts model is a classical spin system model in statistical physics and computer science. It is defined on a finite graph  $G = (V, E)$ , by a set of spins (or colors)  $[q] = \{1, \dots, q\}$  and an edge weight or inverse temperature parameter  $\beta > 0$ . A configuration  $\sigma \in \{1, \dots, q\}^V$  of the model is an assignment of spins to the vertices of  $V$ . The probability of  $\sigma$  is given by the Gibbs distribution:

$$\mu_{G, \beta, q}(\sigma) = \frac{1}{Z_{G, \beta, q}} \exp(-\beta D(\sigma)), \quad (1.1)$$

where  $D(\sigma) = |\{\{v, w\} \in E : \sigma(v) \neq \sigma(w)\}|$  is the number of edges whose endpoints have different spins in  $\sigma$ , and  $Z_{G, \beta, q}$  is a normalizing factor known as the partition function. The Ising model of ferromagnetism corresponds to the case where  $q = 2$ .

Sampling from the Potts Gibbs distribution (1.1) is one of the most frequently encountered problems when running simulations in statistical physics or when solving a variety of inference tasks in computer science; see e.g. [25, 26, 33, 34, 51, 54, 56] and the references therein for a sample of these applications. There is a family of powerful sampling algorithms for the Potts model that are based on its *random-cluster representation*, defined subsequently. Such algorithms, which include the Glauber dynamics of the random-cluster model and the widely-used Swendsen–Wang dynamics, are an attractive option computationally since they are often efficient at “low-temperatures” (large  $\beta$ ), a parameter regime where standard Markov chains for the Potts model (including the canonical Glauber dynamics) often converge exponentially slowly; see, e.g., [11, 13, 14, 19].

To be more precise, the *random-cluster model* on a finite graph  $G = (V, E)$ , is defined by an edge probability parameter  $p \in (0, 1)$  and a cluster weight  $q > 0$ . The set of configurations of the model is the set of all subsets of edges  $\omega \subseteq E$ . The probability of each configuration  $\omega$  is given by the Gibbs distribution:

$$\pi_{G, p, q}(\omega) = \frac{1}{Z_{G, p, q}} p^{|\omega|} (1-p)^{|E|-|\omega|} q^{c(\omega)}, \quad (1.2)$$

where  $c(\omega)$  is the number of connected components (also called clusters) in the subgraph  $(V, \omega)$ , and  $Z_{G, p, q}$  is the corresponding partition function. The random-cluster model was introduced by Fortuin and Kasteleyn [27] as a unifying framework for studying random graphs, spin systems, and electrical networks, and it is also known as the *FK-representation* of the Ising and Potts model.

For integer  $q \geq 2$ , a sample  $\omega \subseteq E$  from the random-cluster Gibbs distribution  $\pi_{G,p,q}$  can be easily transformed into one for the ferromagnetic  $q$ -state Potts model with inverse temperature  $\beta(p) = -\ln(1-p)$ , by independently assigning a random spin from  $\{1, \dots, q\}$  to (all vertices in) each connected component of  $(V, \omega)$ ; see, e.g., [24, 27, 38]. As such, any sampling algorithm for the random-cluster model yields one for the ferromagnetic Potts model with essentially no computational overhead. This has led to significantly improved sampling algorithms for the Potts model in various low-temperature settings [12, 30, 43, 49, 58, 60] and more generally, to a broad interest in dynamics for the random-cluster model [4–8, 15, 39].

In this paper, we focus on the Glauber dynamics of the random-cluster model, which for easy distinction we will henceforth call the *FK-dynamics*. From a configuration  $\omega_t \subseteq E$ , one step of this Markov chain transitions to a new configuration  $\omega_{t+1} \subseteq E$  as follows:

- (1) Choose an edge  $e_t \in E$  uniformly at random;
- (2) Set  $\omega_{t+1} = \omega_t \cup \{e_t\}$  with probability  $\begin{cases} \hat{p} := \frac{p}{q(1-p)+p} & \text{if } e_t \text{ is a “cut-edge” in } (V, \omega_t); \\ p & \text{otherwise;} \end{cases}$
- (3) Otherwise set  $\omega_{t+1} = \omega_t \setminus \{e_t\}$ .

Here, we say  $e$  is a *cut-edge* in  $(V, \omega_t)$  if changing the state of  $e_t$  changes the number of connected components  $c(\omega_t)$  in  $(V, \omega_t)$ . The probabilities in step (2) are exactly the conditional probabilities of  $e_t$  being in the configuration  $\omega_t$  given the remainder of  $\omega_t$ . As such, this Markov chain is reversible with respect to  $\pi_{G,p,q}$  and converges to it. We are interested in its *mixing time*  $t_{\text{MIX}}$ ; i.e., the number of steps until the dynamics is within variation distance  $1/4$  of  $\pi_{G,p,q}$ , starting from the worst possible initial configuration.

As mentioned, the FK-dynamics is by now well-studied in its own right, though sharp analyses of its mixing time are only available on certain structured graphs like the complete graph [6, 8, 37], boxes in the infinite integer lattice graph  $\mathbb{Z}^d$  [5, 7, 14, 32, 35, 36, 41], and trees [2]. Recently, in [4], the authors studied the FK-dynamics on random regular graphs and established optimal  $\Theta(n \log n)$  mixing time for the FK-dynamics throughout the entire high-temperature tree uniqueness regime.

Our aim in this paper is to study the FK-dynamics in settings in which the maximum degree of the underlying graph is much larger than its *average* degree. In such settings, high-degree vertices are an obstruction to the fast convergence of the Ising/Potts Glauber dynamics. For instance, we later prove (see Section 1.2) that on a general class of random graphs on  $n$  vertices with maximum degree  $d_{\text{MAX}}$ , the Ising/Potts Glauber dynamics requires  $n \cdot \exp(\Omega(d_{\text{MAX}}))$  steps to converge at high temperatures.

We reveal here that, for the same general family of random graphs, random-cluster based algorithms are not affected by the presence of high-degree vertices; both their mixing times and fast mixing parameter regimes are determined instead by the *average degree* of the graph. This reveals a novel and significant computational advantage of random-cluster based algorithms for sampling from the ferromagnetic Potts model *at high temperatures*. Indeed, prior to this work, random-cluster based sampling algorithms were only found to be more efficient than Ising/Potts Glauber dynamics at low temperatures.

More precisely, we study the mixing time of the FK-dynamics on random graphs of average branching  $\gamma > 0$  in the full uniqueness (high-temperature) regime  $p < p_u(q, \gamma)$ . At integer  $\gamma$ , the threshold  $p_u(q, \gamma)$ , formally defined in (2.1), was identified in [40] as a uniqueness/non-uniqueness phase transition point of the random-cluster model on the *wired*  $\gamma$ -ary tree, i.e., where the leaves are externally wired to be in the same connected component. For us,  $p_u(q, \gamma)$  is the natural extension of that function to non-integer  $\gamma$ , which we show corresponds to the high-temperature uniqueness threshold of the random-cluster model on general trees of average branching  $\gamma$  for all  $q \geq 1$  (see Corollary 3.4 in Section 3).

Before we describe our general results for random graph models with fixed degree sequence (which we define in the next subsection) we present a special case of our main result of particular interest concerning the FK-dynamics on sparse Erdős–Rényi random graphs.

**Theorem 1.1.** *Fix  $q \geq 1$ ,  $\gamma > 0$  and  $p < p_u(q, \gamma)$ . If  $\mathcal{G}$  is an Erdős–Rényi random graph  $\mathcal{G} \sim G(n, \gamma/n)$ , then with probability  $1 - o(1)$ ,  $\mathcal{G}$  is such that the FK-dynamics on  $\mathcal{G}$  satisfies  $t_{\text{MIX}} = \Theta(n \log n)$ .*

This yields a sampler for the Potts distribution on Erdős–Rényi random graphs with near-optimal running time. Let  $\beta_u(q, \gamma) = -\ln(1 - p_u(q, \gamma))$  be the corresponding uniqueness point for the Potts model.

**Corollary 1.2.** Fix  $q \geq 2$ ,  $\gamma > 0$  and  $\beta < \beta_u(q, \gamma)$ . There is an MCMC sampling algorithm that, with probability  $1 - o(1)$  over the choice of an Erdős–Rényi random graph  $\mathcal{G} \sim G(n, \gamma/n)$ , outputs a configuration whose distribution is within total-variation distance  $\delta > 0$  of  $\mu_{\mathcal{G}, \beta, q}$  in time  $O(n(\log n)^3 \log(1/\delta))$ .

Corollary 1.2 is a direct consequence of Theorem 1.1 and the aforementioned connection between the random-cluster model and the Potts model. The extra  $O((\log n)^2)$  factor in the running time of the algorithm comes from the (amortized) cost of checking whether the chosen edge is a cut-edge in each step of the FK-dynamics (see, e.g., [44, 59]).

To the best of our knowledge, this is the first polynomial-time sampling algorithm for the Potts model on Erdős–Rényi random graphs for  $q \geq 3$  and  $\beta = \Omega(1)$ . Even for the better understood  $q = 2$  case (i.e., the Ising model), Corollary 1.2 provides the fastest known sampling algorithm, improving upon the running time of samplers based on the Glauber dynamics which, for the Ising model, is known to converge in  $n^{1+\Theta(\frac{1}{\log \log n})}$  steps for all  $\beta < \beta_u(2, \gamma)$  [53].

We mention that the thresholds  $p_u(q, \gamma)$  and  $\beta_u(q, \gamma)$  should be sharp, in the sense that the FK-dynamics is conjectured to undergo polynomial or exponential slowdowns (depending on  $q$ ) at the point  $p_u(q, \gamma)$  (and when  $q > 2$  in a whole critical window  $(p_u, p'_u)$ ). This is by analogy with the FK-dynamics on the complete graph [37] and on random regular graphs [18]; see also [20, 31, 43].

**1.1. Results on random graphs with general degree sequences.** We next provide our main results on random graph models with a fixed degree sequence. Let  $\mathbf{d}_n = (d_1, \dots, d_n)$  be the degree sequence giving the degree of each vertex  $v \in \{1, \dots, n\}$ . Our results will hold for uniform random graphs with degree sequence  $\mathbf{d}_n$  under certain mild conditions on this degree sequence. The first condition we make on  $\mathbf{d}_n$  is that the sequence is *graphical*: i.e., that there exists at least one simple graph having degree sequence  $\mathbf{d}_n$ .

Given a graphical sequence  $\mathbf{d}_n$ , we define  $\mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  as the uniform distribution over all simple graphs on  $n$  vertices having degree sequence  $\mathbf{d}_n$ . The governing quantity in this degree sequence, in terms of the uniqueness thresholds for the Potts and random-cluster models on  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$ , will be what we call the *effective offspring distribution*  $\mathbb{P}_{\mathbf{d}_n}$ , which is defined as the distribution over the set  $\mathcal{M}(\mathbf{d}_n) = \{d_v - 1 : v \in \{1, \dots, n\}\}$  where  $x \in \mathcal{M}(\mathbf{d}_n)$  is assigned probability:

$$\mathbb{P}_{\mathbf{d}_n}(x) = \frac{\sum_v (x+1) \mathbf{1}_{\{d_v=x+1\}}}{\sum_v d_v}. \quad (1.3)$$

In words, the distribution  $\mathbb{P}_{\mathbf{d}_n}$  corresponds to choosing  $d_v - 1$  with probability proportional to the total degree of vertices having degree  $d_v$ . This distribution governs the offspring distribution corresponding to the random trees one obtains when looking at balls of small radius around a vertex of a random graph  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$ . Specifically, a vertex of degree  $d$  is selected to be the next vertex added to the random tree with probability proportional to the total degree of all such vertices, and once it is selected and connected to its parent, it has  $d - 1$  available edges to connect to other randomly chosen vertices.

Our results will apply to graphical degree sequences whose effective offspring distribution has a certain mean, and has bounded finite moments, as we detail next.

**Definition 1.3.** Let  $\mathcal{D}_{\gamma, \kappa}$  be the set of graphical degree sequences  $(\mathbf{d}_n)_n$  such that  $D \sim \mathbb{P}_{\mathbf{d}_n}$  has mean that is uniformly bounded away from  $\gamma$  and uniformly bounded  $\kappa$ -th moment. Formally,

$$\limsup_n \mathbb{E}_{\mathbf{d}_n}[D] < \gamma \quad \text{and} \quad \limsup_n \mathbb{E}_{\mathbf{d}_n}[D^\kappa] < \infty.$$

Let us finally assume that  $\sum_{1 \leq v \leq n} d_v = \Omega(n)$ ; this is not strictly necessary, but will simplify presentation.

This framework is fairly standard in the random graphs literature [10] and is similar to e.g., the setting of [29] for studying sampling from Potts on random graphs with fixed degree sequences at sufficiently low temperatures. While Definition 1.3 yields a fairly general family of random graphs, we draw attention to some well-studied examples which fall under its umbrella.

**Example 1.4.**  *$\Delta$ -regular random graph.* In this case,  $\mathbf{d}_n = (\Delta, \dots, \Delta)$  and the effective offspring distribution simply assigns probability 1 to  $\Delta - 1$ ; thus  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  for every  $\gamma > \Delta - 1$  and every  $\kappa$ .

**Example 1.5.** *Erdős–Rényi random graph  $G(n, \lambda/n)$ .* It was shown in [47] that if  $\mathbf{d}_n$  is drawn as an i.i.d. sequence of Poisson random variables of mean  $\lambda$ , then  $\mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  is contiguous with respect to  $G(n, \lambda/n)$ . (Two random graph models are *contiguous* when any sequence of events that has a probability of  $1 - o(1)$  in one has a probability of  $1 - o(1)$  in the other model as well.) Hence, it suffices to prove the desired results with high probability over such  $\mathbf{d}_n$  (see Lemma 4.8). Standard concentration estimates for Poisson random variables (see Lemma 4.9) then give that for every  $\gamma > \lambda$  and every  $\kappa$ , with high probability,  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ .

Our main result is an optimal mixing time bound for the FK-dynamics on  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$ , which applies to all the examples above and more generally to random graphs with degree sequences in  $\mathcal{D}_{\gamma, \kappa}$ .

**Theorem 1.6.** *Fix  $q \geq 1$ ,  $\gamma > 0$ , and  $p < p_u(q, \gamma)$ . There exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then with probability  $1 - o(1)$ , the FK-dynamics on  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  satisfies  $t_{\text{MIX}} = \Theta(n \log n)$ .*

This parameter regime in Theorem 1.6 is tight as FK-dynamics have been very recently shown [18] to exponentially slow down as soon as  $p > p_u(q, \gamma)$  for random regular graphs (Example 1.4) at integer  $q > 2$ .

The proof of the upper bound in Theorem 1.6 is the main content of this paper. As mentioned, the special case of the  $\Delta$ -regular random graph (i.e.,  $\mathbf{d}_n = (\Delta, \dots, \Delta)$ ) was the content of an earlier paper [4]. However, as soon as the degree sequence is not homogeneous, substantial further obstacles arise.

First, even the uniqueness threshold for the random-cluster model on wired heterogeneous trees (specifically, with offspring distribution  $\mathbb{P}_{\mathbf{d}_n}$ ) had not been established. In our proof of Theorem 1.6 we require something much stronger; namely, an exponential decay of connectivities with the correct rate (see Lemma 2.7). In the regular case, the fact that  $p_u(q, \gamma)$  is the uniqueness threshold goes back to the work of Häggström [40] (see also [3, 46]), and the exponential decay rate was established in [4]. To establish analogous results for the heterogeneous case, we combine the approach of [50] (which considered the special case of the Ising model  $q = 2$ ) with ideas from [3], so as to recurse, not on the marginal of an edge of the tree, but rather on a nice functional of its probability of downwards connection to infinity.

The second technical obstacle concerns establishing that the FK-dynamics on  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  *shatters*, i.e., that its components have size at most  $O(n^\varepsilon)$  after  $O(n)$  steps of the dynamics. This is proved using a delicate revealing procedure for the random graph with the FK-dynamics configuration on top of it, a technique introduced in [4] for the case of random regular graphs. The heterogeneity of the degrees in the current setting, however, introduces extra correlations between the underlying graph and the FK-dynamics configuration, necessitating substantial modifications to the revealing procedure from [4].

The changes we make to deal with the above-described dependencies include: (i) modifications to the revealing process so that it is based on half-edges rather than vertices and the dynamics is run in continuous time, and (ii) a new criteria to truncate potentially unbounded increments in the revealing procedure. The more robust procedure yields a notable further improvement: we show that the shattering time is  $O(n)$  (as opposed to  $O(n \log n)$  in [4]). Though this improvement has no impact on the eventual mixing time bound, the more precise understanding of the shattering phase may be useful in other settings.

A more detailed proof sketch of this theorem and the new complications that arise is provided in Section 2 and Remark 5.11.

**1.2. Slowdown for the corresponding Potts Glauber dynamics.** Returning to the advantage of FK-dynamics in the presence of high-degree vertices, the following theorem establishes that in the same setting as Theorem 1.6 the Ising/Potts Glauber dynamics slows down exponentially in the maximum degree.

**Theorem 1.7.** *Fix  $q \geq 1$ ,  $\gamma > 0$  and  $\beta < \beta_u(q, \gamma)$ . Then there exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then with probability  $1 - o(1)$ ,  $\mathcal{G} \sim \mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  is such that the Glauber dynamics for the Potts model on  $\mathcal{G}$  has  $t_{\text{MIX}} = n \cdot \exp(\Omega(\|\mathbf{d}_n\|_\infty))$ .*

Intuitively, the slowdown comes from the fact that the neighborhood of a vertex of degree  $\|\mathbf{d}_n\|_\infty$  is a star graph, in which the Ising/Potts Glauber dynamics mixes slowly when  $\beta \gg \frac{1}{\|\mathbf{d}_n\|_\infty}$ . In a random graph at high temperatures (i.e., when  $\beta < \beta_u(q, \gamma)$ ) there is essentially no interference with this effect from the

remainder of the graph. Note that the FK-dynamics in the star graph is fast mixing at all temperatures, so this obstruction is not present.

**Remark 1.8.** We remark that under various decay of correlation conditions (see, e.g., [16, 22, 23, 42]) the mixing time of this chain is known to be  $\text{poly}(n)$  when (roughly)  $\beta \leq 1/\|\mathbf{d}_n\|_\infty$ . This does not contradict Theorem 1.7, which holds when  $\beta = \Omega(1)$ . In fact, if one tracks the dependence on  $\beta$  in our proof, it gives  $t_{\text{MIX}} = n \cdot \exp(\Omega(\beta^2 \|\mathbf{d}_n\|_\infty))$ .

The known  $n^{1+\Omega(\frac{1}{\log \log n})}$  slowdown of the Ising/Potts Glauber dynamics on the Erdős–Rényi random graph [52, 53] is a special case of Theorem 1.7 where  $\|\mathbf{d}_n\|_\infty = \Theta(\frac{\log n}{\log \log n})$ . Below are a few examples where this slowdown can be even more dramatic, indeed stretched exponential in the total number of vertices.

**Example 1.9. Power-law degree distributions.** Consider graphical sequences  $(\mathbf{d}_n)_n$  satisfying item (1) in Definition 1.3, and for which the fraction of degrees of size  $\ell$  is  $\Theta(\ell^{-\zeta})$ . For every  $\kappa$ , if  $\zeta > \kappa + 2$ , one would have  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ . In such situations,  $\|\mathbf{d}_n\|_\infty = \Theta(n^{1/\zeta})$  and  $t_{\text{MIX}} = \exp(\Omega(n^{1/\zeta}))$ .

**Example 1.10. Planted high-degree vertices.** Consider a random  $\Delta$ -regular random graph and change the degree of one vertex to  $\Theta(n^\varepsilon)$ . If  $\varepsilon < 1/(\kappa + 1)$  and  $\gamma > \Delta - 1$ , then  $(\mathbf{d}_n) \in \mathcal{D}_{\gamma, \kappa}$  and  $t_{\text{MIX}} = \exp(\Omega(n^\varepsilon))$ .

In the above instances where the maximum degree is polynomial in  $n$ , there is an exponential vs. polynomial difference in the high-temperature mixing times of the Ising/Potts Glauber dynamics and of the FK-dynamics. At this level, the computational benefits of random-cluster based sampling methods also extend to the often implemented Swendsen–Wang dynamics [58]. In particular, using the comparison inequalities from [60] the upper bounds of Theorems 1.1 and 1.6 translate into  $O(n^2 \log n)$  upper bounds on the mixing time of the Swendsen–Wang dynamics in those settings.

**Acknowledgements.** The authors thank the anonymous referee for their helpful comments. The research of A.B. was supported in part by NSF grants CCF-1850443 and CCF-2143762. R.G. thanks the Miller Institute for Basic Research in Science for its support.

## 2. PROOF OUTLINE

In this section, we present the main technical contributions in our paper, and describe how they combine to yield the mixing time upper bound of Theorem 1.6.

**Notational disclaimers.** Throughout the paper, a subset  $\omega \subset E$  is naturally identified with an assignment of  $\{0, 1\}$ , or closed and open, to  $E$ , via  $\omega(e) = 1$  if and only if  $e \in \omega$ . The parameters  $p, q, \gamma$  will always be fixed quantities, and all constants in little-o, big-O, etc. notations may depend on these. As such, we also drop  $p, q$  from subscripts when understood from context, e.g.,  $\pi_G = \pi_{G, p, q}$ . All our results should be understood to hold uniformly over all sufficiently large  $n$ . We use  $C$  to generally denote the existence of a constant (possibly depending on fixed parameters such as  $p, q, \gamma$ ) such that the relevant statement holds for all large  $n$ ; for ease of notation, this constant  $C$  may change from line to line.

**2.1. Random graphs.** We start by describing the locally treelike structure and exponential rate of volume growth of random graphs with fixed degree sequence  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ . It will be convenient to work with the *configuration model*, a useful and standard tool for studying random graphs with fixed degree sequence. The configuration model  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is a distribution over multigraphs on  $n$  vertices with degree sequence  $\mathbf{d}_n$ . It is defined by giving  $d_v$  half-edges to every vertex  $v$  and drawing a uniform at random perfect matching on the  $\sum_v d_v$  many half-edges to form the  $\frac{1}{2} \sum_v d_v$  edges of the graph [9]. It is a standard fact that for any  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , and any sequence of sets  $A_n$  of simple graphs on  $n$  vertices, we have

$$\mathbb{P}_{\text{RG}(\mathbf{d}_n)}(\mathcal{G} \in A_n) = o(1) \quad \text{if and only if} \quad \mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \in A_n) = o(1) :$$

see [9, 28]. It thus suffices to prove Theorems 1.6–1.7 for  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ .

For a graph  $G = (V, E)$  and a vertex  $v \in V$ , we define the ball of radius  $R$  around  $v$  as:

$$B_R(v) := \{w \in V : d(w, v) \leq R\},$$



where  $d(\cdot, \cdot)$  is the graph distance. For a set  $B \subset V$  define  $E(B) = \{\{v, w\} \in E : v, w \in B\}$ .

**Definition 2.1.** We say that a graph  $G = (V, E)$  is  $L$ -Treelike if there is a set  $H \subset E$  with  $|H| \leq L$  such that the graph  $(V, E \setminus H)$  is a tree. We say that  $G$  is  $(L, R)$ -Treelike if for every  $v \in V$  the subgraph  $(B_R(v), E(B_R(v)))$  is  $L$ -Treelike.

The following lemma says that small balls of the random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  are close to trees. Indeed, for  $R/\log_\gamma n$  uniformly less than  $1/2$ , the ball  $B_R(v)$  in  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is typically a random tree with offspring distribution approximately  $\mathbb{P}_{\mathbf{d}_n}$ , defined in (1.3).

**Lemma 2.2.** *There exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  the following holds. For every  $\delta > 0$ , there exists  $L = L(\delta)$  such that if  $1 \leq R \leq (\frac{1}{2} - \delta) \log_\gamma n$ , we have*

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \text{ is } (L, R)\text{-Treelike}) = 1 - o(n^{-10}).$$

Using standard concentration estimates for the volume of Galton–Watson trees (see Lemma 3.6), we establish that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  has average exponential rate  $\gamma$  of volume growth.

**Definition 2.3.** A graph  $G = (V, E)$  on  $n$  vertices is said to have  $(\gamma, \varepsilon)$ -volume growth if for every  $v \in V$  and every integer  $r \in [\varepsilon \log_\gamma n, \frac{1}{2} \log_\gamma n]$  the graph has  $|B_r(v)| \leq \gamma^r$ .

**Lemma 2.4.** *Fix  $\varepsilon \in (0, \frac{1}{2})$ . There exists  $\kappa(\varepsilon)$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then*

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \text{ has } (\gamma, \varepsilon)\text{-growth}) \geq 1 - o(n^{-10}).$$

**2.2. Exponential decay and uniqueness on general trees and treelike graphs.** Given the local tree structure of the random graphs from  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , to control the decay rate of connectivities of the random-cluster model on  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , we need to first understand how these connectivities decay on heterogeneous (i.e., non-regular) trees. The relevant random-cluster measure on the tree requires the addition of *boundary conditions* mimicking the possible presence of open edges in the random graph outside of the treelike ball. Towards this, let us formally define boundary conditions.

**Definition 2.5.** A random-cluster *boundary condition*  $\xi$  on  $G = (V, E)$  is a partition of  $V$ , such that the vertices in each element of the partition are identified with one another. The random-cluster measure with boundary conditions  $\xi$ , denoted  $\pi_{G, p, q}^\xi$ , is the same as in (1.2) except the number of connected components  $c(\omega) = c(\omega; \xi)$  would be counted with this vertex identification, i.e., if  $v, w$  are in the same element of  $\xi$ , they are always counted as being in the same connected component of  $\omega$  in (1.2). The boundary condition can alternatively be seen as external “wirings” of the vertices in the same element of  $\xi$ .

**Remark 2.6.** The *free* boundary condition,  $\xi = 0$ , corresponds to the case of no external wirings; i.e., its partition is the one consisting of only of singletons. For a subset  $\partial V \subset V$ , the *wired* boundary condition on  $\partial V$ , denoted  $\xi = 1$ , is the one whose partition has all vertices of  $\partial V$  in the same element (and all vertices of  $V \setminus \partial V$  as singletons); i.e.,  $\xi = \{\partial V\} \cup \bigcup \{v : v \in V \setminus \partial V\}$ . For boundary conditions  $\xi, \xi'$  we say  $\xi \leq \xi'$  if  $\xi$  is a finer partition than  $\xi'$ . When  $q \geq 1$ , the random-cluster model has the following monotonicity property: for any two boundary conditions  $\xi \geq \xi'$ ,  $\pi_{G, p, q}^\xi \geq \pi_{G, p, q}^{\xi'}$  where  $\geq$  denotes stochastic domination [38].

Now define the threshold

$$p_u(q, \gamma) := 1 - \frac{1}{1 + \inf_{y \geq 1} h(y)}, \quad \text{where} \quad h(y) := \frac{(y-1)(y^\gamma + q - 1)}{y^\gamma - y}. \quad (2.1)$$

The work [40] studied the random-cluster measure on homogeneous,  $d$ -ary trees, with wired boundary conditions and identified  $p_u(q, d)$  as the *uniqueness threshold* such that whenever  $p < p_u(q, d)$ , the probability that the root is connected to a distance  $h$  in the wired  $d$ -ary tree goes to zero as  $h \rightarrow \infty$ ; a different proof was given in [3]. In [4], it was shown that this decay is in fact exponential with rate  $\hat{p} = p/(p + q(1 - p))$ . However, the methods of those papers do not easily extend to the non-regular setting, where there may be vertices of unbounded degree, but one would expect the threshold for connectivity decay to only depend

on the *average* branching rate. In [50], it was shown that the analogue  $\beta_u(2, \gamma)$  of (2.1) gives the correct uniqueness threshold in the case of the Ising model  $q = 2$ , for general (non-homogenous) trees of average branching  $\gamma$ . However, the argument there recursed over the single-site spin marginals, and relied on the fact that it was an Ising model whose interactions are nearest-neighbor. In the case of the random-cluster model, interactions between edge-marginals are non-local, and we therefore have to work with a more complicated functional encoding the probability of an edge being downward connected to the wired boundary. Combining ideas from [50] and [3], we are then able to establish uniqueness, and that connectivities decay exponentially with rate  $\hat{p}$  on general heterogeneous trees of average branching factor  $\gamma$  for all  $q \geq 1$  and all  $p < p_u(q, \gamma)$ . When  $p < p_u(q, \gamma)$ , we have  $\hat{p} < 1/\gamma$  (see e.g., [40, Theorem 1.5]); this indicates by a union bound why there will typically be no connections to the boundary in a tree of average branching  $\gamma$ .

More formally, let  $\mathcal{T}_h = (V(\mathcal{T}_h), E(\mathcal{T}_h))$  be an arbitrary finite tree, rooted at  $\rho$ , and of height  $h$ . Let  $\partial\mathcal{T}_h \subset V(\mathcal{T}_h)$  be the set of vertices of  $\mathcal{T}_h$  at distance exactly  $h$  from  $\rho$ . For  $v \in V(\mathcal{T}_h)$ , let  $\mathcal{T}_v$  be the subtree of  $\mathcal{T}_h$  rooted at  $v$ , let  $h(v)$  denote the height of  $\mathcal{T}_v$ , and let  $\partial\mathcal{T}_v = \partial\mathcal{T}_h \cap \mathcal{T}_v$ . For a random-cluster configuration  $\omega$  on  $\mathcal{T}_h$ , let  $\mathcal{C}_\rho(\omega)$  denote the connected component of  $\omega$  that contains the root  $\rho$  of  $\mathcal{T}_h$ . Finally, let  $(1, \mathfrak{G})$  denote the boundary condition that wires all vertices of  $\partial\mathcal{T}_h$  together, and also wires them up to the root, and let  $\pi_{\mathcal{T}_h}^{(1, \mathfrak{G})}$  be the random-cluster measure with this boundary condition.

**Lemma 2.7.** *Fix  $q \geq 1$ ,  $\gamma > 1$ ,  $p < p_u(q, \gamma)$ , and  $\varepsilon \in [0, 1)$ . Suppose that  $|\partial\mathcal{T}_v| \leq \gamma^{h(v)}$  for every  $v \in V(\mathcal{T}_h)$  with  $h(v) > \varepsilon h$ . Then, there exists a constant  $C = C(p, q, \gamma)$  such that for any  $u \in \partial\mathcal{T}_h$*

$$\pi_{\mathcal{T}_h}^{(1, \mathfrak{G})}(\omega : u \in \mathcal{C}_\rho(\omega)) \leq C\hat{p}^{(1-\varepsilon)h}.$$

We note that the condition that  $|\partial\mathcal{T}_v| \leq \gamma^{h(v)}$  for every  $v \in V(\mathcal{T}_h)$  with  $h(v) > \varepsilon h$  in the lemma holds with high probability for random trees with averaging branching  $\gamma$ : see Corollary 3.8. In addition, the exponential decay rate in Lemma 2.7 is essentially optimal, and together with Lemmas 2.2–2.4, allows us to derive precise estimates on the exponential decay of connectivities on the treelike balls around each vertex of the random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ . We will actually need a sharp bound on the rate of influence decay between the boundary and the center of the ball  $B_R(v)$ ; we find that this is the square of the rate of connectivity decay on a corresponding tree of depth  $R$ . (Intuitively, this is because *two* disjoint paths are required to reach the center of the ball in order for the boundary to have any effect on it.) To be more precise, let  $G = (V, E)$  be a graph and for  $v \in V$ , let  $E_v \subseteq E$  denote the set of edges incident to  $v$ .

**Definition 2.8.** A random-cluster boundary condition  $\xi$  on a graph  $H$  is said to be  $K$ -Sparse if the number of vertices in non-trivial (non-singleton) boundary components of  $\xi$  is at most  $K$ .

**Theorem 2.9.** *Fix  $\gamma > 0$ ,  $q \geq 1$ , and  $p < p_u(q, \gamma)$ . Suppose  $G$  is  $(L, R)$ -Treelike for some  $L$  and some  $R \leq \frac{1}{2} \log_\gamma n$ . Also suppose  $G$  has  $(\gamma, \varepsilon)$ -volume growth for some  $\varepsilon > 0$  sufficiently small. There exists a constant  $C > 0$  such that for every  $v \in G$ , and any two  $K$ -Sparse boundary conditions  $\xi$  and  $\tau$  on  $B_R(v)$ :*

$$\|\pi_{B_R(v)}^\xi(\omega(E_v) \in \cdot) - \pi_{B_R(v)}^\tau(\omega(E_v) \in \cdot)\|_{\text{TV}} \leq C\hat{p}^{(2-CL\sqrt{\varepsilon})R}.$$

A similar influence decay bound was proven for the regular case in [4, Section 5.2].

**2.3. Shattering of the FK-dynamics.** With Theorem 2.9 in hand, the core of our argument becomes establishing that the boundary conditions induced by the FK-dynamics chains from all possible initializations, on balls of radius  $R \leq \frac{1}{2} \log_\gamma n$  are  $K$ -Sparse. This will follow from *shattering* of the FK-dynamics, by which we mean the time at which the connected components of the FK-dynamics configuration are all small, say of size  $n^{o(1)}$ .

**Remark 2.10.** It will be technically convenient to prove our results in continuous time instead of discrete time. In the continuous-time FK-dynamics, each edge of the graph has a rate-1 Poisson clock and every time a clock rings, the corresponding edge is updated as in the discrete-time version of the FK-dynamics; that is, according to the conditional distribution given the configuration off of this edge. It is a standard fact (see

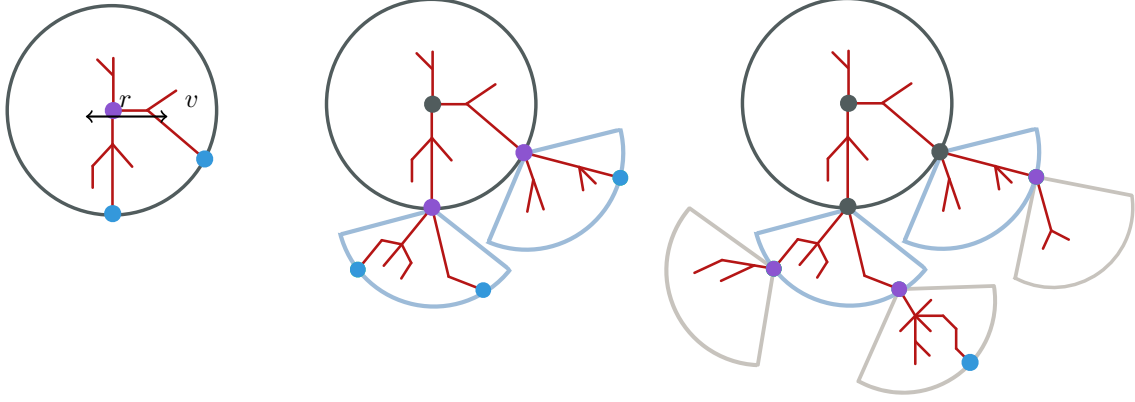


FIGURE 2.1. Three “generations” of the revealing procedure. In each figure, the purple vertices are the current generation of exposed vertices; the revealing procedure reveals the ball of radius  $r$  around such a vertex  $v$ , and a dominating localized FK-dynamics configuration  $\tilde{\omega}(B_r(v))$  on that ball. The next generation of exposed vertices (blue) consists of those on the boundary  $B_r(v)$  that are in the connected component of  $v$  in the configuration  $\tilde{\omega}(B_r(v))$ . Exposed vertices from previous generations are then colored black.

e.g., [48, Theorem 20.3]) that the discrete-time mixing time is comparable to  $|E(\mathcal{G})|$  times the continuous-time mixing time. It therefore suffices for us to establish the mixing time bounds of Theorems 1.1 and 1.6 as  $\Theta(\log n)$  bounds for the continuous-time version of the FK-dynamics. From this point on, we let  $X_t^{x_0}$  denote the continuous-time FK-dynamics on  $\mathcal{G}$  initialized from the configuration  $x_0$ , and use the superscripts 1 and 0 to denote the full (all-open) and empty (all-closed) configurations, respectively.

We now formalize what we mean by a shattered random-cluster configuration, and establish that the FK-dynamics shatters after an  $O(1)$  continuous-time burn in period.

**Definition 2.11.** A random-cluster configuration  $\omega$  on  $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$  is  $(K, R)$ -Sparse if, for every  $v \in V(\mathcal{G})$ , the boundary conditions induced on  $B_R(v)$  by  $\omega(E(\mathcal{G}) \setminus E(B_R(v)))$  are  $K$ -Sparse.

**Theorem 2.12.** Fix  $q \geq 1$ ,  $\gamma > 0$  and  $p < p_u(q, \gamma)$ . For every  $\delta > 0$ , there exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , there exists  $T = T(p, q, \gamma)$  and  $K = K(p, q, \gamma, \delta)$  such that for any  $t \geq T$ , and every  $1 \leq R \leq (\frac{1}{2} - \delta) \log_\gamma n$ , with probability  $1 - o(1)$ ,  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is such that

$$\mathbb{P}(X_t^1 \text{ is } (K, R)\text{-Sparse}) \geq 1 - o(n^{-5}). \quad (2.2)$$

Our starting point for the proof of Theorem 2.12 is a proof of shattering for the FK-dynamics on  $\Delta$ -regular random graphs from [4]. Hence, as in [4], our proof relies on a delicate simultaneous revealing procedure for the random graph, along with the connected component of a vertex  $v$  in  $X_t^1$ , showing that after a burn-in period, the configuration  $X_t^1$  is shattered. The revealing scheme for the component of a vertex  $v$  in the FK-dynamics chain  $X_t^1$  roughly proceeds as follows (see the accompanying Figure 2.1). First “expose” the starting vertex  $v$ , and iteratively, for each exposed vertex  $u$  do the following:

- (1) Reveal the ball  $B_r(u)$  in the random graph for a large  $r = O(1)$ ;
- (2) Reveal a configuration  $\tilde{\omega}(B_r(u))$  that dominates the configuration of the FK-dynamics at time  $t$  on  $B_r(u)$ . This configuration will come from simulating FK-dynamics that ignores all updates outside of  $B_r(u)$  (effectively inducing the wired boundary condition on  $B_r(u)$ ) and thus can be obtained independently of the dynamics on the rest of the graph;
- (3) Add to the set of exposed vertices all vertices of  $\partial B_r(u)$  that get connected to  $u$  in  $\tilde{\omega}(B_r(u))$ .

The key point of the argument is then to stochastically dominate the exposed vertices by a branching process, which can be shown to be sub-critical (see Lemma 2.7). In our setting, the heterogeneity of the degrees causes substantial complications to the argument from [4], because in balls where the branching rate is



locally larger than  $\gamma$ , the overlayed FK-dynamics configuration will actually be highly connected. The presence of high degrees also destroys the  $O(1)$  bounds on the maximum number of new vertices that could possibly get exposed in step (3) above; this complicates relevant concentration arguments, as our branching process martingale will no longer have bounded increments.

**2.4. Organization of the remainder of the paper.** In Section 3, we prove that whenever  $p < p_u(q, \gamma)$ , the random-cluster model on trees of average branching  $\gamma$  is in its uniqueness regime, and deduce Lemma 2.7. In Section 4, we prove key properties of the random-graph model  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , including Lemmas 2.2–2.4. Section 5 contains the proof of shattering of the FK-dynamics, and in particular Theorem 2.12. In Section 6, we bound the rate of influence decay (Theorem 2.9) and mixing time (Lemma 6.7) in treelike graphs with sparse boundary conditions. Section 7 combines these ingredients to conclude the  $\Theta(n \log n)$  bound on the FK-dynamics for Theorem 1.6. Finally, Section 8 proves the exponential (in  $\|\mathbf{d}_n\|_\infty$ ) lower bound on the Potts Glauber dynamics of Theorem 1.7.

### 3. UNIQUENESS AND EXPONENTIAL DECAY ON GENERAL TREES

Our main result in this section is to prove Lemma 2.7. We also use this section to deduce some corollaries about uniqueness of infinite-volume random-cluster and Potts measures on general trees of average branching  $\gamma$ , and apply these results to super-critical Galton–Watson trees.

**3.1. Exponential decay of connectivities on general trees.** We begin by considering the probability  $\varphi(\rho)$  that the root  $\rho$  is connected to  $\partial\mathcal{T}_h$  in  $\omega \sim \pi_{\mathcal{T}_h}^1$ , and show that  $\varphi(\rho) = \varphi_{p,q,\mathcal{T}_h}(\rho)$  decays exponentially with  $h$  for all trees of average branching  $\gamma$ , whenever  $p < p_u(q, \gamma)$ .

**Lemma 3.1.** *Fix  $\gamma \geq 1$  and  $q \geq 1$  and let  $p < p_u(q, \gamma)$ . There exists  $\theta = \theta(p, q, \gamma) \in (0, 1)$  and  $C = C(p, q, \gamma)$  such that if  $|\partial\mathcal{T}_h| \leq \gamma^h$ , then  $\varphi(\rho) \leq C\theta^h$ .*

*Proof.* Fix  $h$  and fix  $\mathcal{T}_h$  having  $|\partial\mathcal{T}_h| \leq \gamma^h$ . Recall that for  $v \in V(\mathcal{T}_h)$ ,  $\mathcal{T}_v$  denotes the subtree of  $\mathcal{T}_h$  rooted at  $v$ . Let  $Z(v) = Z_{\mathcal{T}_v, p, q}$  denote the partition function corresponding to  $\pi_{\mathcal{T}_v}^1$  (the random-cluster measure on  $\mathcal{T}_v$  with all its vertices in  $\partial\mathcal{T}_h$  wired together). Let  $Z_1(v)$  be the contribution to  $Z(v)$  from the configurations on  $\mathcal{T}_v$  that contain an open path between  $v$  and  $\partial\mathcal{T}_h$ . Similarly, let  $Z_0(v)$  denote the contribution from the configurations that do not have such a path. Note that  $Z(v) = Z_0(v) + Z_1(v)$  and  $\varphi(\rho) = \frac{Z_1(\rho)}{Z_0(\rho) + Z_1(\rho)}$ .

For  $v \in V(\mathcal{T}_h)$ , let  $N_v$  denote the set of children of  $v$ . Using tree recurrences, and the definition of (1.2), the following identities can be checked; the proof is similar to that in [3, Lemma 33] and is provided later.

**Fact 3.2.** *Let  $t = p/q + 1 - p$ . For any  $v \in V(\mathcal{T}_h)$ ,*

$$\begin{aligned} Z_1(v) &= q \prod_{w \in N_v} \left( \frac{Z_1(w)}{q} + \frac{tZ_0(w)}{q} \right) - q \prod_{w \in N_v} \left( \frac{(1-p)Z_1(w)}{q} + \frac{tZ_0(w)}{q} \right), \\ Z_0(v) &= q^2 \prod_{w \in N_v} \left( \frac{(1-p)Z_1(w)}{q} + \frac{tZ_0(w)}{q} \right). \end{aligned}$$

Now consider the function  $f : V(\mathcal{T}_h) \rightarrow \mathbb{R}$  defined as

$$f(v) := q \frac{Z_1(v)}{Z_0(v)} + 1.$$

Using the identities in Fact 3.2, one easily sees that

$$f(v) = \prod_{w \in N_v} g(f(w)) \quad \text{for} \quad g(x) := \frac{x + (q-1)(1-p)}{(1-p)x + p + (q-1)(1-p)}.$$

The following calculus bound, which is proved later, holds for the function  $g$ .

**Fact 3.3.** *Fix  $q, \gamma \geq 1$  and  $p < p_u(q, \gamma)$ . There exists  $\xi \in (0, 1/\gamma)$  such that  $g(x) \leq x^{1/\gamma - \xi}$  for all  $x \geq 1$ .*

Now, let  $D_k \subset V(\mathcal{T}_h)$  denote the set of vertices at distance  $k$  from the root  $\rho$  and let  $L_k \subseteq D_k$  be the set of leaves at distance  $k$  from  $\rho$ . Setting  $\zeta = 1/\gamma - \xi$ , and using the facts that  $g(1) = 1$ , and that if  $w$  is a leaf that does not belong to  $\partial\mathcal{T}_h$  then  $Z_1(w) = 0$  and  $g(f(w)) = 1$ , we obtain

$$f(\rho) = \prod_{w \in D_1} g(f(w)) = \prod_{w \in D_1 \setminus L_1} g(f(w)) \leq \prod_{w \in D_1 \setminus L_1} f(w)^\zeta.$$

Iterating, and using the fact that  $g(x) \leq (1-p)^{-1}$  for all  $x \geq 1$ , we have

$$f(\rho) \leq \prod_{w \in D_{h-1} \setminus L_{h-1}} f(w)^{\zeta^{h-1}} \leq (1-p)^{-\zeta^{h-1} |\partial\mathcal{T}_h|}.$$

Then, recalling  $\varphi(\rho) = Z_1(\rho)/(Z_0(\rho) + Z_1(\rho))$ , we get

$$\varphi(\rho) \leq \frac{Z_1(\rho)}{Z_0(\rho)} = \frac{f(\rho) - 1}{q} \leq \frac{1}{q} \left( \frac{1}{1-p} \right)^{\zeta^{h-1} |\partial\mathcal{T}_h|} - \frac{1}{q} \leq \frac{\zeta^h \cdot |\partial\mathcal{T}_h|}{q(1-p)^{1/\zeta}},$$

where the last inequality follows from the fact that  $a^x \leq 1+ax$  when  $a \geq 1$  and  $x \in [0, 1]$  since  $\zeta^h \cdot |\partial\mathcal{T}_h| \leq 1$  when  $|\partial\mathcal{T}_h| \leq \gamma^h$ . The proof is complete by setting  $\theta = 1 - \gamma\xi$ .  $\square$

With Lemma 3.1 on hand, we can now provide the proof of Lemma 2.7, which gives a precise bound on the rate of decay under stronger assumptions for the growth of  $\mathcal{T}_h$ .

**Proof of Lemma 2.7.** Let  $u$  be a vertex in  $\partial\mathcal{T}_h$  and for  $v \in V(\mathcal{T}_h)$  let  $\vartheta(v, u)$  be the probability that  $v$  is connected to  $u$  in  $\mathcal{T}_v$  under  $\pi_{\mathcal{T}_v}^1$ . Let  $\vartheta^\zeta(v, u)$  be the probability of the same event under  $\pi_{\mathcal{T}_v}^{(1, \zeta)}$ .

By monotonicity we have  $\vartheta(\rho, u) \leq \vartheta^\zeta(\rho, u)$  and by a standard comparison between boundary conditions (see e.g., Lemma 6.4), we have  $\vartheta^\zeta(\rho, u) \leq q\vartheta(\rho, u)$ . Hence, it suffices to bound  $\vartheta(\rho, u)$ . Consider the unique path  $P = (\rho = v_0, v_1, \dots, u = v_h)$  between  $\rho$  and  $u$ . Let  $N_v$  denote the set of children of  $v$ . For  $w \in N_{v_0}$ , let  $I_w$  be the indicator function of the event that there is a path from  $v_0$  to  $\partial\mathcal{T}_h$  going through  $w$ ; set  $I = \sum_{w \in N_\rho: w \neq v_1} I_w$ . Then, we can write

$$\vartheta(\rho, u) \leq p \cdot \pi_{\mathcal{T}_h}^1(I \geq 1) \cdot \vartheta^\zeta(v_1, u) + \hat{p}\vartheta(v_1, u) \leq \vartheta(v_1, u) [pq^2 \cdot \pi_{\mathcal{T}_h}^1(I \geq 1) + \hat{p}].$$

In the first inequality, we used the fact that in order for the root to be connected to the vertex  $u$ , it is required that the root is connected to  $v_1$ , and that  $v_1$  is connected to  $u$  in its sub-tree. The former event occurs with probability  $p$  or  $\hat{p}$ , depending on whether or not the root is connected to  $\partial\mathcal{T}_h$  through any child besides  $v_1$ .

Let  $\varphi(w)$  denote the probability that  $w$  is connected to  $\partial\mathcal{T}_w$  under  $\pi_{\mathcal{T}_w}^1$ . Then,  $\pi_{\mathcal{T}_h}^1(I \geq 1) \leq \varphi(v_0)$  and since  $|\partial\mathcal{T}_{v_0}| \leq \gamma^{h(v_0)}$  by assumption, Lemma 3.1 implies that for suitable constants  $\theta = \theta(p, q, \gamma) \in (0, 1)$  and  $C = C(p, q, \gamma) > 0$ , we have  $\pi_{\mathcal{T}_h}^1(I \geq 1) \leq C\theta^{h(v_0)}$ . Thus, setting  $a = \frac{Cpq^2}{\hat{p}}$ , and continuing the recursion we obtain

$$\begin{aligned} \vartheta(\rho, u) &\leq \hat{p} \cdot \vartheta(v_1, u) \left[ 1 + a \cdot \theta^{h(v_0)} \right] \leq \hat{p}^{(1-\varepsilon)h} \prod_{i=0}^{(1-\varepsilon)h} \left[ 1 + a \cdot \theta^{h(v_i)} \right] \\ &\leq \hat{p}^{(1-\varepsilon)h} \exp \left[ a \sum_{i=0}^{(1-\varepsilon)h} \theta^{h(v_i)} \right] \leq A\hat{p}^{(1-\varepsilon)h}, \end{aligned}$$

for a suitable constant  $A = A(p, q, \gamma) > 0$ . Hence,  $\vartheta^\zeta(\rho, u) \leq Aq^2\hat{p}^{(1-\varepsilon)h}$  and the result follows.  $\square$

**3.2. Proofs of auxiliary facts.** We now provide the deferred proofs of Facts 3.2 and 3.3.

**Proof of Fact 3.2.** For  $v \in V(\mathcal{T}_h)$ , let  $N_v$  denote the set of children of  $v$  and let  $\partial\mathcal{T}_v \subseteq \partial\mathcal{T}_h$  be the set of vertices of  $\mathcal{T}_v \cap \partial\mathcal{T}_h$ . We compute  $Z_1(v)$  and  $Z_0(v)$  by partitioning the space of configurations according to which subtrees of  $v$  among  $\{\mathcal{T}_u : u \in N_v\}$  are connected to the  $\partial\mathcal{T}_v$ . For each configuration  $\omega$ , the connectivity of the children of  $v$  to their respective boundaries is encoded by a vector  $a_\omega \in \{0, 1\}^{N_v}$ , where for  $u \in N_v$  we have  $a_\omega(u) = 1$  when  $u$  is connected to  $\partial\mathcal{T}_u$  by a path in  $\mathcal{T}_u$ .

We start by proving the identity for  $Z_1(v)$ . In this case, we only consider configurations such that  $\|a_\omega\|_1 \geq 1$ . 1. For a fixed vector  $a_\omega$  such that  $\|a_\omega\|_1 = k$ , let  $u_1, \dots, u_k \in N_v$  be the neighbors of  $v$  for which  $a_\omega(u_i) = 1$ , and let  $\hat{u}_1, \dots, \hat{u}_l \in N_v$  be the neighbors of  $v$  for which  $a_\omega(\hat{u}_i) = 0$ ; hence  $l = |N_v| - k$ . Any random-cluster configuration  $\omega$  of  $\mathcal{T}_v$ , can be partitioned into the configuration on  $E(\{v\} \cup \bigcup_{u_i} \mathcal{T}_{u_i})$  and the configuration on  $E(\{v\} \cup \bigcup_{\hat{u}_i} \mathcal{T}_{\hat{u}_i})$ .

Given a vector  $a$ , let  $W_1(v, a, 1)$  denote the total weight under the wired boundary condition of the random-cluster configurations on  $E(\{v\} \cup \bigcup_{u_i} \mathcal{T}_{u_i})$  that contain a  $v$  to  $\partial \mathcal{T}_v$  connection and a  $u_i$  to  $\partial \mathcal{T}_{u_i}$  path in  $\mathcal{T}_{u_i}$  for every  $i \in \{1, \dots, k\}$ . Similarly, let  $W_1(v, a, 0)$  denote the total weight of the configurations on  $E(v \cup \bigcup_{\hat{u}_i} \mathcal{T}_{\hat{u}_i})$  in which there is no path between  $\hat{u}_i$  and  $\partial \mathcal{T}_{\hat{u}_i}$  in  $\mathcal{T}_{\hat{u}_i}$  for  $i \in \{1, \dots, l\}$ . Since conditioning on a disconnected configuration on  $E(\{v\} \cup \bigcup_{\hat{u}_i} \mathcal{T}_{\hat{u}_i})$  has no effect on the weight of the configuration on  $E(\{v\} \cup \bigcup_{u_i} \mathcal{T}_{u_i})$ , we have the identity

$$Z_1(v) = \frac{1}{q} \sum_{a \in \{0,1\}^{N_v} : \|a\|_1 \geq 1} W_1(v, a, 1) W_1(v, a, 0). \quad (3.1)$$

Here, the  $1/q$  factor comes from merging the two wired boundary components when  $\|a\|_1 < |N_v|$ ; if  $\|a\|_1 = |N_v|$ , we set  $W_1(v, a, 0) = q$ .

We compute  $W_1(v, a, 1)$  first. We use  $\Omega_1(\mathcal{T}_x)$  (resp.,  $\Omega_0(\mathcal{T}_x)$ ) for the set of all random-cluster configurations on the subtree  $\mathcal{T}_x$  in which there is (resp., there is not) an open path between  $x$  and  $\partial \mathcal{T}_h$  in  $\mathcal{T}_x$ . For a configuration  $\eta_i \in \Omega_0(\mathcal{T}_{u_i}) \cup \Omega_1(\mathcal{T}_{u_i})$ , we use  $\text{wgt}(\eta_i) = p^{|\eta_i|} (1-p)^{|E(\mathcal{T}_{u_i})| - |\eta_i|} q^{c_1(\eta_i)}$  for the weight of the random-cluster configuration on  $\mathcal{T}_{u_i}$  under the wired boundary condition; i.e.,  $c_1(\eta_i)$  corresponds to the number of connected components on  $\eta_i$  taking into consideration the wired boundary condition. Then, accounting also for the configuration in the edges between  $v$  and the  $u_i$ 's, we have

$$W_1(v, a, 1) = \sum_{\eta_1 \in \Omega_1(\mathcal{T}_{u_1})} \cdots \sum_{\eta_k \in \Omega_1(\mathcal{T}_{u_k})} \left( \prod_{i=1}^k \text{wgt}(\eta_i) \right) \frac{1}{q^{k-1}} \left( \sum_{i=1}^k \binom{k}{i} p^i (1-p)^{k-i} \right) \quad (3.2)$$

$$= \frac{1 - (1-p)^k}{q^{k-1}} \prod_{i=1}^k Z_1(u_i). \quad (3.3)$$

The re-scaling in (3.2) by  $\frac{1}{q^{k-1}}$  comes from the fact that the  $k$  boundary components in each subtree are all merged into a single component. By similar reasoning, when  $\|a\|_1 < |N_v|$

$$W_1(v, a, 0) = \sum_{\eta_1 \in \Omega_0(\mathcal{T}_{\hat{u}_1})} \cdots \sum_{\eta_l \in \Omega_0(\mathcal{T}_{\hat{u}_l})} \left( \prod_{i=1}^l \text{wgt}(\eta_i) \right) \frac{1}{q^{l-1}} \left( \sum_{i=0}^l \binom{l}{i} \left( \frac{p}{q} \right)^i (1-p)^{l-i} \right) \quad (3.4)$$

$$= \frac{(1-p + p/q)^l}{q^{l-1}} \prod_{i=1}^l Z_0(u_i). \quad (3.5)$$

Note that in (3.4), in addition to the re-scaling by  $\frac{1}{q^{l-1}}$  from merging the boundary components, any edge between  $v$  and one of its children decreases the number of components by 1; hence the  $q^{-i}$  in the term  $\left(\frac{p}{q}\right)^i$ .

Recall that  $t = 1 - p + p/q$ . Plugging (3.3) and (3.5) into (3.1) we obtain

$$Z_1(v) = q \sum_{a \in \{0,1\}^{N_v} : \|a\|_1 \geq 1} (1 - (1-p)^{\|a\|_1}) \prod_{w \in N_v : a(w)=1} \frac{Z_1(w)}{q} \prod_{w \in N_v : a(w)=0} \frac{tZ_0(w)}{q}.$$

Observe next that

$$\begin{aligned} \sum_{a \in \{0,1\}^{N_v} : \|a\|_1 \geq 1} \prod_{w \in N_v : a(w)=1} \frac{Z_1(w)}{q} \prod_{w \in N_v : a(w)=0} \frac{t \cdot Z_0(w)}{q} \\ = \prod_{w \in N_v} \left( \frac{Z_1(w)}{q} + \frac{t \cdot Z_0(w)}{q} \right) - \prod_{w \in N_v} \frac{t \cdot Z_0(w)}{q}, \end{aligned}$$

and

$$\begin{aligned} \sum_{a \in \{0,1\}^{N_v} : \|a\|_1 \geq 1} (1-p)^{\|a\|_1} \prod_{w \in N_v : a(w)=1} \frac{Z_1(w)}{q} \prod_{w \in N_v : a(w)=0} \frac{t \cdot Z_0(w)}{q} \\ = \prod_{w \in N_v} \left( \frac{(1-p)Z_1(w)}{q} + \frac{t \cdot Z_0(w)}{q} \right) - \prod_{w \in N_v} \frac{t \cdot Z_0(w)}{q}. \end{aligned}$$

Hence,

$$Z_1(v) = q \prod_{w \in N_v} \left( \frac{Z_1(w)}{q} + \frac{t \cdot Z_0(w)}{q} \right) - q \prod_{w \in N_v} \left( \frac{(1-p)Z_1(w)}{q} + \frac{t \cdot Z_0(w)}{q} \right),$$

as claimed. The expression for  $Z_0(v)$  can be derived from an analogous and slightly simpler argument and is thus omitted.  $\square$

**Proof of Fact 3.3.** We first consider the interval  $x \in [1, 1 + \eta]$  for some  $\eta > 0$  small. It can be checked that

$$\begin{aligned} g'(x) &= \frac{p(p+q-pq)}{(-1+q+x-p(-2+q+x))^2}, \\ g''(x) &= \frac{-2p(1-p)(q+(1-p)p)}{((1-p)x+p+(1-p)(q-1))^3}. \end{aligned}$$

Hence,  $g'(1) = \hat{p}$  and  $|g''|$  is decreasing for  $x \geq 1$ . Then, from the Taylor expansion of  $g$  at 1, we get

$$g(x) \leq 1 + \hat{p}(x-1) + c\eta^2, \quad (3.6)$$

where  $c = c(p, q) > 0$  is suitable constant. Similarly, using the Taylor expansion of  $x^{1/\gamma-\xi}$  at 1, we obtain

$$x^{1/\gamma-\xi} \geq 1 + (1/\gamma - \xi)(x-1) - c'\eta^2$$

for a suitable constant  $c' = c'(\gamma, \xi) > 0$ . Since  $\hat{p} < 1/\gamma$  when  $p < p_u(q, \gamma)$ , then for sufficiently small  $\xi$  and  $\eta$  (depending on  $p, q, \gamma$ ) we have  $g(x) \leq x^{1/\gamma-\xi}$  as desired.

We next observe that since  $g(x) \leq \frac{1}{1-p}$ , we have  $g(x) \leq x^{1/\gamma-\xi}$  for all  $x \geq K$  for  $K$  sufficiently large (depending on  $p, q, \gamma$ ), importantly independent of  $\xi$  as long as  $\xi < 1/(2\gamma)$ , say.

It remains to consider the case when  $x \in (1 + \eta, K)$ . For this, let us give an auxiliary form of  $p_u(q, \gamma)$ :

$$p_u(q, \gamma) = \sup \left\{ p : \sup_{x > 1} \{g_p(x) - x^{1/\gamma}\} \leq 0 \right\}. \quad (3.7)$$

(where we have added the  $p$  subscript to  $g$  to emphasize the  $p$  dependence there). Let us first conclude the proof assuming the equality of (3.7). By direct computation, it can be checked that  $\frac{\partial g(x)}{\partial p} > 0$  whenever  $x > 1$ . Hence, fixing  $p' \in (p, p_u(q, \gamma))$  for every  $x > 1$  we have  $g_p(x) < g_{p'}(x)$ , and by continuity  $g_p(x) \leq g_{p'}(x) - \delta$  for a sufficiently small  $\delta > 0$ . By continuity, in fact there exists a uniform choice of  $\delta > 0$  such that

$$g_p(x) < g_{p'}(x) - \delta \quad \text{for all } x \in [1 + \eta, K].$$

At the same time, for  $\xi$  sufficiently small, depending on  $\delta, \gamma, K$ , we have

$$|x^{1/\gamma} - x^{1/\gamma-\xi}| \leq \delta \quad \text{for all } x \in [1 + \eta, K].$$

Combining these two, and using (3.7), we see that for all  $x \in [1 + \eta, K]$ ,

$$g_p(x) \leq g_{p'}(x) - x^{1/\gamma} + x^{1/\gamma-\xi} \leq x^{1/\gamma-\xi} \quad \text{for all } x \in [1 + \eta, K].$$

It remains to establish the equality (3.7). We first rewrite the definition of  $p_u(q, \gamma)$  from (2.1) as

$$p_u(q, \gamma) = \sup \left\{ p : \sup_{y>1} \left\{ p - 1 + \frac{1}{1 + h(y)} \right\} \leq 0 \right\}.$$

It therefore suffices to establish that

$$\sup_{y>1} \left\{ p - 1 + \frac{1}{1 + h(y)} \right\} \leq 0 \iff \sup_{x>1} \{ g(x) - x^{1/\gamma} \} \leq 0.$$

By substituting  $y = x^{1/\gamma}$ , and calculating, this reduces to showing that for every  $y > 1$ ,

$$\frac{-(1-p)y^{\gamma+1} + y^\gamma - [p + (q-1)(1-p)]y + (q-1)(1-p)}{y^\gamma - y + (y-1)(y^\gamma + q-1)} \leq 0$$

if and only if

$$\frac{-(1-p)y^{\gamma+1} + y^\gamma - [p + (q-1)(1-p)]y + (q-1)(1-p)}{(1-p)y^\gamma + p + (q-1)(1-p)} \leq 0.$$

This equivalence follows because the numerators are the same, and the denominators are both positive whenever  $\gamma > 1$ ,  $q \geq 1$  and  $y > 1$ .  $\square$

**3.2.1. Uniqueness in general trees.** As a consequence of the decay of the root-to-leaf connectivity we have established, it follows that there is a unique infinite wired random-cluster measure whenever  $p < p_u(q, \gamma)$  on infinite trees with average branching  $\gamma$ . The random-cluster measure on the infinite wired tree is defined using the Dobrushin-Lanford-Ruelle (DLR) formalism (see, e.g., [38, 40]); in particular, the wired boundary condition corresponds to counting all infinite connected components as one.

Let  $\mathcal{T}$  be an infinite tree, let  $D_h \subset V(\mathcal{T})$  denote the set of vertices at distance  $h$  from the root of  $\mathcal{T}$  and define the branching rate  $Br(\mathcal{T})$  per [50] as:

$$Br(\mathcal{T}) = \inf \left\{ \lambda > 0 : \inf_h |D_h| \lambda^{-h} = 0 \right\}.$$

Observe that if  $Br(\mathcal{T}) < \gamma$ , then  $|D_h| < \gamma^h$  for all sufficiently large  $h$ . We prove the following.

**Corollary 3.4.** *Fix  $q \geq 1$ ,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . Suppose  $\mathcal{T}$  is an infinite tree with  $Br(\mathcal{T}) < \gamma$ . Then, there is a unique infinite-volume random-cluster measure on  $\mathcal{T}$  under the wired boundary condition.*

*Proof.* Let  $\mathcal{T}_h$  denote the subtree of  $\mathcal{T}$  that includes all vertices at distance at most  $h$  from the root  $\rho$  of  $\mathcal{T}$ . Let  $\pi_{\mathcal{T}}^1 = \lim_{h \rightarrow \infty} \pi_{\mathcal{T}_h}^1$ . It was established in [40, Lemma 3.1] that the limiting measure  $\pi_{\mathcal{T}}^1$  is a random-cluster measure with parameters  $p$  and  $q$  and, moreover, that any other random-cluster measure on  $\mathcal{T}$  with the same parameters is stochastically dominated by  $\pi_{\mathcal{T}}^1$ . (We note that Lemma 3.1 from [40] is stated for the case when  $\mathcal{T}$  is a homogeneous tree, but the proof there does not use this assumption, and the result clearly extends to general trees.) Now, since  $Br(\mathcal{T}) < \gamma$  we have that  $|D_h| < \gamma^h$  for sufficiently large  $h$ , and so Lemma 3.1 implies that  $\pi_{\mathcal{T}}^1(\rho \leftrightarrow \infty) = \lim_{h \rightarrow \infty} \pi_{\mathcal{T}_h}^1(\rho \leftrightarrow \partial \mathcal{T}_h) = 0$ . This implies that the conditional probability that any edge  $e$  is present, given the configuration outside of  $e$ , is  $\hat{p}$  with probability 1 (see, e.g., the proof of Theorem 1.8 in [40]). Hence,  $\pi_{\mathcal{T}}^1$  corresponds to the i.i.d. distribution on  $\{0, 1\}^{E(\mathcal{T})}$  with edge probability  $\hat{p}$ . By the same argument, the same is true for any other random-cluster  $\mu$  since  $\mu \leq \pi_{\mathcal{T}}^1$ , and the result follows.  $\square$

**Corollary 3.5.** *Fix  $q \geq 2$  integer,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . Suppose  $\mathcal{T}$  is an infinite tree with  $Br(\gamma) < \gamma$ . Then there is a unique infinite-volume Potts measure on  $\mathcal{T}$ .*



**3.3. Galton–Watson trees: volume and uniqueness.** As corollaries of our results on general trees, we can obtain exponential decay and uniqueness results for the random-cluster model on a Galton–Watson random tree. Let  $\nu$  denote the progeny distribution for a Galton–Watson tree. For  $\ell \geq 0$  let  $Z_\ell$  be the number of vertices in  $\ell$ -th generation so that  $Z_0 = 1$  and  $Z_1 \sim \nu$ . Our first result provides a tail bound for  $Z_\ell$  (under mild moment assumptions on  $\nu$ ). This bound will allow us to argue that the Galton–Watson tree satisfies the volume assumptions of Lemma 2.7, and deduce uniqueness of the random-cluster measure on super-critical Galton–Watson trees when  $p < p_u(q, \gamma)$ .

**Lemma 3.6.** *Let  $N \sim \nu$ ,  $\kappa \geq 1$  and suppose  $\mathbb{E}[N] \geq 1$  and  $\mathbb{E}[N^\kappa] < \mathbf{m}_\kappa$  for some constant  $\mathbf{m}_\kappa$ . If  $\kappa$  is a power of 2, there exists  $C = C(\kappa, \mathbb{E}[N], \mathbf{m}_\kappa)$  such that for every  $\gamma > 0$  and every  $1 \leq \ell \leq h$ ,*

$$\mathbb{P}(Z_\ell \geq \gamma^h) \leq Ch^{2\kappa} \left( \frac{\mathbb{E}[N]}{\gamma} \right)^{\kappa h}.$$

*Proof.* Let  $\mathbb{E}[N] = \mathbf{m}$  and  $W_\ell = Z_\ell / \mathbf{m}^\ell$ . From the definition of the Galton–Watson tree we have that  $Z_{\ell+1} = \sum_{i=1}^{Z_\ell} N_{\ell,i}$  for all  $\ell \geq 1$ , where the  $N_{\ell,i}$  are independent copies of  $N \sim \nu$ . Then,

$$W_{\ell+1} - W_\ell = \frac{1}{\mathbf{m}^\ell} \sum_{i=1}^{Z_\ell} (W_{1,i} - 1), \quad (3.8)$$

where the  $W_{1,i}$ ’s are i.i.d.’s instances of  $W_1 = N/\mathbf{m}$ . From (3.8) we deduce that

$$\mathbb{E}[|W_{\ell+1} - W_\ell|^\kappa \mid Z_\ell = L] = \frac{1}{\mathbf{m}^{\kappa\ell}} \mathbb{E} \left[ \left| \sum_{i=1}^L (W_{1,i} - 1) \right|^\kappa \right].$$

Since the  $W_{1,i}$ ’s are i.i.d.’s with  $\mathbb{E}[W_{1,i} - 1] = 0$ , and  $\mathbb{E}[|W_{1,i} - 1|^\kappa] \leq 2^\kappa (\frac{1}{\mathbf{m}^\kappa} \mathbb{E}[N^\kappa] + 1)$  is finite, it follows from the Marcinkiewicz–Zygmund inequality that

$$\mathbb{E} \left[ \left| \sum_{i=1}^L (W_{1,i} - 1) \right|^\kappa \right] \leq B_\kappa \cdot \mathbb{E} \left[ \left( \sum_{i=1}^L (W_{1,i} - 1)^2 \right)^{\kappa/2} \right],$$

where  $B_\kappa = (2\lceil \kappa/2 \rceil)^{1/2})^\kappa$  (see Section 10.3 in [17]). For  $\kappa > 2$ , Jensen’s inequality then provides the bound

$$\mathbb{E} \left[ \left( \sum_{i=1}^L (W_{1,i} - 1)^2 \right)^{\kappa/2} \right] \leq L^{\kappa/2-1} \sum_{i=1}^L \mathbb{E}[|W_{1,i} - 1|^\kappa]. \quad (3.9)$$

Combining these inequalities and taking expectations we obtain

$$\mathbb{E}[|W_{\ell+1} - W_\ell|^\kappa] \leq \frac{B_\kappa}{\mathbf{m}^{\kappa\ell}} \mathbb{E}[Z_\ell^{\kappa/2}] \mathbb{E}[|W_1 - 1|^\kappa] = \frac{B_\kappa}{\mathbf{m}^{\kappa\ell/2}} \mathbb{E}[W_\ell^{\kappa/2}] \mathbb{E}[|W_1 - 1|^\kappa],$$

and since  $\mathbb{E}[|W_1 - 1|^\kappa] \leq 2^\kappa (\frac{1}{\mathbf{m}^\kappa} \mathbb{E}[N^\kappa] + 1)$ , for a suitable constant  $C = C(\mathbf{m}, \kappa, \mathbf{m}_\kappa)$  we have

$$\mathbb{E}[|W_{\ell+1} - W_\ell|^\kappa] \leq \frac{C}{\mathbf{m}^{\kappa\ell/2}} \mathbb{E}[W_\ell^{\kappa/2}]. \quad (3.10)$$

Now, let  $Y_1 = W_1$  and for  $i \geq 2$  let  $Y_i = W_i - W_{i-1}$ , so that  $W_\ell = \sum_{i=1}^\ell Y_i$ . Using the triangle and Jensen’s inequalities (as in (3.9)) we deduce that

$$\mathbb{E}[|W_\ell|^\kappa] \leq \mathbb{E} \left[ \left( \sum_{i=1}^\ell |Y_i| \right)^\kappa \right] \leq \ell^{\kappa-1} \sum_{i=1}^\ell \mathbb{E}[|Y_i|^\kappa],$$

and from the bound in (3.10) we get

$$\mathbb{E}[W_\ell^\kappa] \leq C \ell^{\kappa-1} \sum_{i=1}^\ell \frac{\mathbb{E}[W_i^{\kappa/2}]}{\mathbf{m}^{\kappa i/2}}.$$

From this, letting  $\rho(h, \kappa) = \max_{i \leq h} \mathbb{E}[W_i^\kappa]$ , we obtain the recurrence

$$\rho(h, \kappa) \leq C h^{\kappa-1} \rho(h, \kappa/2) \sum_{i=1}^h \frac{1}{\mathbf{m}^{\kappa i/2}} \leq C \cdot h^\kappa \cdot \rho(h, \kappa/2),$$

since  $\mathbf{m} \geq 1$  by assumption. Since  $\kappa$  is a power of two, iteratively, for a suitable constant  $C_1 = C_1(\kappa, \mathbf{m}, \mathbf{m}_\kappa)$ ,

$$\rho(h, \kappa) \leq C_1 \cdot \rho(h, 1) \cdot \prod_{i=0}^{\log_2 \kappa} h^{\kappa/2^i} \leq C_1 \cdot h^{2\kappa}. \quad (3.11)$$

Finally, we note that by Markov's inequality, for any  $\ell \leq h$

$$\mathbb{P}(Z_\ell \geq \gamma^h) = \mathbb{P}\left(W_\ell^\kappa \geq \left(\frac{\gamma}{\mathbf{m}}\right)^{h\kappa}\right) \leq C_1 h^{2\kappa} \left(\frac{\mathbf{m}}{\gamma}\right)^{\kappa h},$$

as claimed.  $\square$

We show next that Galton-Watson trees satisfy (with high probability) a certain growth condition that would allow us to apply the sharp decay of connectivities in random-cluster configurations from Lemma 2.7. We define the following volume growth condition for the random tree, which is stronger than the assumption of Lemma 2.7, and will also reappear later in the paper.

**Definition 3.7.** We say a tree  $\mathcal{T}_h = (V(\mathcal{T}_h), E(\mathcal{T}_h))$  satisfies the  $(\gamma, \varepsilon)$ -tree-growth condition if for every  $v \in V(\mathcal{T}_h)$  with  $h(v) > \varepsilon h$ , and every  $k$  such that  $\varepsilon h < k \leq h(v)$ , we have  $|V(\mathcal{T}_v(k))| \leq \gamma^k$ , where  $\mathcal{T}_v(k)$  denotes the subtree of  $\mathcal{T}_h$  of height  $k$  rooted at  $v$ .

**Corollary 3.8.** Let  $N \sim \nu$  and  $\kappa \geq 1$ . Suppose  $1 \leq \mathbb{E}[N] < \gamma$  and that there exists a constant  $\mathbf{m}_\kappa$  such that  $\mathbb{E}[N^\kappa] < \mathbf{m}_\kappa$ . Then, if  $\kappa$  is a sufficiently large power of 2, there exists a constant  $\theta = \theta(\gamma, \kappa, \mathbb{E}[N], \mathbf{m}_\kappa) \in (0, 1)$  such that the Galton-Watson tree truncated at height  $h$  with progeny distribution  $\nu$  has  $(\gamma, \varepsilon)$ -tree-growth with probability at least  $1 - \theta^{\varepsilon \kappa h}$  for  $h$  sufficiently large.

*Proof.* Let  $\{X_k^{(j)}\}_{j \geq 1}$  be i.i.d. random variables corresponding to the total number of vertices in a Galton-Watson tree of height  $k$ . By a union bound, the probability that the Galton-Watson tree does not satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition is at most:

$$\begin{aligned} & \sum_{l=0}^{(1-\varepsilon)h} \mathbb{P}\left(\bigcup_{j=1}^{Z_l} \bigcup_{\varepsilon h \leq k \leq h-l} \{X_k^{(j)} \geq \gamma^k\}\right) \\ & \leq (2\gamma)^h \sum_{l=0}^{(1-\varepsilon)h} \sum_{\varepsilon h \leq k \leq h-l} \mathbb{P}\left(X_k^{(1)} \geq \gamma^k\right) + \sum_{l=0}^{(1-\varepsilon)h} \mathbb{P}\left(Z_l \geq (2\gamma)^h\right). \end{aligned} \quad (3.12)$$

From Lemma 3.6, we know that there exists a constant  $C = C(\kappa, \mathbb{E}[N], \mathbf{m}_\kappa)$  such that

$$\sum_{l=0}^{(1-\varepsilon)h} \mathbb{P}\left(Z_l \geq (2\gamma)^h\right) \leq \frac{C h^{2\kappa+1}}{2^{\kappa h}}. \quad (3.13)$$

Now, observe that  $X_k^{(1)}$  has the same distribution as  $\sum_{j=0}^k Z_j$ . Hence, Lemma 3.6 and a union bound imply that there exists  $\hat{\gamma} \in (\mathbb{E}[N], \gamma)$  such that

$$\mathbb{P}\left(X_k^{(1)} \geq \gamma^k\right) = \mathbb{P}\left(\sum_{j=0}^k Z_j \geq \gamma^k\right) \leq \sum_{j=0}^k \mathbb{P}(Z_j \geq \hat{\gamma}^k) \leq C_1 k^{2r+1} \left(\frac{\mathbb{E}[N]}{\hat{\gamma}}\right)^{\kappa k}$$

for a suitable constant  $C_1 = C_1(\kappa, \mathbb{E}[N], \mathbf{m}_\kappa) > 0$  and  $k$  large enough. Then,

$$\begin{aligned} \sum_{l=0}^{(1-\varepsilon)h} \sum_{\varepsilon h \leq k \leq h-l} \mathbb{P}\left(X_k^{(1)} \geq \gamma^k\right) &\leq C_1 \sum_{l=0}^{(1-\varepsilon)h} \sum_{\varepsilon h \leq k \leq h-l} k^{2r+1} \left(\frac{\mathbb{E}[N]}{\hat{\gamma}}\right)^{rk} \\ &\leq C_2 h^{2\kappa+2} \left(\frac{\mathbb{E}[N]}{\hat{\gamma}}\right)^{\varepsilon\kappa h}, \end{aligned}$$

for a suitable constant  $C_2 > 0$ . Plugging this bound and (3.13) into (3.12), we obtain that the probability that the Galton–Watson tree does not satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition is at most  $1 - \theta^{\varepsilon\kappa h}$  for a suitable  $\theta = \theta(\gamma, \kappa, \mathbb{E}[N], \mathbf{m}_\kappa) \in (0, 1)$  as claimed.  $\square$

**3.3.1. Uniqueness in Galton–Watson trees.** Let  $\mathcal{T}$  be a Galton–Watson tree with progeny distribution  $\nu$  and let  $N \sim \nu$ . By Lemma 3.6 and the Borel–Cantelli lemma, with probability one over  $\mathcal{T}$ , for any  $\gamma > \mathbb{E}[N]$ , we have  $Z_h \leq \gamma^h$  for all sufficiently large  $h$ . In particular, with probability one,  $Br(\mathcal{T}) < \gamma$  for any  $\gamma > \mathbb{E}[N]$ . As such, Corollary 3.4 implies that there is a unique random-cluster measure on  $\mathcal{T}$  under the wired boundary condition when  $p < p_u(q, \gamma)$ .

**Corollary 3.9.** *Fix  $q \geq 1$ ,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . Let  $N \sim \nu$ ,  $\kappa \geq 1$  and suppose  $1 \leq \mathbb{E}[N] < \gamma$  and that there exist a constant  $\mathbf{m}_\kappa$  such that  $\mathbb{E}[N^\kappa] < \mathbf{m}_\kappa$ . With probability one over  $\mathcal{T}$ , there is a unique random-cluster distribution on  $\mathcal{T}$  under the wired boundary condition. Similarly, at integer  $q$ , with probability one over  $\mathcal{T}$ , there is a unique Potts distribution on  $\mathcal{T}$ .*

#### 4. RANDOM-GRAPH ESTIMATES

In this section, we describe the standard revealing scheme for the configuration model with degree sequence  $\mathbf{d}_n$ . We also formalize the mechanism to translate probability  $1 - o(1)$  events for  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  to  $1 - o(1)$  events for  $\mathbb{P}_{\text{RG}(\mathbf{d}_n)}$  and for the Erdős–Rényi random graph model; we use this to provide a proof of Theorem 1.1 given Theorem 1.6. We then use the revealing scheme for the configuration model to prove the random graph estimates of Lemmas 2.2 and 2.4.

**4.1. Configuration model with general degree sequence.** We begin by describing a revealing procedure for the configuration model with degree sequence  $\mathbf{d}_n$ . To do so, we begin with an important definition providing the *state space* for our revealing procedures of the configuration model. Recall that a *matching* on a graph is an edge-subset such that no vertex belongs to more than one edge. A *perfect matching* is an edge-subset in which every vertex belongs to exactly one edge.

**Definition 4.1.** Given a degree sequence  $\mathbf{d}_n = (d_v)_{1 \leq v \leq n}$ , to each vertex  $v \in \{1, \dots, n\}$ , assign  $d_v$  *half-edges*. Consider an auxiliary complete graph  $K_{\|\mathbf{d}_n\|_1}$  whose  $\|\mathbf{d}_n\|_1$  vertices are identified with these half-edges. Let  $\mathfrak{M}_{\mathbf{d}_n}$  be the set of all matchings (not necessarily perfect) on  $K_{\|\mathbf{d}_n\|_1}$ , and let  $\overline{\mathfrak{M}}_n(\mathbf{d}_n)$  be the set of all perfect matchings on  $K_{\|\mathbf{d}_n\|_1}$ .

We are now in position to formally define the configuration model of random graphs.

**Definition 4.2.** Given a degree sequence  $\mathbf{d}_n$ , the configuration model  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is the uniform distribution over  $\overline{\mathfrak{M}}_n(\mathbf{d}_n)$ , i.e., it is a uniform perfect matching of the  $\|\mathbf{d}_n\|_1$  half-edges assigned to the vertices  $\{1, \dots, n\}$ . This is naturally identified with a multigraph on  $\{1, \dots, n\}$  by identifying all half-edges with the vertex they come from, so that the edges in the matching become edges of the graph between the corresponding vertices. In this manner, with a slight abuse of notation, elements  $E \in \overline{\mathfrak{M}}_n(\mathbf{d}_n)$  are simply the edge-sets of the multigraph  $\mathcal{G} = (V, E)$ .

**Remark 4.3.** The definitions of the random-cluster model (1.2), and the FK-dynamics extend naturally to multigraphs, where  $G = (V, E)$  is such that  $V$  is identified with  $\{1, \dots, n\}$  and  $E \in \overline{\mathfrak{M}}_n$  is a multiset. The random-cluster model and FK-dynamics then live over subsets of  $E$ , identified with  $\omega : E \rightarrow \{0, 1\}$ , and connected components of a configuration  $\omega$  are understood naturally.

**4.2. Revealing procedure for the configuration model.** We now describe a simple revealing procedure for generating a sample from the configuration model distribution given fixed degree sequence  $\mathbf{d}_n$ .

**Process 4.4.** Fix a degree sequence  $\mathbf{d}_n$  with  $\sum_v d_v$  even. Suppose  $f$  is a (possibly random) function from matchings  $A \in \mathfrak{M}_n$ , to a half-edge not matched in  $A$ .

- (1) Initialize the set  $A_0 = \emptyset$
- (2) For every  $t \geq 0$ , if  $A_t \notin \overline{\mathfrak{M}}_n$  (i.e., there exist un-matched half-edges), construct  $A_{t+1}$  as follows:
  - (a) Let  $\hat{e}_{t+1}$  be the half-edge  $f(A_t)$
  - (b) Pick another un-matched half-edge in  $A_t$  uniformly at random, and match it with  $\hat{e}_{t+1}$  in  $A_{t+1}$ .

For natural choices of the function  $f$ , we can reveal, for example, a ball in the random graph without revealing any information about the remainder of the random graph. The next definitions give an example of such an  $f$  that we will use repeatedly.

**Definition 4.5.** Given a matching  $A \in \mathfrak{M}_n$ , the set of *exposed* half-edges of  $A$  is the set of un-matched half-edges that belong to the same vertex (among  $V = \{1, \dots, n\}$ ) as some half-edge that is matched in  $A$ . Denote this set by  $\hat{E}(A)$ .

**Process 4.6.** The breadth-first exploration of a ball  $B_r(v) \subset E(\mathcal{G})$  is constructed using Process 4.4 with the following choice of  $f$ . For each  $A$ ,  $f(A)$  is an arbitrarily chosen exposed half-edge among  $\hat{E}(A)$  whose distance in  $(V, A)$  to  $v$  is at most  $r$ .

**4.3. Contiguity with simple random graphs.** The configuration model described above gives a uniform at random *multigraph* with prescribed degree sequence  $\mathbf{d}_n$ . In the sparse regime of bounded average degree, this happens to be a very useful model for studying random *simple* graphs (i.e., has no self-loops or multi-edges), most notably  $\Delta$ -regular random graphs, but also a uniformly chosen random simple graph with degree sequence  $\mathbf{d}_n$  (as long as the sequence is graphical).

**4.3.1. General degree sequence.** It is well established that in the sparse regime of bounded average degree, the configuration model will have probability uniformly bounded away from zero of being simple, and on that event it is exactly a uniform simple graph with degree sequence  $\mathbf{d}_n$ . This contiguity can be summarized as follows (see e.g., [45]).

**Lemma 4.7.** Fix any  $\gamma$  and  $\kappa$ . Suppose  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  and  $\|\mathbf{d}_n\|_1 = \Omega(n)$ . Then for any sequence of sets  $A_n$  of simple graphs on  $n$  vertices, we have

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \in A) = o(1) \quad \text{if and only if} \quad \mathbb{P}_{\text{RG}(\mathbf{d}_n)}(\mathcal{G} \in A) = o(1).$$

**4.3.2. Erdős–Rényi random graph.** In the case of the Erdős–Rényi random graph  $G(n, d/n)$ , the degree of a vertex  $v$  is not fixed, but rather is distributed as  $\text{Bin}(n-1, d/n)$ . Nonetheless, there is a way to first randomly sample  $\mathbf{d}_n$  then draw a configuration model on  $\mathbf{d}_n$ , such that the resulting random graph is contiguous to the Erdős–Rényi distribution. Let  $\mathbb{P}_{\text{Poi}(d)}$  be the distribution over  $\mathbf{d}_n = (d_1, \dots, d_n)$  where  $d_i$  are i.i.d. Poisson( $d$ ) random variables. The following was established in [47].

**Lemma 4.8.** For any  $d = \Theta(1)$ , for every sequence of sets  $A_n$  of simple graphs on  $n$  vertices, we have

$$\mathbb{E}_{\text{Poi}(d)}[\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \in A)] = o(1) \quad \text{if and only if} \quad \mathbb{P}_{G(n, d/n)}(\mathcal{G} \in A) = o(1).$$

In the above lemma, on the event that  $\mathbf{d}_n$  does not have  $\|\mathbf{d}_n\|_1$  even, as a matter of convention, we take the probability in the expectation to be zero. Overloading notation slightly, let  $\mathbb{P}_{\text{Poi}(d)}$  be the product distribution over  $\mathbf{d}_n \sim \mathbb{P}_{\text{Poi}(d)}$  for each  $n$ .

**Lemma 4.9.** For every  $0 < d < \gamma$  and every  $\kappa \geq 1$ ,

$$\mathbb{P}_{\text{Poi}(d)}((\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}) = 1.$$

*Proof.* Recall by definition of  $\mathbb{P}_{\mathbf{d}_n}, \mathbb{E}_{\mathbf{d}_n}$ , that

$$\mathbb{E}_{\mathbf{d}_n}[D] = \frac{1}{\|\mathbf{d}_n\|_1} \sum_v d_v(d_v - 1) = \frac{\sum_v d_v^2}{\|\mathbf{d}_n\|_1} - 1.$$

Let  $0 < \varepsilon < \gamma - d$ . Then for every  $n$  large, we have

$$\begin{aligned} \mathbb{P}_{\text{Poi}(d)}(\mathbb{E}_{\mathbf{d}_n}[D] < \gamma - \varepsilon) &\leq \mathbb{P}_{\text{Poi}(d)}\left(\frac{1}{\|\mathbf{d}_n\|_1} \sum_v d_v^2 \geq \gamma - \varepsilon + 1\right) \\ &\leq \mathbb{P}\left(\frac{1}{n} \sum_v d_v^2 > d(d+1) + n^{-\frac{1}{2}+\delta}\right) + \mathbb{P}\left(\frac{1}{n} \sum_v d_v < d - n^{-\frac{1}{2}+\delta}\right). \end{aligned} \quad (4.1)$$

To bound either of these terms, notice by Markov's inequality, that

$$\mathbb{P}\left(\left|\sum_v d_v^k - \mathbb{E}[\sum_v d_v^k]\right| > \lambda\right) \leq \frac{\mathbb{E}[|\sum_v (d_v^k - \mathbb{E}[d_v^k])|^l]}{\lambda^l}.$$

The numerator on the right-hand side is a sum of i.i.d. mean-zero random variables, each of which have all finite moments. As such, for any fixed  $l$ , the right-hand side above is at most

$$Cn^{l/2}\lambda^{-l} \leq Cn^{-l\delta}.$$

Taking  $l > 5\delta^{-1}$ , the right-hand side above is  $O(n^{-5})$ . Therefore, the sum over  $n$  of the probabilities of the left-hand side of (4.1) is finite, and by Borel–Cantelli, with probability one, eventually almost surely,  $\mathbb{E}_{\mathbf{d}_n}[D] < \gamma - \varepsilon$ , so that  $\limsup \mathbb{E}_{\mathbf{d}_n}[D] < \gamma$ . A similar argument yields the uniform boundedness of the  $\kappa$ 'th moments  $\mathbb{E}_{\mathbf{d}_n}[D^\kappa]$  for any  $\kappa$ , yielding the desired and concluding the proof.  $\square$

Given Lemmas 4.8–4.9, our Theorem 1.1 becomes a corollary of Theorem 1.6.

**Proof of Theorem 1.1 given Theorem 1.6.** Fix  $q \geq 1, \gamma > 0$  and  $p < p_u(q, \gamma)$ . Suppose  $\mathcal{G} \sim G(n, \gamma/n)$ . Fix a large constant  $K$  and let  $A$  be the set of simple graphs  $\mathcal{G}$  such that the mixing time of FK-dynamics on  $\mathcal{G}$  at parameters  $p, q$  satisfies  $K^{-1} \log n \leq t_{\text{MIX}} \leq K \log n$ . By Lemma 4.8, it suffices to show that

$$\mathbb{E}_{\text{Poi}(\gamma)}[\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \notin A)] = o(1).$$

Considering this quantity, for any  $\gamma'$ ,

$$\begin{aligned} \limsup_n \mathbb{E}_{\text{Poi}(\gamma)}[\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \notin A)] \\ \leq \mathbb{P}_{\text{Poi}(\gamma)}((\mathbf{d}_n)_n \notin \mathcal{D}_{\gamma', \kappa}) + \sup_{(\mathbf{d}_n) \in \mathcal{D}_{\gamma', \kappa}} \limsup_n \mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \notin A). \end{aligned}$$

The first term on the right-hand side is zero for all  $\gamma' > \gamma$  and all  $\kappa$  by Lemma 4.9. By Theorem 1.6 and Lemma 4.7, the second term is zero if  $\gamma' > \gamma$  is such that  $p < p_u(q, \gamma')$ , and if  $\kappa$  and  $K$  are sufficiently large (depending on  $p, q, \gamma'$ ). By continuity of  $p_u(q, \gamma)$ , if  $p < p_u(q, \gamma)$ , there also exists  $\gamma' > \gamma$  such that  $p < p_u(q, \gamma')$ , concluding the proof.  $\square$

**4.4. Local domination of the configuration model by random trees.** We now dominate balls of volume  $o(n^{1/2})$  of the random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  by branching processes whose progeny are approximately given by  $\mathbb{P}_{\mathbf{d}_n}$ . To be more precise, we define the following.

**Definition 4.10.** Define the *truncated empirical distribution* by letting  $\underline{\mathbf{d}}_n = \mathbf{d}_n \setminus A_{\mathbf{d}_n}$ , where  $A_{\mathbf{d}_n}$  are the smallest  $2\sqrt{n}$  elements of  $\mathbf{d}_n$ , and the set subtraction is done in the multi-set sense. Then let  $\mathbb{P}_{\underline{\mathbf{d}}_n}$  be the corresponding effective offspring distribution of  $\underline{\mathbf{d}}_n$ , i.e., for  $x \in \{d_v - 1 : d_v \in \underline{\mathbf{d}}_n\}$ ,

$$\mathbb{P}_{\underline{\mathbf{d}}_n}(x) = \frac{\sum_{v: d_v \in \underline{\mathbf{d}}_n} (x+1) \mathbf{1}_{\{d_v = x+1\}}}{\|\underline{\mathbf{d}}_n\|_1}.$$

Let  $\underline{D} \sim \mathbb{P}_{\underline{\mathbf{d}}_n}$ , and let  $\mathbb{E}_{\underline{\mathbf{d}}_n}$  be the corresponding expectation.



**Lemma 4.11.** *If  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then  $(\underline{\mathbf{d}}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ .*

*Proof.* Let  $A_{\mathbf{d}_n}$  be the set of  $2n^{1/2}$  smallest degrees of  $\mathbf{d}_n$ . We first of all claim that  $\|\mathbf{d}_n\|_1 \leq (1 + o(1))\|\underline{\mathbf{d}}_n\|_1$ . Indeed this follows from the calculation

$$\frac{\|\mathbf{d}_n\|_1 - \|\underline{\mathbf{d}}_n\|_1}{\|\underline{\mathbf{d}}_n\|_1} \leq \frac{\sum_{v: d_v \in A_{\mathbf{d}_n}} d_v}{\sum_{v: d_v \notin A_{\mathbf{d}_n}} d_v} \leq \frac{2\sqrt{n} \max\{d_v : d_v \in A_{\mathbf{d}_n}\}}{(n - 2\sqrt{n}) \min\{d_v : d_v \notin A_{\mathbf{d}_n}\}} \leq O(n^{-1/2}).$$

We then can observe that

$$\mathbb{E}_{\underline{\mathbf{d}}_n}[D] = \frac{\|\mathbf{d}_n\|_1}{\|\underline{\mathbf{d}}_n\|_1} \frac{1}{\|\mathbf{d}_n\|_1} \sum_{v: d_v \notin A_{\mathbf{d}_n}} d_v(d_v - 1) \leq (1 + o(1))\mathbb{E}_{\mathbf{d}_n}[D].$$

We now wish to prove the desired moment conditions. Those follow by analogous reasoning:

$$\mathbb{E}_{\underline{\mathbf{d}}_n}[D^k] \leq \frac{\|\mathbf{d}_n\|_1}{\|\underline{\mathbf{d}}_n\|_1} \frac{1}{\|\mathbf{d}_n\|_1} \sum_{v: d_v \notin A_{\mathbf{d}_n}} d_v(d_v - 1)^k \leq (1 + o(1))\mathbb{E}_{\mathbf{d}_n}[D^k].$$

Altogether, these give the desired implications of the lemma.  $\square$

We now wish to show that the balls of the random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}}(\mathbf{d}_n)$  are stochastically dominated by random trees with offspring distribution  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ , even conditionally on an already revealed portion  $H \in \mathfrak{M}_n$  of the random graph. However, this will only hold if  $|H| \leq n^{1/2}$  and the ball does not intersect  $H$ . We now formalize this notion.

**Process 4.12.** For a subgraph  $H = (V(H), E(H))$ , let  $\hat{E}(H)$  be the set of half-edges incident to  $H$  but not matched in  $H$ . (Notice that this definition aligns with the use of  $\hat{E}(A)$  for the exposed half-edges of  $A \in \mathfrak{M}_n$  when taking  $H = (V(A), A)$ .) For a half-edge  $\hat{e}$  in  $\hat{E}(H)$ , define  $B_r(\hat{e}; H^c)$  as the ball of radius  $r$  “out of  $H$ ”. More formally,  $B_r(\hat{e}; H^c)$  is obtained by

- (1) Matching  $\hat{e}$  to a vertex  $w$ .
- (2) Running the breadth-first revealing of  $B_{r-1}(w)$  from Process 4.6 but where  $f(A)$  cannot be in  $\hat{E}(H)$  (i.e., it will be an arbitrarily chosen half-edge of  $\hat{E}(A) \setminus \hat{E}(H)$  at distance at most  $r$  from  $w$  in  $A$ ).

Due to the extra edge from matching  $\hat{e}$ , let us say a *single-source* Galton–Watson tree is a Galton–Watson tree whose first generation deterministically has exactly one child.

**Proposition 4.13.** *Consider any degree sequence  $\mathbf{d}_n$ . Let  $\underline{\mathbf{d}}_n$  be as per Definition 4.10. Let  $\hat{\mathcal{T}}_h(\underline{\mathbf{d}}_n)$  be a single-source Galton–Watson tree of at depth  $h$  (meaning it is truncated at depth  $h$ ) with offspring distribution  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ . Fix an arbitrary  $H = (V(H), E(H)) \in \mathfrak{M}_n$ , and consider  $\hat{e} \in \hat{E}(H)$ . Then, conditionally on  $\{E(H) \subset E(\mathcal{G})\}$ , we have the stochastic domination*

$$|B_r(\hat{e}; H^c)| \mathbf{1}_{\{|E(H) \cup E(B_r(\hat{e}; H^c))| \leq n^{1/2}\}} \leq |\hat{\mathcal{T}}_r(\underline{\mathbf{d}}_n)|,$$

*On the event that  $B_r(\hat{e}; H^c)$  is a tree, there is an isometry between the graphs such that  $B_r(\hat{e}; H^c)$  is a subset of  $\hat{\mathcal{T}}_r(\underline{\mathbf{d}}_n)$ .*

*Proof.* We appeal to the revealing procedure of Process 4.4 with the choice of breadth-first revealing described in Processes 4.6 and 4.12. Begin the single-source Galton–Watson tree with a root vertex and a single child, corresponding to  $\hat{e}$ . Iteratively, when a half-edge  $\hat{f}$ , corresponding to a vertex  $x$  in the single-source Galton–Watson tree, gets matched in the revealing procedure to a vertex  $w$ ,

- (1) If  $w$  had not been exposed yet, identify the other  $d_w - 1$  half-edges of  $w$  with the children of  $x$  in the single-source Galton–Watson tree.
- (2) If  $w$  is an exposed vertex, do nothing.

(We say a vertex is exposed if one of its half-edges has already been matched, whether in  $H$  or in the revealing.) Uniformly over any subset of at most  $2n^{1/2}$  matched half-edges (forming  $n^{1/2}$  edges), the distribution  $d_w - 1$  is easily seen to be stochastically below  $\mathbb{P}_{\underline{\mathbf{d}}_n}$  (in which the *smallest*  $2n^{1/2}$  half-edges have been removed). Notice then that on the indicator

$$\mathbf{1}_{\{|E(H) \cup E(B_r(\hat{e}; H^c))| \leq n^{1/2}\}},$$

throughout the breadth-first revealing process, the number of matched half-edges will always be at most  $2n^{1/2}$ . Thus, we see that this process maintains the desired stochastic domination relation as compared to the single-source Galton–Watson tree until the number of matched half-edges exceeds  $n^{1/2}$ .

When  $B_r(\hat{e}; H^c)$  is a tree, item (2) above never happens, and the isometry goes by identifying the edge containing  $\hat{f}$  in  $E(\mathcal{G})$  with the edge connecting the corresponding vertex in  $\hat{\mathcal{T}}_r(\underline{\mathbf{d}}_n)$  to its parent.  $\square$

With Proposition 4.13, we can translate the volume growth bounds of Lemma 3.6 into the desired volume growth estimate for the random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ . In this proof, and other proofs relying on the random graph revealing procedure, it will be useful to have an  $\ell^\infty$  bound on the degrees. For this, note that  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  implicitly places a constraint on  $\|\mathbf{d}_n\|_\infty$ , since  $\mathbb{P}_{\mathbf{d}_n}$  chooses  $\|\mathbf{d}_n\|_\infty - 1$  with probability  $\Omega(\|\mathbf{d}_n\|_\infty/n)$ . More precisely, we have the following.

**Fact 4.14.** *If  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*}$  for  $\varepsilon_*(\kappa) = 2/(\kappa + 1)$ .*

**Proof of Lemma 2.4.** We will take a union bound over the probabilities that for a fixed vertex  $v \in \{1, \dots, n\}$ , and a fixed  $r \geq \varepsilon \log_\gamma n$ , the graph  $\mathcal{G}$  has  $|B_r(v)| \leq C\gamma^r$ . Fix any such  $v, r$  and take  $H = (\{v\}, \emptyset)$ , so that  $\hat{E}(H)$  are exactly the half-edges of  $v$ . Evidently, for  $\hat{e} \in \hat{E}(H)$

$$\mathbb{P}(|B_r(v)| \geq \gamma^r) \leq d_v \mathbb{P}(|B_{r-1}(\hat{e})| \geq d_v^{-1} \gamma^r).$$

Consider the probability on the right. For each  $\hat{e} \in \hat{E}(H)$ , by Proposition 4.13,

$$|B_{r-1}(\hat{e})| \mathbf{1}_{\{|B_{r-1}(\hat{e})| \leq n^{1/2}\}} \leq Z_r,$$

where  $Z_r \sim |\hat{\mathcal{T}}_{r-1}(\underline{\mathbf{d}}_n)|$ , where we recall this is the single-source Galton–Watson tree of depth  $r - 1$  whose offspring distribution is  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ . Now using a union bound,

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\mathcal{G} \text{ does not have } (\gamma, \varepsilon)\text{-volume growth}) \leq \sum_v d_v \sum_{r=\varepsilon \log_\gamma n}^{\frac{1}{2} \log_\gamma n} \mathbb{P}(|Z_r| \geq d_v^{-1} \gamma^r).$$

Since  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , there exists  $\eta > 0$  such that it also is in  $\mathcal{D}_{\gamma-\eta, \kappa}$ . Fix such an  $\eta$ .

Let  $\kappa(\gamma, \eta, \varepsilon)$  be large, to be chosen later, and let  $\varepsilon_*(\kappa)$  be such that  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*}$  per Fact 4.14. Then the right-hand side is at most

$$n^{1+\varepsilon_*} \sum_{r=\varepsilon \log n}^{\frac{1}{2} \log_\gamma n} \mathbb{P}(|Z_r| \geq n^{-\varepsilon_*} \gamma^r) = n^{1+\varepsilon_*} \sum_{r=\varepsilon \log n}^{\frac{1}{2} \log_\gamma n} \mathbb{P}(|Z_r| \geq (\gamma^{1-\varepsilon_* \log \gamma / \varepsilon})^r).$$

Let  $\tilde{\gamma} = \gamma^{1-\varepsilon_* \log \gamma / \varepsilon}$  and take  $\kappa$  to be sufficiently large (so that  $\varepsilon_*$  is sufficiently small) that  $\tilde{\gamma} > \gamma - \eta/2$ . By Lemma 3.6, then, the right-hand side above is at most

$$C n^{1+\varepsilon_*} \sum_{r=\varepsilon \log n}^{\frac{1}{2} \log_\gamma n} r^\kappa \left( \frac{\gamma - \eta}{\gamma - \eta/2} \right)^{r\kappa}.$$

One then sees that if  $\kappa$  is large enough, the right-hand side will be  $o(n^{-10})$  as desired.  $\square$

**4.5. Treelike nature of the configuration model.** We can also use the breadth-first revealing procedures together with the volume growth estimates, to establish that the random graph given by the configuration model is typically  $(L, R)$ -Treelike for  $L = O(1)$  and  $R \leq \frac{1}{2} \log_\gamma n$ .

**Proof of Lemma 2.2.** By Lemma 2.4, with probability  $1 - o(n^{-10})$  the random graph  $\mathcal{G}$  has  $(\gamma, \varepsilon)$  volume growth, say for  $\varepsilon = 1/4$ , as long as  $\kappa$  is sufficiently large. Let us work on that event, so that  $|B_R(v)| \leq n^{1/2-\delta}$  for all  $v \in \{1, \dots, n\}$ .

Now fix any  $v$  and perform the breadth-first revealing of  $B_R(v)$  per Process 4.6. In order for  $B_R(v)$  to not be  $L$ -Treelike, it must be the case that for more than  $L$  different steps  $m$  in the first  $n^{1/2-\delta}$  steps, the half-edge  $f(A_{m-1})$  is being matched to a half-edge of  $\hat{E}(A_{m-1})$ . Call such a step *bad*. (If there were at most  $L$  bad steps, then the removal of the at-most  $L$  edges formed by those at-most  $L$  matchings in the revealing scheme, evidently leaves a tree.)

Uniformly over  $A_{m-1}$ , the probability of the  $m$ 'th step being bad is at most

$$(m \|\mathbf{d}_n\|_\infty) / (\|\mathbf{d}_n\|_1 - m).$$

We thus find that for every  $\ell \geq 1$ ,

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(B_R(v) \text{ is not } \ell\text{-Treelike}) \leq \mathbb{P}\left(\text{Bin}\left(n^{1/2-\delta}, \frac{n^{1/2-\delta} \|\mathbf{d}_n\|_\infty}{\|\mathbf{d}_n\|_1 - n^{1/2-\delta}}\right) > \ell\right). \quad (4.2)$$

Recall that the standard Chernoff bound applied to a binomial distribution with mean  $\mu = Np$  says that for every  $s \geq \mu$ ,

$$\mathbb{P}(\text{Bin}(N, p) \geq s) \leq e^{s-\mu} \left(\frac{s}{\mu}\right)^{-s}. \quad (4.3)$$

Using the assumption that  $\|\mathbf{d}_n\|_1 = \Omega(n)$  and recalling from Fact 4.14 that  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*(\kappa)}$ , (4.3) implies that the right-hand side of (4.2) is at most  $(Cn^{-2\delta+\varepsilon_*})^\ell$ . As a consequence, taking  $\kappa$  large enough that  $\varepsilon_* < \delta$ , and choosing  $L > 11\delta^{-1}$ , we would find that the probability of  $B_R(v)$  not being  $L$ -Treelike is  $o(n^{-11})$  for all  $v$ , and a union bound over  $v \in \{1, \dots, n\}$  implies the desired result.  $\square$

## 5. THE FK-DYNAMICS SHATTERS QUICKLY ON RANDOM GRAPHS

Our first goal in this section is to prove the following theorem establishing the existence of  $T = O(1)$  (in continuous-time) such that for  $t \geq T$ , the FK-dynamics chain on the random graph  $\mathcal{G}$  initialized from the all-wired configuration (i.e., all edges are open), denoted  $X_{\mathcal{G},t}^1$ , is shattered, i.e., all the connected components of the FK-dynamics configuration are small; recall Definition 2.11 for a precise formulation.

**Theorem 5.1.** Fix  $q \geq 1$ ,  $\gamma > 1$ , and  $p < p_u(q, \gamma)$ . For every  $\varepsilon > 0$ , there exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , the following holds. There exists  $T = O(1)$  such that for every  $t \geq T$  and every  $v$ , with probability  $1 - o(n^{-10})$ ,  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is such that

$$\mathbb{P}(|\mathcal{C}_v(X_{\mathcal{G},t}^1)| \geq n^\varepsilon) \leq o(n^{-10}).$$

We will then use this to conclude Theorem 2.12, demonstrating that if  $t \geq T$ , the boundary condition  $X_{\mathcal{G},t}^1$  induces on any ball of volume  $o(\sqrt{n})$  is  $O(1)$ -sparse.

By monotonicity of the FK-dynamics, for every  $\mathcal{G}$ , we have that  $X_{\mathcal{G},t}^1 \geq \pi_{\mathcal{G}}$ , from which it follows that Theorem 5.1 holds under  $\pi_{\mathcal{G}}$ .

**Corollary 5.2.** Fix  $q \geq 1$ ,  $\gamma > 1$ , and  $p < p_u(q, \gamma)$ . For every  $\varepsilon > 0$ , there exists  $\kappa$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , then for every  $v$ , with probability  $1 - o(n^{-10})$ ,  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is such that

$$\pi_{\mathcal{G}}(|\mathcal{C}_v(X_{\mathcal{G},t}^1)| \geq n^\varepsilon) \leq o(n^{-10}).$$

While we do not use this corollary here, it may find applications elsewhere.

**5.1. Couplings and revealing schemes for the FK-dynamics on random graphs.** In this section, we define our central revealing procedure for exposing the random graph together with a family of coupled FK-dynamics on subsets of the random graph  $\mathcal{G}$ , which together stochastically dominate  $X_{\mathcal{G},t}^1$ . This revealing procedure is essential to the proof of shattering for  $X_{\mathcal{G},t}^1$  in the uniqueness region after  $O(1)$  continuous-time.

A similar revealing scheme of random graphs with an FK-dynamics chain on top of it was introduced in [4]. The revealing scheme we use here builds on that, but makes some key modifications to deal with the non-uniformity of the degrees and the lack of deterministic control on the volume of small balls of  $\mathcal{G}$ . These changes are explicitly laid out in Remark 5.11.

**5.1.1. Grand coupling of localized FK-dynamics.** In this section, we define a grand coupling of FK-dynamics on all possible edge subsets of the random graph  $\mathcal{G}$  in such a way that all monotonicities of the model are maintained.

Recall from Definition 4.1 that we use  $\mathfrak{M}_n$  as the set of all (not necessarily perfect) matchings of the complete graph on the  $\|\mathbf{d}_n\|_1$  many half-edges. The matching  $A$  is naturally identified with a set of edges on the original vertex set  $\{1, \dots, n\}$ , each pairing of two half-edges becoming an edge between the vertices they belong to. Abusing notation, we will understand  $A$  both as a matching element of  $\mathfrak{M}_n$  and as a multiset of edges on  $\{1, \dots, n\}$ .

**Definition 5.3.** For an element  $A \in \mathfrak{M}_n$ , let  $\partial A$  be the set of vertices in  $\{1, \dots, n\}$  having half-edges that are not matched in  $A$ . Let  $\pi_A^1$  be the random-cluster measure on the edge set  $A$  with wired boundary conditions wiring all vertices of  $\partial A$ . Let  $(X_{A,t}^1)_{t \geq 0}$  be the continuous-time FK-dynamics initialized from the all wired configuration on  $A$  (as well as outside  $A$ ), and making updates in  $A$  according to  $\pi_A^1$ .

We next place all the chains  $(\mathcal{X}_t^1)_{t \geq 0} = \{(X_{A,t}^1)_{t \geq 0}\}_{A \in \mathfrak{M}_n}$  on all possible matchings  $A \in \mathfrak{M}_n$ , in the same probability space, and construct an explicit coupling of them.

**Definition 5.4.** The probability space we consider will consist of the following sources of randomness:

- (1) Independently assign each possible edge  $e$  (i.e., each possible pairing of two half-edges), a sequence of times  $\mathfrak{T}_e = (T_1^e, T_2^e, \dots)$  given by the rings of a rate-1 Poisson clock; and
- (2) Independently assign each possible  $e$  a sequence of  $\text{Unif}[0, 1]$  random-variables  $\mathfrak{U}_e = (U_1^e, U_2^e, \dots)$ .

We denote by  $\mathcal{F}_t$  the  $\sigma$ -algebra generated by the processes  $(\mathfrak{T}_e)_e$  up to time  $t$ , as well as the corresponding set of random variables in  $(\mathfrak{U}_e)_e$ .

**Definition 5.5.** From  $(\mathfrak{T}, \mathfrak{U})$  construct the processes  $(X_{A,t}^1)_{t \geq 1}$  for all  $A \in \mathfrak{M}_n$  as follows:

- (1) Let  $0 < t_1 < t_2 < \dots$  be the (almost surely distinct) times in  $\bigcup_i \bigcup_e \{T_i^e\}$  in increasing order.
- (2) Initialize  $X_{A,0}^1(e) = 1$  for all  $e$ ; i.e., the all wired configuration.
- (3) For each  $i \geq 1$ , let

$$X_{A,t}^1 = X_{A,t_{i-1}}^1 \quad \text{for all } t \in [t_{i-1}, t_i).$$

Then, let  $(e_i, k_i)$  be the unique pair for which  $t_i = T_{k_i}^{e_i}$  and define  $X_{A,t_i}^1$  by setting

$$X_{A,t_i}^1(e) = X_{A,t_{i-1}}^1(e) \quad \text{for all } e \in A \setminus \{e_i\}$$

and

$$X_{A,t_i}^1(e_i) = \begin{cases} 1 & \text{if } e_i \in A \text{ and } U_{e_i, k_i} \leq \varrho; \\ 0 & \text{if } e_i \in A \text{ and } U_{e_i, k_i} > \varrho; \end{cases}$$

for

$$\varrho = \pi_A^1(\omega(e_i) = 1 \mid \omega(A \setminus \{e_i\}) = X_{A,t_{i-1}}^1(A \setminus \{e_i\}));$$

i.e., if  $e_i \in A$ , we resample  $e_i$  given the remainder of the configuration on  $A$ , together with the wired boundary condition on  $\partial A$ , using the same  $U_{e_i, k_i}$  for every  $X_{A,t_i}^1$  such that  $e_i \in A$ .

The following two observations are elementary to observe, but of central importance to our analysis.

**Observation 5.6.** *The coupling defined in Definition 5.5 is a monotone coupling. In particular, we have  $X_{A',t}^1 \leq X_{A,t}^1$  for any two matchings  $A, A' \in \mathfrak{M}_n$  with  $A \subset A'$ . As such, we have for every  $\mathcal{G}$  that*

$$X_{\mathcal{G},t}^1(e) \leq \min_{A \in \mathfrak{M}_n: A \subset E(\mathcal{G})} X_{A,t}^1(e), \quad \text{for all } e \in E(\mathcal{G}) \text{ and all } t \geq 0.$$

**Observation 5.7.** *For every  $A$ , the configuration  $X_{A,t}^1$  depends only on  $(\mathfrak{T}_e, \mathfrak{U}_e)_{e \in A}$ , and in fact only on their restriction to  $\mathcal{F}_t$  (the  $\sigma$ -algebra generated by elements of  $\mathfrak{T}, \mathfrak{U}$  before time  $t$ ).*

We now use the coupling of Definition 5.5 to design a coupling of FK-dynamics chains on random graphs.

**Definition 5.8.** Let  $\mathbb{P}_t^1$  be the distribution over pairs  $(\mathcal{G}, \omega_t)$  where  $\omega_t$  is a random-cluster configuration on  $\mathcal{G}$  that results by first drawing  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , then drawing  $\omega_t \sim \mathbb{P}(X_{\mathcal{G},t}^1 \in \cdot)$ . Likewise, for every set  $A \in \mathfrak{M}_n$ , let  $\mathbb{P}_{A,t}^1$  be the distribution over pairs  $(\mathcal{G}, \omega_{A \cap E(\mathcal{G}),t})$  where  $\omega_{A,t} \sim \mathbb{P}(X_{A,t}^1 \in \cdot)$ . Couple, under the distribution  $\mathbb{P}$ , the family of distributions  $(\mathbb{P}_{A,t}^1)_{A \in \mathfrak{M}_n, t \geq 0}$  by selecting the same random graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  for all of them, then using the coupling of Definition 5.5 for the family of FK-dynamics  $(X_{A,t}^1)_{A \in E(\mathcal{G}), t \geq 0}$  on  $\mathcal{G}$ .

In this manner, we have constructed a monotone coupling of  $(\mathcal{G}, (X_{A,t}^1)_{t \geq 1})_{A \in E(\mathcal{G})}$ . Note that we use this coupling for  $A$  which we know have  $A \subset E(\mathcal{G})$ , so that the randomness of the graph is only over the edges of  $E(\mathcal{G}) \setminus A$ , which we note  $X_{A,t}^1$  is independent of; thus the role of this coupling is only to put the random graphs with their random-cluster configurations on the same probability space. We emphasize that by Observation 5.7, if  $A \cap B = \emptyset$ , then  $X_{A,t}^1$  and  $X_{B,t}^1$  are independent.

**5.1.2. The joint revealing procedure.** We now construct a revealing procedure for  $\mathcal{G}$  and a configuration  $\tilde{\omega}_t$  on  $\mathcal{G}$  that stochastically dominates  $X_{\mathcal{G},t}^1$ . This will be inspired by the simultaneous revealing procedure first introduced in [4], with significant modifications that streamline that argument, and deal with the heterogeneity of degrees and volumes of balls in  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ .

**Definition 5.9.** Given a degree sequence  $\mathbf{d}_n$ , a vertex set  $\mathcal{V} \subset \{1, \dots, n\}$ , and a matching  $\mathcal{A} \in \mathfrak{M}_n$ , let  $\hat{E}(\mathcal{V}, \mathcal{A})$  be the set of half-edges incident to  $\mathcal{V}$ , and not matched in  $\mathcal{A}$ .

We note that  $\hat{E}(V(\mathcal{A}), \mathcal{A}) = \hat{E}(\mathcal{A})$  from Definition 4.5.

For a matching on half-edges,  $\mathcal{A}_0 \in \mathfrak{M}_n$ , so that  $\mathcal{A}_0 \subset E(\mathcal{G})$ , and a subset of vertices  $\mathcal{V}_0 \subset V(\mathcal{G})$ , we construct a procedure to expose (a set containing) the connected components  $\mathcal{C}_{\mathcal{V}_0}(X_{\mathcal{G},t}^1(E(\mathcal{G}) \setminus \mathcal{A}_0))$ , i.e., the union of all the connected components of the vertices in  $\mathcal{V}_0$  in the configuration  $X_{\mathcal{G},t}^1(E(\mathcal{G}) \setminus \mathcal{A}_0)$ . The two examples to have in mind are

- (1)  $\mathcal{A}_0 = \emptyset$  and  $\mathcal{V}_0 = \{v\}$ , used to prove Theorem 5.1;
- (2)  $\mathcal{A}_0 = E(B_R(v))$ , and  $\mathcal{V}_0 = \partial B_R(v)$ , used to prove Theorem 2.12.

In this revealing procedure, the index  $m$  will count the number of “steps”, and  $k$  will track the number of “generations”. We will keep track of the following variables through our revealing process:

- $\mathcal{A}_m$ : the element of  $\mathfrak{M}_n$  that has been shown to be a subset of  $E(\mathcal{G})$  through step  $m$ ;
- $\tilde{\omega}_m$ : the random-cluster configuration revealed through step  $m$ ;
- $\hat{\mathcal{E}}_k$ : the set of half-edges we want to explore out of in the  $k$ -th generation.

For  $\mathcal{A} \in \mathfrak{M}_n$ , recall from Process 4.12 that  $B_r(\hat{e}; \mathcal{A}^c)$  is revealed in a breadth-first manner, with the breadth-first exploration rejecting branches through vertices in  $V(\mathcal{A})$ .

The revealing process, with parameters  $(p, q, \gamma, r, t)$ , and input  $(\mathcal{V}_0, \mathcal{A}_0)$  is defined as follows: see Figure 5.1–5.2 for a depiction to accompany the below.

**Process 5.10.****Inputs:**  $(p, q, \gamma); \quad t \geq 0; \quad r \geq 1; \quad \mathcal{V}_0 \subset \{1, \dots, n\}; \quad \mathcal{A}_0 \in \mathfrak{M}_n;$ **Initialize:**  $k = 0; \quad m = 1; \quad \hat{\mathcal{E}}_0 = \hat{E}(\mathcal{V}_0, \mathcal{A}_0); \quad \tilde{\omega}_0 = \emptyset;$ **for each**  $k \geq 0$  **while**  $\hat{\mathcal{E}}_k \neq \emptyset$ **for each**  $\hat{e} \in \hat{\mathcal{E}}_k,$ 

1. Reveal the ball of radius  $r$  out from  $\hat{e}$  in the random graph  $\mathcal{G}$ :
  - (a) Set  $\hat{e}_m = \hat{e}$ . Conditionally on  $\mathcal{A}_{m-1}$ , reveal  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  per Process 4.12.
  - (b) Set  $\mathcal{A}_m = \mathcal{A}_{m-1} \cup B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$ .
  - (c) Let  $A_m := \mathcal{A}_m \setminus \mathcal{A}_{m-1}$  be the set of new edges revealed to belong to  $E(\mathcal{G})$ .
2. Simulate the FK-dynamics up to time  $t$  on the newly revealed edge set  $A_m$ :
  - (a) Reveal  $\mathfrak{F}_{A_m, t} := \{(\mathfrak{T}^e)_{e \in A_m}, (\mathfrak{U}^e)_{e \in A_m}\} \cap \mathcal{F}_t$  (as defined in Definition 5.4).
  - (b) Generate  $X_{A_m, t}^1$  from  $\mathfrak{F}_{A_m, t}$  per Definition 5.5.
3. Update the configuration  $\tilde{\omega}_m$ , the boundary half-edges  $\hat{\mathcal{E}}_k$ , and the step count  $m$ :
  - (a) Concatenate  $X_{A_m, t}^1$  with  $\tilde{\omega}_{m-1}$  to obtain a new configuration  $\tilde{\omega}_m$  on  $\mathcal{A}_m$ .
  - (b) Add to  $\hat{\mathcal{E}}_{k+1}$  all un-matched half-edges of vertices in  $\partial \mathcal{A}_m$  that are in the component of  $\mathcal{V}_0$  in  $\tilde{\omega}_m(\mathcal{A}_m)$  and are not in  $\hat{\mathcal{E}}_j$  for any  $j \leq k$ .
  - (c) Increase  $m$  by 1.

**Remark 5.11.** Before proceeding, let us describe the specific differences between the current revealing scheme and that of [4], as well as why these changes are needed to overcome difficulties arising from heterogeneity of the underlying degree sequence. The main changes are as follows:

- (1) The revealing process is based on half-edges rather than vertices: this ensures that the revealing of the ball  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  does not reveal the degrees of the vertices from which the exploration proceeds (which could potentially have high-degree and introduce correlations between generations).
- (2) The revealing of the ball  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  does not continue exploring if it encounters any vertex of  $V(\mathcal{A}_{m-1}^c)$ . This is important because if  $B_r(\hat{e}_m)$  intersects a dense region of  $\mathcal{A}_{m-1}^c$  that has already been revealed, then the volume of  $B_r(\hat{e}_m)$  would not be independent of  $\mathcal{A}_{m-1}^c$ .
- (3) The FK-dynamics is simulated in continuous time, rather than discrete time. This introduces additional independence so that the number of updates taken by each of the localized FK-dynamics chains  $X_{A_m, t}^1$  are truly independent of one another.

For ease of notation, let  $k_\emptyset$  be the first  $k$  such that  $\hat{\mathcal{E}}_k = \emptyset$ , i.e., the total number of generations of the revealing procedure. Let

$$\mathfrak{m}_k = \sum_{0 \leq j \leq k} |\hat{\mathcal{E}}_j|,$$

be the total number of half-edges for which  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  was revealed in step 1.a) of Process 5.10, so that  $\mathfrak{m}_{k_\emptyset}$  counts the total number of half-edges out of which a ball is ever revealed. Let

$$\tilde{\omega} = \tilde{\omega}_{\mathfrak{m}_{k_\emptyset}}(\mathcal{A}_{\mathfrak{m}_{k_\emptyset}} \setminus \mathcal{A}_0)$$

be the random-cluster configuration revealed when the process terminates. The following key observation is a direct consequence of Observation 5.6 and the construction of Process 5.10.

**Observation 5.12.** *Under the procedure of Process 5.10, we have*

$$\tilde{\omega}(\mathcal{A}_{\mathfrak{m}_{k_\emptyset}} \setminus \mathcal{A}_0) \geq X_{\mathcal{G}, t}^1(\mathcal{A}_{\mathfrak{m}_{k_\emptyset}} \setminus \mathcal{A}_0).$$



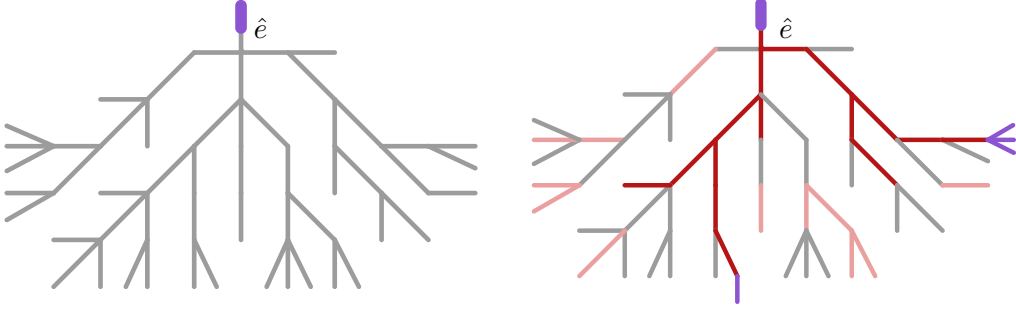


FIGURE 5.1. Left: We initialize the revealing process with  $r = 6$  from  $\mathcal{V}_0 = \{v\}$ ,  $\mathcal{A}_0 = \emptyset$  and the half-edge  $\hat{e} = \hat{e}_1 = \hat{\mathcal{E}}_0$  (purple). The process begins by revealing  $A_1 = B_r(\hat{e}; \mathcal{A}_0^c)$ , depicted in gray. Right: The process then reveals the configuration  $X_{A_1,t}^1$  (open edges shown in red/pink). Half-edges belonging to vertices in  $\partial A_1$  that are in the  $X_{A_1,t}^1$ -connected component of  $v$  (red) are added to form  $\hat{\mathcal{E}}_1$  (purple).

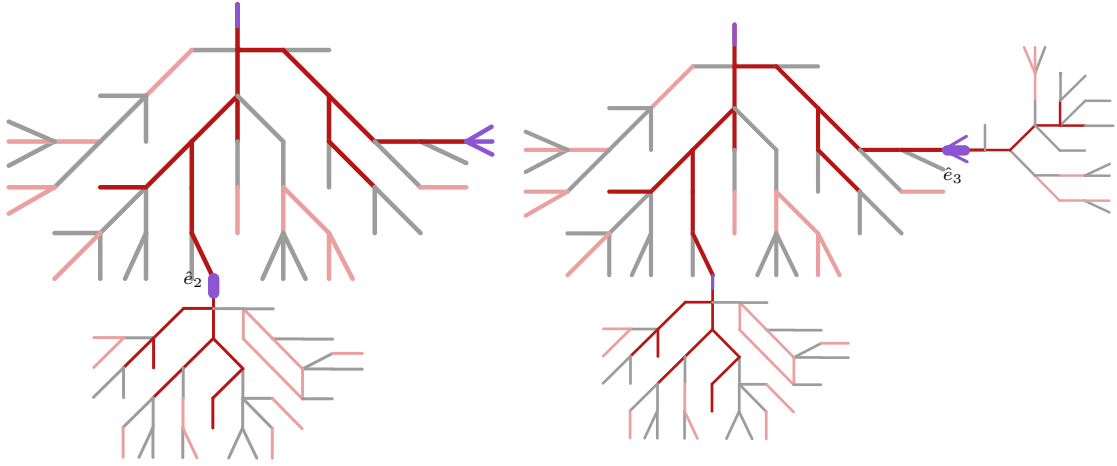


FIGURE 5.2. Left: Proceeding from above, in the next generation, starting from  $\hat{e}_2 \in \hat{\mathcal{E}}_1$ , the process reveals the edges of  $A_1 = B_r(\hat{e}_2, \mathcal{A}_1^c)$  in  $\mathcal{G}$ ; in this case, this is not a tree as it contains a single cycle. The FK-dynamics configuration  $X_{A_2,t}^1$  is then generated and concatenated with  $\tilde{\omega}_1$  to form  $\tilde{\omega}_2$ . Right: The process continues with  $\hat{e}_3$ , revealing  $B_r(\hat{e}_3; \mathcal{A}_2^c)$  with the FK-dynamics configuration  $X_{A_3,t}^1$  on top of it.

In particular, the connected component of each vertex in  $\mathcal{V}_0$  in  $X_{\mathcal{G},t}^1(E(\mathcal{G}) \setminus \mathcal{A}_0)$  is a subset of a connected component of a vertex in  $\mathcal{V}_0$  in  $\tilde{\omega}$ .

Thus, if  $N_\omega(A)$  denotes the number of vertices in non-trivial (i.e., non-singleton) components of the boundary condition induced by  $\omega(E(\mathcal{G}) \setminus A)$  on  $A$ , then

$$N_{X_{\mathcal{G},t}^1}(\mathcal{A}_0) \leq N_{\tilde{\omega}}(\mathcal{A}_0).$$

With Observation 5.12 in hand, we focus on obtaining the stretched exponential tail bound of Theorem 5.1 for the size of  $\mathcal{C}_v(\tilde{\omega})$  (the component of  $v$  in  $\tilde{\omega}$ ) and likewise, the sparsity bound of Theorem 2.12.

5.1.3. *Constructing a dominating branching process.* Towards proving Theorem 5.1 and 2.12, we construct a branching process (ours will be a size-dependent one but we use the terminology nonetheless) which we will show stochastically dominates the sequence  $(\hat{\mathcal{E}}_k)_{k \geq 0}$  of our joint revealing process. This process  $(Z_k)_{k \geq 0}$  will then be shown to be sub-critical, with good tail bounds.

**Definition 5.13.** Initialize  $Z_0 = |\hat{\mathcal{E}}_0|$ . Let  $(Z_k)_{k \geq 0}$  be the branching process, which for each  $k \geq 0$ , has progeny  $(\chi_{i,k})_{i \leq Z_k}$ , i.e.,

$$Z_{k+1} = \sum_{i \leq Z_k} \chi_{i,k}.$$

The progeny  $\chi_{i,k}$  are distributed as follows. First, let  $(\hat{\mathcal{T}}_r^{i,k})_{i,k}$  be i.i.d. single-source Galton–Watson trees of depth  $r$ , with offspring distribution  $\mathbb{P}_{\underline{\mathbf{d}}_n}$  (from Definition 4.10); recall the single-source here refers to the fact that this is a tree of depth  $r$  whose first generation deterministically has one offspring; beyond that first edge, it is simply a Galton–Watson tree of depth  $r - 1$  with offspring distribution  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ . Then the offspring distribution (parametrized by  $p, q, \gamma, \varepsilon$  and  $r, t$ ), is as follows:

- (1) With probability  $n^{-1/2}$ , let  $\chi_{i,k} = \|\mathbf{d}_n\|_\infty^r (\sum_{j < k} Z_j + \sum_{j < i} \chi_{j,k})$ ;
- (2) Otherwise,
  - (a) If  $\hat{\mathcal{T}}_r^{i,k}$  does not satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition (Definition 3.7), let  $N_t = |E(\hat{\mathcal{T}}_r^{i,k})|$  and let  $\chi_{i,k}$  be a sum of  $N_t$  independent random variables drawn from  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ .
  - (b) If  $\hat{\mathcal{T}}_r^{i,k}$  does satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition, first generate a configuration on  $\hat{\mathcal{T}}_r^{i,k}$  by running FK-dynamics with  $(1, \mathcal{C})$  boundary conditions, initialized from  $\omega_0 \equiv 1$  for time  $t$ . Let  $N_t$  be the number of vertices of  $\partial \hat{\mathcal{T}}_r^{i,k}$  that are connected to the root, and let  $\chi_{i,k}$  be a sum of  $N_t$  independent random variables drawn from  $\mathbb{P}_{\underline{\mathbf{d}}_n}$ .

Let us motivate the above construction. Item (1) in the definition of  $\chi_{i,k}$  corresponds to cases when either

- The ball  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  is not a tree, or
- The ball  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c) \setminus \{\hat{e}_m\}$  intersects some already exposed vertex in  $\mathcal{A}_{m-1}$ .

The  $n^{-1/2}$  probability of item (1) is because we will need to take  $\mathcal{A}_0$  possibly as large as  $n^{-\frac{1}{2}-o(1)}$ . On the latter of these two events, the connected component of  $\mathcal{V}_0$  may, in one step, incorporate many edges of  $\mathcal{A}_{m-1}$ , by virtue of an already revealed large connected component of  $\mathcal{A}_{m-1}$ . In this case, the best a priori bound we can place on the progeny is the total number of edges revealed up to that point.

In case (2), the newly revealed ball is indeed a tree and does not intersect any already exposed vertex of  $\mathcal{A}_{m-1}$ . On the indicator of this event, by Proposition 4.13, the ball is stochastically below  $\hat{\mathcal{T}}_r^{i,k}$ ; cases (2a)–(2b) then distinguish whether or not the dominating tree satisfies the  $(\gamma, \varepsilon)$ -tree-growth condition. This is important because if the tree does not satisfy the condition,  $p < p_u(q, \gamma)$  will *not* be sub-critical for the  $(1, \mathcal{C})$  random-cluster model on  $\hat{\mathcal{T}}_r^{i,k}$ , and we can only take the full boundary of the tree as our bound on the size of the component of the tree's root.

**5.1.4. Dominating the revealing process by the branching process.** We are now in position to state the main two lemmas of this section, comparing the revealing procedure to the branching process of Definition 5.13, and then establishing its sub-criticality.

Recall that  $m_0 = |\hat{\mathcal{E}}_0|$  and for each  $k \geq 1$ ,  $m_{k+1} = m_k + |\hat{\mathcal{E}}_{k+1}|$ , i.e., in each generation  $k$ ,  $m_k$  is the number of half-edges we explore out from. This will be the quantity which we compare to the population of the branching process  $(Z_k)_k$  of Definition 5.13. For notational simplicity, write  $\mathcal{A}_\infty = \mathcal{A}_{m_{k_\infty}}$ .

**Lemma 5.14.** *For every  $\mathcal{A}_0, \mathcal{V}_0$  such that  $|\mathcal{A}_0|, |\mathcal{V}_0| \leq n^{\frac{1}{2}-\delta}$  for  $\delta > 0$ , and every  $\ell \geq 1$ ,*

$$(|\hat{\mathcal{E}}_j| \mathbf{1}_{\{m_j \leq n^{1/2-\delta/2}\}})_{j \leq \ell} \leq (Z_j)_{j \leq \ell}.$$

Furthermore, we have

$$|\mathcal{A}_\infty \setminus \mathcal{A}_0| \mathbf{1}_{\{m_\infty \leq n^{1/2-\delta/2}\}} \leq \gamma^r \sum_{j=0}^{\infty} Z_j.$$

The proof of Lemma 5.14 is briefly deferred to the next subsection; before that proof, we observe that the lemma reduces the analysis of the set of exposed vertices through the revealing process of  $(\mathcal{G}, \tilde{\omega})$ , and

thus, the clusters of  $X_{G,t}^1$ , to the analysis of the process  $(Z_k)$ , which for most steps is a simple branching process with progeny distribution typically dictated by connectivity probabilities in the wired measure on trees satisfying a  $(\gamma, \varepsilon)$ -tree-growth condition, but occasionally makes large state-dependent jumps.

Our claim is that if  $r$  and  $t$  are chosen to be sufficiently large, but  $O(1)$ , the dominating branching process will be sub-critical. To formalize this claim, let

$$\bar{\tau}_{\text{MIX}} := \max_{\mathbb{T}_r \text{ of } (\gamma, \varepsilon)\text{-tree-growth}} t_{\text{MIX}}(\mathbb{T}_r, (1, \mathcal{O})) . \quad (5.1)$$

i.e., the maximum over all possible trees of depth  $r$  satisfying the  $(\gamma, \varepsilon)$ -tree-growth condition, of the (continuous-time) mixing time with  $(1, \mathcal{O})$  boundary conditions. Now define the *burn-in time*

$$T_{\text{BURN}} = T_{\text{BURN}}(C_0, r) := C_0 \gamma^r \bar{\tau}_{\text{MIX}} . \quad (5.2)$$

**Lemma 5.15.** *Fix  $q \geq 1$ ,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . For  $\varepsilon$  sufficiently small and  $C_0, r$  and  $\kappa$  sufficiently large, if  $t \geq T_{\text{BURN}}(C_0, r)$  and  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , the branching process of Definition 5.13 satisfies the following tail bound: if  $Z_0 \leq n^{\frac{1}{2}-\delta}$ , then for every  $M \geq 1$ , and every  $\lambda : \lambda Z_0 \leq n^{\frac{1}{2}-\frac{\delta}{2}}$ ,*

$$\mathbb{P}\left(\sum_{j \geq 0} Z_j \geq (1 + \lambda)Z_0\right) \leq CM \exp\left(\frac{\lambda^{1/M} Z_0}{C \|\mathbf{d}_n\|_{\infty}^{(2+M)r}}\right) + Cn^{-\delta M/2} .$$

Roughly, the constant  $M$  can be thought of as the number of times the “bad” offspring distribution of item (1) in Definition 5.13 is selected, allowing the total population to double, and away from such “bad” updates, we will show that the branching process indeed satisfies exponential tails.

**5.2. Coupling the revealing process to the branching process.** We next prove the desired stochastic domination relation between the revealing process  $(\hat{\mathcal{E}}_k)_k$  and  $Z_k$  by constructing a coupling between the two such that the former is below the latter while the total population is at most  $n^{1/2-\delta/2}$ .

**Proof of Lemma 5.14.** We proceed by induction over  $\ell \geq 0$ . The base case,  $Z_0 = |\hat{\mathcal{E}}_0|$ , is by construction. Now fix  $\ell \geq 1$  and suppose by way of induction that the following stochastic domination holds:

$$(|\hat{\mathcal{E}}_j| \mathbf{1}_{\{\mathbf{m}_j \leq n^{1/2-\delta/2}\}})_{j \leq \ell-1} \leq (Z_j)_{j \leq \ell-1} .$$

Thus, there exists a monotone coupling of the sequence on the left-hand side, such that it is below the sequence  $(Z_j)_{j \leq \ell-1}$  in the natural element-wise ordering on the sequence. Working on that coupling, it suffices for us to then show that on the event  $\{\mathbf{m}_{\ell-1} \leq n^{1/2-\delta/2}\}$ , for every  $m \in \{\mathbf{m}_{\ell-1} + 1, \dots, \mathbf{m}_{\ell}\}$ , the distribution of the children of  $\hat{e}_m$  is stochastically below the progeny distribution of Definition 5.13. Here, by children of  $\hat{e}_m$ , we mean the set of half-edges added in step 3.(b) of Process 5.10. In what follows, denote that set by  $\Xi(\hat{e}_m)$ .

Define the event  $\Gamma_{\text{good}}$  on the revealed ball  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  as the event that

$$V(B_r(\hat{e}_m; \mathcal{A}_{m-1}^c) \setminus \{\hat{e}_m\}) \cap V(\mathcal{A}_{m-1}) = \emptyset \quad \text{and} \quad B_r(\hat{e}_m; \mathcal{A}_{m-1}^c) \text{ is a tree} .$$

On the bad event  $\Gamma_{\text{good}}^c$ , we take the a priori bound of  $\hat{E}(\mathcal{A}_m)$  on the set  $\Xi(\hat{e}_m)$ , namely assuming that in the worst-case all exposed half-edges of  $\mathcal{A}_m$ , both those in  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$ , and those of  $\mathcal{A}_{m-1}$  become connected up to  $\mathcal{V}_0$  in  $\tilde{\omega}_m$ . By the inductive hypothesis, the number of such half-edges is at most  $\|\mathbf{d}_n\|_{\infty}^r$  times the population of the branching process up to that step, given by  $\sum_{j < k} Z_j + \sum_{j < i} \chi_{j,k}$ , where the  $\|\mathbf{d}_n\|_{\infty}^r$  comes from assuming that in each of these steps the corresponding ball we revealed in the graph has maximal size. We claim that the probability of  $\Gamma_{\text{good}}^c$  is at most  $n^{-1/2}$ . To see this, notice that in the breadth-first revealing of  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$ , the probability that the next half-edge that gets matched is matched either with a vertex having an edge in  $\mathcal{A}_{m-1}$ , or an already exposed vertex of  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  is at most

$$\frac{|\hat{E}(\mathcal{A}_m)|}{\|\mathbf{d}_n\|_1 - |\hat{E}(\mathcal{A}_m)|} \leq \frac{\mathbf{m}_{\ell} \|\mathbf{d}_n\|_{\infty}}{\|\mathbf{d}_n\|_1 - \mathbf{m}_{\ell}} .$$

Assuming that  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  (for  $\kappa$  to be chosen sufficiently large later), by Fact 4.14,  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*(\kappa)}$ . Using this upper bound, the bound  $m_\ell \leq n^{1/2-\delta/2}$  (as otherwise the indicator on the left-hand side of the desired stochastic domination would be zero), and the lower bound of  $\|\mathbf{d}_n\|_1 \geq \Omega(n)$ , we see that this probability is at most  $n^{1/2+\delta/2-\varepsilon_*r}$ . Through the revealing of  $B_r(\hat{e}_m, \mathcal{A}_{m-1}^c)$  we make at most  $\|\mathbf{d}_n\|_\infty^r$  attempts at such a bad matching, and thus a union bound implies that

$$\mathbb{P}(B_r(\hat{e}_m; \mathcal{A}_{m-1}^c) \in \Gamma_{\text{good}}^c \mid \mathfrak{F}_{m-1}, m_\ell \leq n^{1/2-\delta/2}) \leq n^{-1/2-\delta/2+2\varepsilon_*r},$$

where  $\mathfrak{F}_{m-1}$  is the filtration generated by the randomness of the revealing procedure through the  $(m-1)$ 'th step. The right-hand side above is at most  $n^{-1/2}$  as long as  $\kappa$  is large enough that  $\varepsilon_*(\kappa) < \frac{1}{2}\delta(2r)^{-1}$ .

Now work on the event that  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c) \in \Gamma_{\text{good}}$ , and recall from Proposition 4.13 that in this case the ball is stochastically dominated by (and in particular there exists a coupling such that it can be embedded as a subset of) a Galton–Watson tree of depth  $r$  with offspring distribution  $\mathbb{P}_{\mathbf{d}_n}$ . This is the law of  $\hat{\mathcal{T}}_r^{i, \ell-1}$ .

Evidently, on the event that  $\hat{\mathcal{T}}_r^{i, \ell-1}$  does not satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition, the number of children  $|\Xi(\hat{e}_m)|$  is at most the number of half-edges emanating from  $\partial B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$ , which is at most  $|\partial \hat{\mathcal{T}}_{r+1}^{i, \ell-1}|$ , or a sum of  $|\partial \hat{\mathcal{T}}_r^{i, \ell-1}| \leq N_t$  independent draws from  $\mathbb{P}_{\mathbf{d}_n}$ .

Finally, suppose we are on the event that  $\hat{\mathcal{T}}_r^{i, \ell-1}$  does satisfy the  $(\gamma, \varepsilon)$ -tree-growth condition, so that  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  does as well. In that case, by Proposition 4.13, there is a coupling such that the graph  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  is a subgraph of  $\hat{\mathcal{T}}_r^{i, \ell-1}$ . One can then couple the FK-dynamics chain  $X_{A_m, t}^1$  to  $Y_t^1$ , the FK-dynamics chain on  $\hat{\mathcal{T}}_r^{i, \ell-1}$  with its  $(1, \emptyset)$  boundary conditions run for time  $t$  initialized from  $Y_0 \equiv 1$  such that  $X_{A_m, t}^1$  is below  $Y_t^1$  with probability one. In particular, the vertices of  $\partial A_m$  which are in the open cluster of  $\hat{e}_m$  in  $X_{A_m, t}^1$ , are a subset of the vertices of  $\partial \hat{\mathcal{T}}_r^{i, \ell-1}$  which are in the open cluster of the root in  $Y_t^1$ , so that the number of them, call it  $N(X_{A_m, t}^1)$  is less than  $N_t = N(Y_t)$ . Since the law of  $X_{A_m, t}^1$  is independent of the choice of boundary vertices, and thus degree sequence at  $\partial A_m$ , the number of half-edges added in step 3.(b) of Process 5.10 is a sum of  $N(X_{A_m, t}^1)$  independent draws from the empirical degree distribution at that point, which is stochastically below  $\mathbb{P}_{\mathbf{d}_n}$ . Therefore, this establishes the domination on this event of the number of children of  $\hat{e}_m$  by item (2b) of the dominating branching process.

In order to then deduce the domination of  $|\mathcal{A}_\infty \setminus \mathcal{A}_0|$  by  $\gamma^r$  times the total population of the dominating branching process, we make the following observation. On the event  $\Gamma_{\text{good}}^c$ , we were already bounding  $\Xi(\hat{e}_m)$  by  $|\mathcal{A}_m \setminus \mathcal{A}_0|$ , and that in turn by  $\chi_{i, k}$ , even without the factor of  $\gamma^r$ . Similarly in the case of (2a). In the event of (2b), we notice that by the  $(\gamma, \varepsilon)$ -tree-growth condition, the number of edges in  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  is at most  $\gamma^r$ .  $\square$

**5.3. Sub-criticality of an auxiliary branching process.** The branching process of Definition 5.13 is not a branching process in a traditional sense, as when it follows item (1) in the definition, its offspring count is state-dependent. Such offspring can create large jumps in the total population, and lead to difficulties in the analysis. We analyze the process by means of an auxiliary branching process that captures the behavior of  $(Z_j)_j$  in between its rare state-dependent steps. More formally, we say an offspring of the branching process of Definition 5.13 is bad if item (1) of Definition 5.13 is taken.

**Definition 5.16.** Consider the *auxiliary* branching process  $(\tilde{Z}_j)_j$  which is defined exactly as in Definition 5.13, except its offspring are conditioned to never be bad. Namely, let  $(\tilde{\chi}_{i, k})_{i, k}$  be a sequence of i.i.d. draws from item (2) of Definition 5.13, and for a fixed  $\tilde{Z}_0$ , construct  $(\tilde{Z}_j)_j$  iteratively by  $\tilde{Z}_j = \sum_{i \leq Z_{j-1}} \tilde{\chi}_{i, j}$ .

The following lemma establishes sub-criticality and tail bounds for the auxiliary branching process—in other words, the branching process during the epochs between the bad updates of  $(Z_j)_j$ .

**Lemma 5.17.** Fix  $q \geq 1$ ,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . For  $\varepsilon$  sufficiently small and  $C_0, r$  and  $\kappa$  sufficiently large, if  $t \geq T_{\text{BURN}}(C_0, r)$  and  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , the auxiliary branching process  $(\tilde{Z}_j)_j$  is uniformly sub-critical, i.e.,

$\limsup_n \mathbb{E}[\tilde{\chi}] < 1$ . Furthermore, it satisfies the following tail bound: for all  $\lambda$  sufficiently large,

$$\mathbb{P}\left(\sum_{j \geq 0} \tilde{Z}_j \geq \lambda \tilde{Z}_0\right) \leq C \exp\left(-\frac{\lambda \tilde{Z}_0}{C \|\mathbf{d}_n\|_\infty^{2r}}\right).$$

**Proof of Lemma 5.17: sub-criticality.** Let us begin by calculating the mean of the offspring distribution of the auxiliary branching process, which corresponds to the offspring distribution of Definition 5.13 conditional on being from item (2). By construction,

$$\begin{aligned} \mathbb{E}[\tilde{\chi}_{i,k}] &= \mathbb{E}[\mathbf{1}\{\hat{\mathcal{T}}_r^{i,k} \notin (\gamma, \varepsilon)\text{-tree-growth}\} |\hat{\mathcal{T}}_r^{i,k}|] \mathbb{E}_{\mathbf{d}_n}[D] \\ &\quad + \mathbb{E}[\mathbf{1}\{\hat{\mathcal{T}}_r^{i,k} \in (\gamma, \varepsilon)\text{-tree-growth}\} |\mathcal{C}_\rho(Y_t^1) \cap \partial \hat{\mathcal{T}}_r^{i,k}|] \mathbb{E}_{\mathbf{d}_n}[D]. \end{aligned}$$

We can bound the first term by Cauchy–Schwarz as

$$\begin{aligned} \mathbb{E}[\mathbf{1}\{\hat{\mathcal{T}}_r^{i,k} \notin (\gamma, \varepsilon)\text{-tree-growth}\} |\hat{\mathcal{T}}_r^{i,k}|] \mathbb{E}_{\mathbf{d}_n}[D] \\ < \gamma \cdot \mathbb{P}(\hat{\mathcal{T}}_r^{i,k} \notin (\gamma, \varepsilon)\text{-tree-growth})^{1/2} \mathbb{E}[|\hat{\mathcal{T}}_r^{i,k}|^2]^{1/2}. \end{aligned}$$

The probability on the right-hand side is at most  $\eta^{-\kappa \varepsilon r}$  for some  $\eta$  small, by Corollary 3.8 if  $(\mathbf{d}_n) \in \mathcal{D}_{\gamma, \kappa}$ . The expectation above is at most  $C\gamma^r$  using the moment bound of (3.11). Thus taking  $\kappa$  large depending on  $\varepsilon$ , we see that this product is exponentially small in  $r$ , and can be taken as close to 0 as desired by taking  $r$  sufficiently large.

Turning to the second term in the expansion of  $\mathbb{E}[\tilde{\chi}_{i,k}]$ , we can first bound it by

$$\begin{aligned} \mathbb{E}_{\mathbf{d}_n}[D] \mathbb{E}[\mathbf{1}\{\hat{\mathcal{T}}_r^{i,k} \in (\gamma, \varepsilon)\text{-tree-growth}\} |\mathcal{C}_\rho(Y_t^1) \cap \partial \hat{\mathcal{T}}_r^{i,k}|] \\ < \gamma \cdot \sup_{\mathbb{T}_r \in (\gamma, \varepsilon)\text{-tree-growth}} \mathbb{E}[|\mathcal{C}_\rho(Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1) \cap \partial \mathbb{T}_r|]. \end{aligned}$$

where  $(Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1)_{t \geq 0}$  is a continuous-time FK-dynamics on the tree  $\mathbb{T}_r$  with  $(1, \mathcal{O})$  boundary conditions, initialized from the all-wired configuration. Now recall that the stationary measure of  $Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1$  is  $\pi_{\mathbb{T}_r}^{(1, \mathcal{O})}$ . By the  $(\gamma, \varepsilon)$ -tree-growth condition,  $|\mathbb{T}_r| \leq \gamma^r$ . As such, there exists some  $C_{r, \gamma} > 0$  such that the mixing time of  $Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1$ , i.e.,  $\bar{\tau}_{\text{MIX}}$  as defined in (5.1), is at most  $C_{r, \gamma}$ . By sub-multiplicativity of total-variation distance (see e.g., [48]), then, if  $t \geq T_{\text{BURN}}(C_0, r) = C_0 \gamma^r \bar{\tau}_{\text{MIX}}$  as defined in (5.2), we have

$$\sup_{\mathbb{T}_r \in (\gamma, \varepsilon)\text{-tree-growth}} \|\mathbb{P}(Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1 \in \cdot) - \pi_{\mathbb{T}_r}^{(1, \mathcal{O})}\|_{\text{TV}} \leq C \exp(-C_0 \gamma^r / C).$$

Using this, for every  $\mathbb{T}_r \in (\gamma, \varepsilon)$ -tree-growth, we can bound the expectation

$$\begin{aligned} \mathbb{E}[|\mathcal{C}_\rho(Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1) \cap \partial \mathbb{T}_r|] &\leq \mathbb{E}_{\pi_{\mathbb{T}_r}^{(1, \mathcal{O})}}[|\mathcal{C}_\rho(\omega) \cap \partial \mathbb{T}_r|] + |\partial \mathbb{T}_r| \|\mathbb{P}(Y_{\mathbb{T}_r^{(1, \mathcal{O})}, t}^1 \in \cdot) - \pi_{\mathbb{T}_r}^{(1, \mathcal{O})}\|_{\text{TV}} \\ &\leq |\partial \mathbb{T}_r| \max_{v \in \partial \mathbb{T}_r} \pi_{\mathbb{T}_r}^{(1, \mathcal{O})}(v \in \mathcal{C}_\rho(\omega)) + C |\partial \mathbb{T}_r| e^{-C_0 \gamma^r / C}. \end{aligned}$$

Using the fact that  $\mathbb{T}_r$  has  $(\gamma, \varepsilon)$ -tree-growth and using the bound of Corollary 2.7 to bound the probability of a leaf being in the component of the root, we bound the above by

$$C \gamma^r \hat{p}^{(1-\varepsilon)r} + C \gamma^r e^{-C_0 \gamma^r / C}.$$

Recall that when  $p < p_u(q, \gamma)$ , we have  $\hat{p} < 1/\gamma$ , from which it follows that for sufficiently small  $\delta, \varepsilon$  and sufficiently large  $C_0$ , uniformly over large  $r$  the above quantity is strictly less than  $1/\gamma$ , so that when multiplied by  $\mathbb{E}_{\mathbf{d}_n}[D] < \gamma$ , it is strictly less than 1. Combining this with the bound on the first term in the expectation, we find that there exist  $\varepsilon(p, q, \gamma)$  and  $C_0(\gamma)$  such that for all  $r$  sufficiently large, we have  $\limsup_n \mathbb{E}[\tilde{\chi}_{i,k}] < 1$  as desired.  $\square$

**Proof of Lemma 5.17: tail bounds.** Having established sub-criticality of the dominating branching process, we now wish to boost this to tail bounds on the number of generations, and total population of the branching process. For this, we use the traditional random-walk exploration of a branching process. Namely, the population beyond  $\tilde{Z}_0$  can be expressed as a sum of i.i.d.'s and we can write the active population in the branching process beyond the first generation as the killed random walk

$$\tilde{Z}_0 + \sum_{i \leq N_0} (\tilde{\chi}_i - 1), \quad \text{where} \quad N_0 = \inf \left\{ j : \tilde{Z}_0 + \sum_{i \leq j} (\tilde{\chi}_i - 1) = 0 \right\},$$

where  $(\tilde{\chi}_i)_i$  are i.i.d. copies from the offspring distribution of Definition 5.13. Observe that with this representation, the *total* population of the branching process is exactly  $N_0$ . Then, we can express tail bounds for this branching process's total population as

$$\mathbb{P} \left( \sum_{0 \leq j < \infty} \tilde{Z}_j \geq \lambda \tilde{Z}_0 \right) \leq \mathbb{P}(N_0 > \lambda \tilde{Z}_0) \leq \mathbb{P} \left( \tilde{Z}_0 + \sum_{i \leq \lambda \tilde{Z}_0} (\tilde{\chi}_i - 1) > 0 \right).$$

Consider the random variable  $\tilde{\chi}_i - 1$ ; its mean satisfies  $\mathbb{E}[\tilde{\chi}_1 - 1] \leq -\eta$  for some  $\eta > 0$ , by the sub-criticality established in the previous proof. Thus this is a sum of  $\lambda \tilde{Z}_0$ -many i.i.d. random variables, the sum has mean smaller than  $-\eta \lambda \tilde{Z}_0$ , and the increments are bounded in  $\ell_\infty$  by  $\|\mathbf{d}_n\|_\infty^r$ . Thus,

$$\mathbb{P} \left( \sum_{i \leq \lambda \tilde{Z}_0} (\tilde{\chi}_i - 1) > -\tilde{Z}_0 \right) \leq \mathbb{P} \left( \left| \sum_{i \leq \lambda \tilde{Z}_0} (\tilde{\chi}_i - 1) - \mathbb{E}[\tilde{\chi}_1 - 1] \right| > \eta(\lambda - \eta^{-1}) \tilde{Z}_0 \right)$$

As long as  $\lambda > 2\eta^{-1}$ , by Hoeffding's inequality, this gives

$$\mathbb{P} \left( \sum_{i \leq \lambda \tilde{Z}_0} (\tilde{\chi}_i - 1) > -\tilde{Z}_0 \right) \leq C \exp \left( -\frac{\lambda \tilde{Z}_0}{C \|\mathbf{d}_n\|_\infty^{2r}} \right)$$

as desired.  $\square$

**5.4. Controlling the original branching process by a sum of auxiliary branching processes.** Given the sub-criticality and tail bounds for the auxiliary branching process, we can now obtain tail bounds on the original dominating branching process  $(Z_j)_j$  as required by Lemma 5.15. Let us now construct a process out of i.i.d. copies of the auxiliary branching process, that stochastically dominates the original branching process. Let  $(\tilde{Z}_j^{(i)})_j$  be i.i.d. copies of the branching process of Definition 5.16, with initializations  $\tilde{Z}_0^{(1)} = Z_0$  and  $\tilde{Z}_0^{(i)} = \|\mathbf{d}_n\|_\infty^r \sum_{j \geq 0} \tilde{Z}_j^{(i-1)}$  given by the total population of the previous auxiliary branching process.

In what follows, for fixed  $\lambda$ , consider the stopping generation

$$\kappa_\lambda = \min \left\{ k : \sum_{0 \leq j \leq k} Z_j > \lambda Z_0 \right\}.$$

Let  $\Gamma_{\text{bad}, M}$  be the event that there are at most  $M$  many bad offspring in the first  $\kappa_\lambda$  many generations of the branching process  $(Z_j)_j$ . The following stochastic domination is self-evident by construction.

**Claim 5.18.** *Given the above construction,  $(\sum_{j \geq 0} Z_j) \mathbf{1}_{\Gamma_{\text{bad}, M}} \leq \sum_{1 \leq i \leq M} \sum_{j \geq 0} \tilde{Z}_j^{(i)}$ .*

Given this stochastic domination, we can now establish Lemma 5.15.

**Proof of Lemma 5.15.** By a union bound, we have

$$\mathbb{P} \left( \sum_{j \geq 0} Z_j \geq \lambda Z_0 \right) \leq \mathbb{P}(\Gamma_{\text{bad}, M}^c) + \mathbb{P} \left( \sum_{1 \leq i \leq M} \sum_{j \geq 0} \tilde{Z}_j^{(i)} \geq \lambda Z_0 \right).$$

The first probability is bounded as follows: for every  $\lambda : \lambda Z_0 \leq n^{\frac{1}{2} - \frac{\delta}{2}}$ , we have

$$\mathbb{P}(\Gamma_{\text{bad}, M}^c) \leq \mathbb{P} \left( \text{Bin}(\lambda Z_0, n^{-1/2}) \geq M \right) \leq C n^{-\delta M/2}.$$



The second probability above can be bounded as

$$\mathbb{P}\left(\sum_{1 \leq i \leq M} \sum_{j \geq 0} \tilde{Z}_j^{(i)} \geq \lambda Z_0\right) \leq \sum_{1 \leq i \leq M} \mathbb{P}\left(\sum_{j \geq 0} \tilde{Z}_j^{(i)} \geq \frac{1}{M^{1/M} \|\mathbf{d}_n\|_\infty^{r_i}} \lambda^{1/M} \tilde{Z}_0^{(i)}\right)$$

Indeed, if for every  $i$ ,  $\sum_{j \geq 0} \tilde{Z}_j^{(i)} \leq M^{-1/M} \lambda^{1/M} \|\mathbf{d}_n\|_\infty^{-r_i} \tilde{Z}_0^{(i)}$ , then  $\sum_{1 \leq i \leq M} \sum_{j \geq 0} \tilde{Z}_j^{(i)} \leq \lambda Z_0$ . In order to now bound the right-hand side, we use the tail bounds of Lemma 5.17 to deduce that

$$\mathbb{P}\left(\sum_{1 \leq i \leq M} \sum_{j \geq 0} \tilde{Z}_j^{(i)} \geq \lambda Z_0\right) \leq CM \exp\left(-\frac{\lambda^{1/M} Z_0}{CM^{1/M} \|\mathbf{d}_n\|_\infty^{(2+M)r}}\right).$$

Combined with the bound on  $\mathbb{P}(\Gamma_{\text{bad}, M}^c)$ , we obtain the desired result.  $\square$

**5.5. Tail bounds on cluster sizes, and shattering of the dynamics.** We are now in a position to conclude the proof of the tail bounds on clusters of  $X_{\mathcal{G}, t}^1$ , and use that to deduce that  $X_{\mathcal{G}, t}^1$  is  $(K, R)$ -Sparse, except with probability  $o(n^{-5})$ . We begin by using Lemmas 5.14–5.15 to prove the following tail bound on  $|\mathcal{A}_\infty|$ , which we recall counts the number of edges exposed through the revealing process of Process 5.10.

**Lemma 5.19.** *Fix  $\delta > 0$  and consider the revealing procedure for any initial pair  $(\mathcal{V}_0, \mathcal{A}_0)$  having  $|\mathcal{A}_0|, |\mathcal{V}_0|$  and  $|\hat{\mathcal{E}}_0|$  all at most  $n^{\frac{1}{2}-\delta}$ . There exist  $C_0(p, q, \gamma), r(p, q, \gamma)$  in the definition of  $T_{\text{BURN}}$  in (5.2) and  $\kappa(p, q, \gamma)$  such that for all  $t \geq T_{\text{BURN}}$  the following holds. For all  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ ,  $M \geq 1$  and  $\lambda : \lambda |\hat{\mathcal{E}}_0| \leq n^{\frac{1}{2}-\delta}$ ,*

$$\mathbb{P}\left(|\mathcal{A}_\infty| \geq |\mathcal{A}_0| + \gamma^r (\lambda |\hat{\mathcal{E}}_0|)\right) \leq CM \exp\left(\frac{\lambda^{1/M} Z_0}{C \|\mathbf{d}_n\|_\infty^{(M+2)r}}\right) + Cn^{-\delta M/2}.$$

*Proof.* Define the following stopping generation

$$\varsigma = \inf \left\{ \ell : \mathfrak{m}_{\ell-1} > \lambda |\hat{\mathcal{E}}_0| \right\}.$$

Similarly define  $\varsigma_Z$  as the first  $\ell : \sum_{j \leq \ell-1} Z_j > \lambda Z_0$ . Under the monotone coupling of Lemma 5.14, if  $\varsigma_Z = \infty$ , then  $\varsigma = \infty$ , the indicators in the lemma are both 1, and both

$$(|\hat{\mathcal{E}}_j|)_j \leq (Z_j)_j, \quad \text{and} \quad |\mathcal{A}_\infty \setminus \mathcal{A}_0| \leq \gamma^r \sum_{j=0}^{\infty} Z_j,$$

hold. Therefore, we obtain

$$\mathbb{P}\left(|\mathcal{A}_\infty \setminus \mathcal{A}_0| \geq \gamma^r (\lambda |\hat{\mathcal{E}}_0|)\right) \leq \mathbb{P}\left(\sum_{k \geq 0} Z_k \geq \lambda Z_0\right).$$

Lemma 5.15 then implies the desired result.  $\square$

We next use Lemma 5.19 and Observation 5.12 to deduce tail estimates on the volume and radius of the cluster in  $X_{\mathcal{G}, t}^1$  containing  $v$ , when  $t \geq T_{\text{BURN}}$ .

**Proof of Theorem 5.1.** Fix some  $v \in \{1, \dots, n\}$ , let  $\mathcal{A}_0 = \emptyset$  and let  $\mathcal{V}_0 = \{v\}$  in Process 5.10. In this case  $\hat{\mathcal{E}}_0$  is the set of half-edges out from  $v$ , and thus  $|\hat{\mathcal{E}}_0| = d_v$ . By Observation 5.12, for each  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , the cluster of  $v$  in the configuration  $X_{\mathcal{G}, t}^1$ , denoted  $\mathcal{C}_v(X_{\mathcal{G}, t}^1)$  is a subset of  $\mathcal{C}_v(\tilde{\omega})$ , which in turn is a subset of  $V(\mathcal{A}_{\mathfrak{m}_\emptyset} \setminus \mathcal{A}_0)$ . Let  $C_0, r$  be sufficiently large constants and take  $t \geq T = T_{\text{BURN}}(C_0, r)$ . Then, we have

$$|\mathcal{C}_v(X_{\mathcal{G}, t}^1)| \leq |\mathcal{C}_v(\tilde{\omega})| \leq |V(\mathcal{A}_{\mathfrak{m}_\emptyset} \setminus \mathcal{A}_0)| \leq 2|\mathcal{A}_{\mathfrak{m}_\emptyset} \setminus \mathcal{A}_0|.$$

By Lemma 5.19 and the above, if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , there exists  $C(p, q, \gamma)$  such that

$$\mathbb{P}\left((\mathcal{G}, X_{\mathcal{G}, t}^1) : |\mathcal{C}_v(X_{\mathcal{G}, t}^1)| \geq \gamma^r (\lambda d_v)\right) \leq CM \exp\left(\frac{\lambda^{1/M} d_v}{C \|\mathbf{d}_n\|_\infty^{(M+2)r}}\right) + Cn^{-\delta M/2}.$$

Let  $\delta = 1/4$  and let  $M = 200$ , for instance. For any fixed small  $\varepsilon > 0$ , by taking  $\kappa$  sufficiently large, by Fact 4.14,  $\|\mathbf{d}_n\|_\infty^{(M+2)r} < n^{\varepsilon/4rM}$ ; then taking  $\lambda = n^{(M-1)\varepsilon/4M}$ , we satisfy that  $\lambda d_v \leq n^{\frac{1}{2}-\delta}$ . Then we see that

$$\gamma^r(\lambda d_v) < n^\varepsilon, \quad \text{and} \quad \frac{\lambda^{1/M} d_v}{C \|\mathbf{d}_n\|_\infty^{(M+2)r}} > n^{\varepsilon/4M}/C.$$

In turn, the probability above is at most  $o(n^{-24})$ . Observing that  $\mathbb{P}((\mathcal{G}, X_{\mathcal{G},t}^1) : X_{\mathcal{G},t}^1 \in \cdot) = \mathbb{E}_{\text{CM}(\mathbf{d}_n)}[\mathbb{P}(X_{\mathcal{G},t}^1 \in \cdot)]$ , we can use Markov's inequality to write

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}\left(\mathcal{G} : \mathbb{P}\left(X_{\mathcal{G},t}^1 : |\mathcal{C}_v(X_{\mathcal{G},t}^1)| \geq n^\varepsilon\right) \geq n^{-12}\right) \leq n^{-12},$$

implying the desired result.  $\square$

We next establish that the  $(K, R)$ -Sparse property for the random-cluster configuration on  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  holds with high probability for all  $t \geq T_{\text{BURN}}$ . Towards this, we introduce the following notation.

**Definition 5.20.** Given a graph  $\mathcal{G}$ , a vertex subset  $\mathcal{V}_0$ , an edge subset  $\mathcal{A}_0$ , and a configuration  $\omega$  on  $E(\mathcal{G})$ , define  $\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(\omega)$  as the subset of vertices in  $\mathcal{V}_0$  in non-singleton components in the boundary condition induced by  $\omega(E(\mathcal{G}) \setminus \mathcal{A}_0)$ .

**Lemma 5.21.** Fix  $q \geq 1$ ,  $\gamma > 1$ ,  $p < p_u(q, \gamma)$ , and  $\delta > 0$ . Let  $R \leq (\frac{1}{2} - \delta) \log_\gamma n$ . There exist  $\kappa$ ,  $K$  as well as  $C_0$  and  $r$ , such that for every  $v \in \{1, \dots, n\}$  for all  $t \geq T_{\text{BURN}}(C_0, r)$  and all  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  the following holds for  $\mathcal{A}_0 = B_R(v)$  and  $\mathcal{V}_0 = \partial B_R(v)$ :

$$\mathbb{P}((\mathcal{G}, X_{\mathcal{G},t}^1) : |\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(X_{\mathcal{G},t}^1)| > K) \leq o(n^{-10}).$$

We use Theorem 5.1 to bound the number of chances the revealing process of Process 5.10 has to reconnect to the vertices of  $\mathcal{V}_0 = \partial B_R(v)$ . Intuitively, since the components of  $\tilde{\omega}$  have (stretched) exponential tail bounds, the number of chances at reconnecting is of the same order as  $|\mathcal{V}_0|$ ; since  $R$  is such that  $|\mathcal{V}_0| \leq n^{1/2-\delta}$ , the number of such connections (each possibly inducing a non-trivial boundary component) will be dominated by an  $\text{Bin}(n^{1/2+\delta}, n^{-1/2-\delta})$  random variable, yielding the desired tail bound on the probability of this exceeding some large  $K$ .

**Proof of Lemma 5.21.** Fix  $v \in \{1, \dots, n\}$  and  $\delta > 0$ , and any  $R \leq (\frac{1}{2} - \delta) \log_\gamma n$ . First of all, we recall from Lemma 2.4, that if we let  $\Gamma$  be the event that  $\mathcal{G}$  has the  $(\gamma, \varepsilon)$ -volume-growth property, then

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\Gamma^c) \leq o(n^{-10}).$$

We will henceforth work on the event  $\Gamma$ . Reveal the sub-graph  $B_R(v)$  on the event  $\Gamma$  (such that its volume is at most  $\gamma^R$ ) and initialize  $\mathcal{V}_0 = \partial B_R(v)$  and  $\mathcal{A}_0 = E(B_R(v))$ . We apply the revealing procedure of Process 5.10 with this initialization. Recall from Observation 5.12 that the FK-clusters of  $\mathcal{V}_0$  induced by  $\tilde{\omega}(E(\mathcal{G}) \setminus \mathcal{A}_0)$  are a subset of  $\mathcal{A}_{\mathbf{m}_{k\emptyset}} \setminus \mathcal{A}_0$ , and the configuration  $\tilde{\omega}$  satisfies  $\tilde{\omega}(\mathcal{A}_{\mathbf{m}_{k\emptyset}} \setminus \mathcal{A}_0) \geq X_{\mathcal{G},t}^1(\mathcal{A}_{\mathbf{m}_{k\emptyset}} \setminus \mathcal{A}_0)$ . Thus, the sets  $\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(\tilde{\omega})$  and  $\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(X_{\mathcal{G},t}^1)$ , are subsets of  $\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(\mathcal{A}_{\mathbf{m}_{k\emptyset}} \setminus \mathcal{A}_0)$ .

Through the revealing process of Process 5.10, for each  $m$ , the edges of  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$  are revealed one at a time via the breadth-first revealing per Processes 4.4 and 4.12. Therefore,  $|\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(\mathcal{A}_{\mathbf{m}_{k\emptyset}} \setminus \mathcal{A}_0)|$  is at most the number of times during the revealing of  $\mathcal{A}_{\mathbf{m}_{k\emptyset}}$ , that a half-edge is matched up to a half-edge belonging to a vertex that had already been discovered. For a fixed  $m$ , consider the revealing of  $B_r(\hat{e}_m; \mathcal{A}_{m-1}^c)$ . Conditionally on a discovered edge set  $\mathcal{A}$  the law of the next half-edge to be matched is uniform amongst all un-matched half-edges. Thus, uniformly over the history of the revealing process up to that point, the probability that the next half-edge to be matched is matched up to a vertex of  $V(\mathcal{A})$  is at most

$$\frac{|\mathcal{A}_{\mathbf{m}_{k\emptyset}}| \|\mathbf{d}_n\|_\infty}{\|\mathbf{d}_n\|_1 - |\mathcal{V}_{\mathbf{m}_{k\emptyset}}| \|\mathbf{d}_n\|_\infty}.$$

We thus obtain for a sufficiently large constant  $\Lambda$  (depending on  $p, q, \gamma, r$ ), for all  $L \geq 1$ ,

$$\begin{aligned} \mathbb{P}\left((\mathcal{G}, \tilde{\omega}) : \mathcal{G} \in \Gamma, |\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(X_{\mathcal{G}, t}^1)| > L\right) \\ \leq \mathbb{P}\left(\Gamma, |\mathcal{A}_{\mathfrak{m}_k \emptyset}| > n^{\frac{1}{2} - \frac{\delta}{2}}\right) + \mathbb{P}\left(\text{Bin}\left(n^{\frac{1}{2} - \frac{\delta}{2}} \|\mathbf{d}_n\|_\infty, 2n^{-\frac{\delta}{2} - \frac{1}{2}} \|\mathbf{d}_n\|_\infty\right) > L\right). \end{aligned}$$

By the bound  $|\hat{\mathcal{E}}_0| \leq \|\mathbf{d}_n\|_\infty \gamma^R \leq n^{\frac{1}{2} - \frac{2\delta}{3}}$  as long as  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*}$  for a sufficiently small  $\varepsilon_*$  (which holds as long as  $\kappa$  is sufficiently large in  $\delta, M$  by Fact 4.14), we can apply Lemma 5.19 with a sufficiently large choice of  $M$  to deduce that the first term is at most

$$\mathbb{P}\left(\Gamma, |\mathcal{A}_\infty| > n^{\frac{1}{2} - \frac{\delta}{2}}\right) \leq \mathbb{P}\left(|\mathcal{A}_\infty| \geq |\mathcal{A}_0| + \gamma^r \|\mathbf{d}_n\|_\infty n^{\frac{\delta}{10}}\right) \leq o(n^{-10}).$$

For the second term, notice that the mean of the binomial is  $2n^{-\delta} \|\mathbf{d}_n\|_\infty^2$ . As long as  $\kappa$  is sufficiently large so that  $\|\mathbf{d}_n\|_\infty \leq n^{\varepsilon_*}$  for sufficiently small  $\varepsilon_* < \delta/4$ , this is  $o(n^{-\delta/2})$ . Thus, by the Chernoff bound for the binomial (4.3), for every fixed  $L \geq 1$ ,

$$\mathbb{P}\left((\mathcal{G}, X_{\mathcal{G}, t}^1) : |\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(X_{\mathcal{G}, t}^1)| > L\right) \leq o(n^{-\frac{\delta L}{2} \wedge 10}). \quad (5.3)$$

Choosing  $L$  sufficiently large (depending on  $\delta$ ), we can make the right-hand side here  $o(n^{-10})$  as well.  $\square$

**Proof of Theorem 2.12.** Given Lemma 5.21, it is straightforward to deduce Theorem 2.12. Specifically, take  $K$  sufficiently large so that the right-hand side of Lemma 5.21 is  $o(n^{-10})$ . By a union bound,

$$\mathbb{P}\left((\mathcal{G}, X_{\mathcal{G}, t}^1) : X_{\mathcal{G}, t}^1 \text{ is not } (K, R)\text{-Sparse}\right) \leq \sum_v \mathbb{P}\left((\mathcal{G}, X_{\mathcal{G}, t}^1) : |\mathfrak{V}_{(\mathcal{V}_0, \mathcal{A}_0)}(X_{\mathcal{G}, t}^1)| > K\right) \leq o(n^{-9}). \quad (5.4)$$

By Markov's inequality,

$$\begin{aligned} \mathbb{P}_{\text{CM}(\mathbf{d}_n)}\left(\mathcal{G} : \mathbb{P}(X_{\mathcal{G}, t}^1 \text{ is not } (K, R)\text{-Sparse}) > n^{-6}\right) \\ \leq n^6 \mathbb{E}_{\text{CM}(\mathbf{d}_n)}[\mathbb{P}(X_{\mathcal{G}, t}^1 \text{ is not } (K, R)\text{-Sparse})], \end{aligned}$$

and the conclusion follows from the fact that the expectation on the right-hand side is exactly the probability on the left-hand side of (5.4).  $\square$

Let us conclude with a better bound in the special case of  $R = 0$  from Theorem 2.12; this will be applied to establish our mixing time lower bounds for the Ising/Potts Glauber dynamics.

**Lemma 5.22.** *Fix  $q, \gamma$  and suppose  $p < p_u(q, \gamma)$ . There exists  $\kappa$  such that for all  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , with probability  $1 - o(1)$  over  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ , for every  $v \in \{1, \dots, n\}$  and every  $\eta > 0$ ,*

$$\pi_{\mathcal{G}}\left(\omega(E_v^c) \text{ is not } \eta d_v\text{-Sparse}\right) \leq C \exp(-\eta d_v / C).$$

(Here  $\omega(E_v^c)$  is viewed as a boundary condition induced by  $\omega$  on  $E_v = \{e : e \ni v\}$ .)

*Proof.* Fix a small  $\varepsilon > 0$  and consider the following modification of the revealing process of Process 5.10.

- (1) Label the half-edges of the vertex  $v$   $\hat{e}_v^{(1)}, \dots, \hat{e}_v^{(d_v)}$
- (2) Perform the process of Process 5.10 with  $\mathcal{V}_0 = v$ ,  $\mathcal{A}_0 = \mathcal{A}_0^{(1)} := \emptyset$ , and  $\hat{\mathcal{E}}_0 = \hat{e}_v^{(1)}$ , stopped if either  $|\mathcal{A}_m^{(1)}| \geq n^\varepsilon$  or in step 1.(a) a bad step is taken, i.e., some previously exposed vertex gets matched with.
- (3) For  $i = 1, \dots, d_v$ , if  $\hat{e}_v^{(i)}$  is hitherto un-matched, set  $\mathcal{A}_0^{(i)}$  to be the set of all matched edges to that point, and run the process of Process 5.10 with  $\mathcal{V}_0 = v$ ,  $\mathcal{A}_0 = \mathcal{A}_0^{(i)}$ , and  $\hat{\mathcal{E}}_0 = \hat{e}_v^{(i)}$ , stopped if  $|\mathcal{A}_m^{(i)}| \geq n^\varepsilon$  or a bad step is taken.

Observe that in order for  $\omega(E_v^c)$  to not be  $\eta d_v$ -Sparse, there must have been more than  $\eta d_v/2$  many  $i$ 's for which the revealing process gets stopped (each such  $i$  adds at most two vertices to the set  $\mathfrak{V}_{(v,\emptyset)}(\omega)$  for  $\omega \sim \pi_G$ ). Throughout the entire procedure described above, at most  $d_v n^\varepsilon$  many edges are revealed, which for  $\varepsilon$  small and  $\kappa$  large is at most  $n^{1/4}$ . By Lemma 5.19 with  $M$  taken sufficiently large, for any fixed  $i$ , the probability of reaching  $|\mathcal{A}_m^{(i)}| \geq n^\varepsilon$  is at most  $o(n^{-1/4})$  uniformly over the history of the process up to that point. At the same time, for any fixed  $i$ , the probability of a bad step being taken for that revealing is at most

$$n^\varepsilon \cdot \frac{n^{1/4} \|\mathbf{d}_n\|_\infty}{\|\mathbf{d}_n\|_1 - n^{1/4} \|\mathbf{d}_n\|_\infty} \leq o(n^{-1/2}).$$

Putting the above together, the probability of more than  $\eta d_v/2$  many of the  $i$ 's being stopped is at most

$$\mathbb{P}(\text{Bin}(d_v, n^{-1/2}) \geq \eta d_v/2) \leq C \exp(-\eta d_v/C).$$

This in turn bounds the probability that  $\omega(E_v^c)$  is  $\eta d_v$ -Sparse as desired.  $\square$

## 6. CORRELATION DECAY AND MIXING TIME ON TREELIKE GRAPHS

Theorem 2.12 together with Lemma 2.2 reduce our analysis to treelike balls of radius  $(\frac{1}{2} - o(1)) \log_\gamma n$  with  $K$ -Sparse boundary conditions. In this section, we establish sharp bounds on the rate of correlation decay on such treelike graphs (Theorem 2.9) and bound the mixing time at these local scales (Lemma 6.7).

**6.1. Rate of correlation decay in treelike graphs.** To prove Theorem 2.9 we will closely follow the approach from [4], where an analogous result was proved for regular graphs (specifically see Proposition 3.3 in [4]). The key part of the extension is the use of the  $(\gamma, \varepsilon)$ -volume growth condition to enable the application of Lemma 2.7 to all sufficiently large subsets of the graph that are trees.

Let us fix an arbitrary vertex  $v \in V$  and for ease of notation set  $B := B_R(v)$  and for each  $1 \leq \ell \leq R$ , let  $Q_\ell = \{u \in B : d(u, v) \geq \ell\}$ . For a boundary condition  $\xi$  on  $\partial B$ , similarly to Definition 5.20 denote by  $\mathfrak{V}_{B,\xi}$  the set of vertices in non-trivial components of  $\xi$  (a component is non-trivial when it has at least two vertices). For any  $u \in B$  such that  $d(u, v) = \ell$ , let  $u \xleftrightarrow{Q_\ell} \mathfrak{V}_{B,\xi}$  denote the event that  $u$  is connected to  $\mathfrak{V}_{B,\xi}$  by a path of open edges fully contained in  $Q_\ell$ . Define the event

$$\Upsilon_{B,\xi} := \left\{ \omega \in \{0, 1\}^{E(B)} : |\{u \in B : d(u, v) = \ell, u \xleftrightarrow{Q_\ell} \mathfrak{V}_{B,\xi}\}| \geq 2 \text{ for all } 1 \leq \ell \leq R \right\}.$$

It was proved in [4] that on general graphs, the event  $\Upsilon_{B,\xi}$  controls the propagation of influence from  $\partial B$  to the vertex  $v$ .

Recall that  $E_v$  denotes the set of edges incident to the vertex  $v$ .

**Lemma 6.1** (Lemma 5.3 in [4]). *Fix a graph  $G = (V, E)$ , a vertex  $v \in V$  and consider the ball  $B_R(v)$ ; let  $\xi \geq \tau$  denote two boundary conditions on  $\partial B_R(v) = \{w \in B_R(v) : d(v, w) = R\}$ . Then,*

$$\|\pi_{B_R(v)}^\xi(\omega(E_v) \in \cdot) - \pi_{B_R(v)}^\tau(\omega(E_v) \in \cdot)\|_{\text{TV}} \leq \pi_{B_R(v)}^\xi(\Upsilon_{B_R(v),\xi}).$$

With this lemma in hand, we are able to provide the proof of Theorem 2.9.

**Proof of Theorem 2.9.** By the triangle inequality and Lemma 6.1, we have

$$\begin{aligned} \|\pi_{B_R(v)}^\xi(\omega(E_v) \in \cdot) - \pi_{B_R(v)}^\tau(\omega(E_v) \in \cdot)\|_{\text{TV}} &\leq \|\pi_{B_R(v)}^\xi(\omega(E_v) \in \cdot) - \pi_{B_R(v)}^0(\omega(E_v) \in \cdot)\|_{\text{TV}} \\ &\quad + \|\pi_{B_R(v)}^\tau(\omega(E_v) \in \cdot) - \pi_{B_R(v)}^0(\omega(E_v) \in \cdot)\|_{\text{TV}} \\ &\leq \pi_{B_R(v)}^\xi(\Upsilon_{B_R(v),\xi}) + \pi_{B_R(v)}^\tau(\Upsilon_{B_R(v),\tau}). \end{aligned}$$

Hence, it suffices to bound  $\pi_B^\xi(\Upsilon_{B,\xi})$  for an arbitrary vertex  $v$  of  $G$  and any  $K$ -Sparse boundary condition  $\xi$ . Fix any such  $v$  and let  $B = B_R(v)$ . Let  $H \subset E(B)$  be a set of at most  $L$  edges such that the subgraph  $(B, E(B) \setminus H)$  is a tree; the existence of such a set is guaranteed by the fact that  $B_R(v)$  is  $L$ -Treelike. Let  $\mathcal{Z} = \{d_1, \dots, d_k\}$  be the subset of distances (from  $v$ ) at which  $H$  contains at least one vertex. Observe that

each edge of  $H$  intersects either one or two consecutive depths (distances from  $v$ ) in  $\mathcal{Z}$  and thus  $|\mathcal{Z}| \leq 2L$  since  $B$  is  $L$ -Treelike. Letting  $d_0 = 0$  and  $d_{k+1} = R$ , for  $i = 0, \dots, k$  we define:

$$\mathcal{F}_i := \{u \in B : d_i < d(u, v) < d_{i+1}\}.$$

For each  $0 \leq i \leq k$ , the graph  $\mathcal{F}_i = (\mathcal{F}_i, E(\mathcal{F}_i))$  is a forest; observe that some  $\mathcal{F}_i$ 's might be empty. For each  $i$ , let  $\mathcal{T}_{ij} = (\mathcal{T}_{ij}, E(\mathcal{T}_{ij}))$  for  $j = 0, 1, \dots$  denote the distinct connected components (subtrees) of  $\mathcal{F}_i$  so that  $\mathcal{F}_i = \bigcup_{j \geq 0} \mathcal{T}_{ij}$ .

For  $\Upsilon_{B,\xi}$  to hold, there must exist two sequences of simple paths  $\Gamma = \gamma_0, \dots, \gamma_k$  and  $\Gamma' = \gamma'_0, \dots, \gamma'_k$  such that  $\gamma_i \subset E(\mathcal{T}_{ij})$  and  $\gamma'_i \subset E(\mathcal{T}_{ij'})$  with  $j \neq j'$  such that  $\gamma_i$  (resp.,  $\gamma'_i$ ) connects the root of  $\mathcal{T}_{ij}$  (resp.,  $\mathcal{T}_{ij'}$ ) to one of its leaves.

Observe that any simple path  $\mathcal{P}$  between  $v$  and  $\mathfrak{V}_{B,\xi}$  is completely determined by an ordered sequence of vertices from  $V(H)$  it uses and its endpoint in  $\mathfrak{V}_{B,\xi}$ . Moreover, it is associated to a unique sequence  $\Gamma$ , and each sequence  $\Gamma$  can in turn correspond to at most  $2^{|V(H)|k} \leq 4^{L^2}$  simple paths because there are at most  $2L$  vertices in  $V(H)$ . Since  $\xi$  is  $K$ -Sparse, there are at most  $K$  choices for the endpoint of the path between  $v$  and  $\mathfrak{V}_{B,\xi}$ . In total, we get that there are at most  $4^{L^2} K (2L + 1)!$  possible simple paths  $\Gamma$  (this is a crude upper bound, but it suffices for our purposes). A union bound then implies

$$\pi_B^\xi(\Upsilon_{B,\xi}) \leq [4^{L^2} K (2L + 1)!]^2 \cdot \sup_{\Gamma, \Gamma': V(\Gamma) \cap V(\Gamma') = \emptyset} \pi_B^\xi(\omega(\Gamma \cup \Gamma') = 1). \quad (6.1)$$

Fix any two such paths  $\Gamma, \Gamma'$ , and consider the probability that  $\omega(\Gamma \cup \Gamma') = 1$ . The paths  $\Gamma$  and  $\Gamma'$  are vertex-disjoint by construction, but the events that  $\Gamma$  and  $\Gamma'$  are open (i.e., that all of their paths are open) in  $\omega$  need not be independent. To make them so, we wire all vertices at depths in the set

$$\bigcup_{i=0}^{k+1} \{d_i - 1, d_i, d_i + 1\} \cap [0, R].$$

Let  $\tilde{\pi}_B$  be the resulting random-cluster distribution. The monotonicity of the random-cluster measure implies that

$$\pi_B^\xi(\omega(\Gamma \cup \Gamma') = 1) \leq \tilde{\pi}_B(\omega(\Gamma \cup \Gamma') = 1). \quad (6.2)$$

The distribution  $\tilde{\pi}_B$  is a product measure over the  $\mathcal{T}_{ij}$ 's with boundary condition  $(1, \circlearrowleft)$  in each  $\mathcal{T}_{ij}$ . Hence, since  $\Gamma$  and  $\Gamma'$  are such that for each  $i \geq 0$ ,  $\gamma_i$  and  $\gamma'_i$  belong to distinct subtrees  $\mathcal{T}_{\gamma_i}, \mathcal{T}_{\gamma'_i}$  of the forest  $\mathcal{F}_i$ , and we have

$$\tilde{\pi}_B(\omega(\Gamma \cup \Gamma') = 1) = \prod_{i=0}^k \pi_{\mathcal{T}_{\gamma_i}}^{(1, \circlearrowleft)}(\gamma_i) \prod_{i=0}^k \pi_{\mathcal{T}_{\gamma'_i}}^{(1, \circlearrowleft)}(\gamma'_i).$$

Let  $h_i = d_{i+1} - d_i$  be the height of the trees in  $\mathcal{F}_i$ . Then,

$$\tilde{\pi}_B(\omega(\Gamma \cup \Gamma') = 1) \leq \prod_{i: h_i > \sqrt{\varepsilon}R} \pi_{\mathcal{T}_{\gamma_i}}^{(1, \circlearrowleft)}(\gamma_i) \pi_{\mathcal{T}_{\gamma'_i}}^{(1, \circlearrowleft)}(\gamma'_i)$$

Since  $G$  satisfies the  $(\gamma, \varepsilon)$ -volume-growth condition of Definition 2.3, for each subtree  $\mathcal{T}_{\gamma_i}$  of height at least  $\sqrt{\varepsilon}R$ , for every vertex of  $\mathcal{T}_{\gamma_i}$  at distance at least  $\varepsilon R$  from  $\partial \mathcal{T}_{\gamma_i}$ , we have  $|\partial \mathcal{T}_{\gamma_i}| \leq \gamma^{h_i}$ . Hence, Lemma 2.7 implies that there exists a constant  $A > 0$  such that, uniformly over  $\Gamma, \Gamma'$ ,

$$\begin{aligned} \tilde{\pi}_B(\omega(\Gamma \cup \Gamma') = 1) &\leq A^{2L} \prod_{i: h_i > \sqrt{\varepsilon}R} \hat{p}^{2(1-\sqrt{\varepsilon})h_i} \\ &= A^{2L} \hat{p}^{2(1-\sqrt{\varepsilon}) \sum_{i: h_i > \sqrt{\varepsilon}R} h_i} \\ &\leq A^{2L} \hat{p}^{2(1-\sqrt{\varepsilon})(R-4L-2L\sqrt{\varepsilon}R)} = A' \hat{p}^{2(1-(2L+1)\sqrt{\varepsilon})R}, \end{aligned}$$

for a suitable constant  $A' = A'(A, L, K)$ . Plugging this bound into (6.1)–(6.2), we obtain

$$\pi_B^\xi(\Upsilon_{B,\xi}) \leq A' [K(2L + 1)!]^2 \hat{p}^{2(1-(2L+1)\sqrt{\varepsilon})R},$$

and the result follows taking  $C = 2A'[4^{L^2}K(2L+1)!]^2$ .  $\square$

**6.2. Local mixing of the FK-dynamics.** In this section, we prove the mixing time bound of Lemma 6.7 for treelike graphs with sparse boundary conditions. We start by recalling some standard background concerning mixing times, log-Sobolev inequalities, and the effects of random-cluster boundary conditions on these quantities.

**Log-Sobolev inequalities.** For a Markov chain on a finite state space  $\Omega$  with transition matrix  $P$ , reversible with respect to a distribution  $\mu$ , the Dirichlet form is defined for any function  $f : \Omega \rightarrow \mathbb{R}$  by

$$\mathcal{E}(f, f) := \frac{1}{2} \sum_{\omega, \omega' \in \Omega} \mu(\omega) P(\omega, \omega') (f(\omega) - f(\omega'))^2, \quad (6.3)$$

and its *log-Sobolev constant* is given by

$$\alpha(P) := \min_{f: \text{Ent}_\mu[f^2] \neq 0} \frac{\mathcal{E}(f, f)}{\text{Ent}_\mu[f^2]}, \quad \text{where} \quad \text{Ent}_\mu[f^2] = \mathbb{E}_\mu \left[ f^2 \log \frac{f^2}{\mathbb{E}_\mu[f^2]} \right]. \quad (6.4)$$

A *log-Sobolev inequality* takes the form  $\mathcal{E}(f, f) \geq \alpha \text{Ent}_\mu[f^2]$  for all functions  $f$ . It is a standard fact that this inequality implies exponential convergence with rate  $\alpha$  in total-variation distance to the stationary distribution (see, [21, Eq. (3.3)]).

**Fact 6.2.** Consider an ergodic Markov chain on a finite state space  $\Omega$  with transition matrix  $P$  reversible with respect to the distribution  $\mu$ . If the chain has a log-Sobolev constant  $\alpha = \alpha(P)$ ,

$$\max_{x_0 \in \Omega} \|\mathbb{P}(X_t^{x_0} \in \cdot) - \mu\|_{\text{TV}} \leq \frac{1}{\sqrt{2}} e^{-\alpha t} \left( \log \frac{1}{\min_{x \in \Omega} \mu(x)} \right)^{1/2},$$

where  $X_t^{x_0}$  is the chain after time  $t$ , started from initial state  $x_0$ .

**Boundary conditions and the FK-dynamics.** Two “similar” random-cluster boundary conditions (in terms of the wiring they induce) have similar effects on the underlying random-cluster distribution and on the behavior of the corresponding FK-dynamics. In turn, the Dirichlet form, and log-Sobolev constants of their corresponding dynamics should be “close” to one another. We compile here a number of definitions and results that formalize this idea.

**Definition 6.3** (Definition 2.1 from [5]). For two boundary conditions (partitions)  $\phi \leq \phi'$ , define  $D(\phi, \phi') := c(\phi) - c(\phi')$  where  $c(\phi)$  is the number of components in  $\phi$ . For two partitions  $\phi, \phi'$  that are not comparable, let  $\phi''$  be the smallest partition such that  $\phi'' \geq \phi$  and  $\phi'' \geq \phi'$  and set  $D(\phi, \phi') = c(\phi) - c(\phi'') + c(\phi') - c(\phi'')$ .

The following lemma is then straightforward from the definition of the random-cluster measure (1.2).

**Lemma 6.4** (Lemma 2.2 from [5]). Let  $G = (V, E)$  be an arbitrary graph,  $p \in (0, 1)$  and  $q > 0$ . Let  $\phi$  and  $\phi'$  be any two partitions of  $V$ , i.e., boundary conditions on  $G$ . Then, for all random-cluster configurations  $\omega \in \{0, 1\}^E$ , we have

$$q^{-2D(\phi, \phi')} \pi_G^{\phi'}(\omega) \leq \pi_G^\phi(\omega) \leq q^{2D(\phi, \phi')} \pi_G^{\phi'}(\omega).$$

The following corollary follows immediately from Lemma 6.4, the definition of the transition matrix of the FK-dynamics, and Theorem 4.1.1 in [57].

**Corollary 6.5.** Let  $G = (V, E)$  be an arbitrary graph,  $p \in (0, 1)$  and  $q > 0$ . Consider the FK-dynamics on  $G$  with boundary conditions  $\phi$  and  $\phi'$ , and let  $\alpha, \alpha'$  denote their log-Sobolev constants, respectively. Then,

$$q^{-5D(\phi, \phi')} \alpha' \leq \alpha \leq q^{5D(\phi, \phi')} \alpha'.$$

We now use the above to bound the rate of convergence to equilibrium on  $L$ -treelike balls of radius  $(\frac{1}{2} - \delta) \log_\gamma n$ .



**Lemma 6.6.** *Suppose  $G = (V, E)$  is  $L$ -Treelike. Let  $\xi$  be a  $K$ -Sparse boundary condition on  $G$ . For every  $p \in (0, 1)$  and  $q > 0$ , there exists  $\alpha_0(p, q, L, K) > 0$  (importantly, independent of  $G$ ) such that the log-Sobolev constant of the FK-dynamics on  $G$  with boundary condition  $\xi$  is at least  $\alpha_0$ .*

*Proof.* Observe first that the FK-dynamics on any tree with free boundary condition has log-Sobolev constant  $c_{p,q} = \Omega(1)$ . This follows from the observation that the random-cluster model on a tree with free boundary condition is simply the product measure, where every edge is open independently with probability  $\hat{p}$ , and the standard fact that the entropy tensorizes over product spaces; see, e.g., [1].

Now, let  $H \subset E$  be a set of at most  $L$  edges such that  $(V, E \setminus H)$  is a tree. Consider the tree  $\mathcal{T} = (V, E \setminus H)$  and let  $\phi$  be the boundary condition that includes all the connections from  $\xi$  and adds wirings between  $w$  and  $w'$  for every edge  $\{w, w'\} \in H$ . By Corollary 6.5, the log-Sobolev constant for the FK-dynamics on  $\mathcal{T}$  with boundary condition  $\phi$  is at least  $c_{p,q} \cdot q^{-5(K+L)}$ .

The FK-dynamics on  $G$  with boundary condition  $\phi$  is a product Markov chain on  $\{0, 1\}^{E \setminus H} \times \{0, 1\}^H$  with stationary distribution  $\pi_{\mathcal{T}}^{\phi} \otimes \prod_{i=1}^{|H|} \nu_i$ , where the  $\nu_i$ 's are independent  $\text{Ber}(p)$  distributions. Hence, it follows that the log-Sobolev constant of the FK-dynamics on  $G$  with boundary condition  $\phi$  is at least  $\hat{c}_{p,q} \cdot q^{-5(K+L)}$  for a suitable constant  $\hat{c}_{p,q} > 0$ . Finally, we note that by Corollary 6.5, the log-Sobolev constant on  $G$  with boundary conditions  $\xi$  (instead of  $\phi$ ) is at least  $\hat{c}_{p,q} q^{-5(K+L)-5L}$ .  $\square$

Combining the above, we arrive at the following bound on the rate of convergence of the FK-dynamics on treelike graphs with sparse boundary conditions.

**Lemma 6.7.** *Consider an  $L$ -Treelike graph  $G = (V, E)$  with a  $K$ -Sparse boundary condition  $\xi$ . For every  $p \in (0, 1)$  and  $q > 0$ , there exists  $\alpha_0 = \alpha_0(p, q, L, K) > 0$  such that*

$$\max_{x_0 \in \Omega} \|\mathbb{P}(X_t^{x_0} \in \cdot) - \pi_G^{\xi}\|_{\text{TV}} \leq \frac{1}{\sqrt{2}} e^{-\alpha_0 t} \left( \log \frac{1}{\min_{x \in \Omega} \pi_G^{\xi}(x)} \right)^{1/2}.$$

**Proof of Lemma 6.7.** This follows by combining Lemma 6.6 and Fact 6.2.  $\square$

## 7. PROOF OF MAIN THEOREM

Given the estimates proven in the preceding sections, we can now prove our main result, Theorem 1.6.

**7.1. Proof of main theorem: upper bound.** We begin with the proof of the upper bound.

**Proof of Theorem 1.6: upper bound.** Fix  $q > 1$ ,  $\gamma > 1$  and  $p < p_u(q, \gamma)$ . (It suffices to consider  $\gamma > 1$  since  $\lim_{\gamma \downarrow 1} p_u(q, \gamma) = 1$ , and if  $\gamma \geq \gamma'$ , then  $\mathcal{D}_{\gamma', \kappa} \subset \mathcal{D}_{\gamma, \kappa}$ .) Let  $R = (\frac{1}{2} - \delta) \log_{\gamma} n$ , where  $\delta > 0$  is a small constant we choose later. For  $K$  and  $L$  fixed positive constants,  $\varepsilon \in (0, 1/2)$  and  $t \geq 0$ , let  $\Gamma_t = \Gamma_t(L, K, \delta, \varepsilon, \gamma)$  be the subset of (multi)graphs on  $n$  vertices with degree sequence  $\mathbf{d}_n$  given by:

$$\Gamma_t = \{\mathcal{G} : \mathcal{G} \text{ is } (L, R)\text{-Treelike, has } (\gamma, \varepsilon)\text{-volume growth} \\ \text{and } \mathbb{P}(X_{\mathcal{G}, t}^1 \text{ is } (K, R)\text{-Sparse}) \geq 1 - n^{-5}\}.$$

By Lemmas 2.2 and 2.4, as well as Theorem 2.12, for every  $\delta \in (0, 1/2)$  and  $\varepsilon \in (0, 1/2)$ , there exist constants  $\kappa(p, q, \gamma, \delta)$ ,  $L(\delta)$ ,  $K(p, q, \gamma, \delta)$ , and  $T(p, q, \gamma)$  such that if  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$  then  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}(\Gamma_T^c) = o(1)$ . Hence, it suffices for us to prove that the mixing time of the FK-dynamics on any  $\mathcal{G} \in \Gamma_T$  is  $O(\log n)$ .

Fix any  $\mathcal{G} \in \Gamma_T$ . Let  $((X_t^{x_0})_{t \geq 0})_{x_0}$  be the family of FK-dynamics initialized from all possible configurations  $x_0$ , coupled via the standard grand coupling for the FK-dynamics; i.e., using the same clock rings and the same uniform random variables to make the edge updates while running the chain from different initializations. Recall that this coupling is monotone when  $q \geq 1$  so that for every  $t \geq 0$ , if  $X_t^{x_0} \leq X_t^{y_0}$ , then  $X_{t'}^{x_0} \leq X_{t'}^{y_0}$  for all  $t' \geq t$ . Using the standard fact that the coupling time provides a bound on the mixing time (see e.g., [48]), by a union bound over the edges, it suffices to show that under this grand coupling,

$$\mathbb{P}(X_T^1(e) \neq X_T^0(e)) \leq o(1/|E(\mathcal{G})|) \quad \text{for every } e \in E(\mathcal{G}). \quad (7.1)$$

Now fix any such  $e = \{u, v\}$  and for ease of notation, set  $B_v = E(B_R(v))$  and  $B_v^c = E(\mathcal{G}) \setminus B_v$ . Consider two auxiliary copies of the FK-dynamics  $Y_t^1$  and  $Y_t^0$  that censor (ignore) all updates on edges of  $B_v^c$  after time  $T$ . The censoring inequality from [55] applied to the FK-dynamics [36, Theorem 2.5] implies that  $Y_t^1 \geq X_t^1$  and  $Y_t^0 \leq X_t^0$  for all  $t \geq 0$  and thus

$$\mathbb{P}(X_t^1(e) \neq X_t^0(e)) \leq \mathbb{P}(X_t^1(e) = 1) - \mathbb{P}(X_t^0(e) = 1) \leq \mathbb{P}(Y_t^1(e) = 1) - \mathbb{P}(Y_t^0(e) = 1).$$

Let  $\mathcal{H}_v$  be the set of configurations on  $B_v^c$  such that the boundary conditions they induce on  $B_v$  are  $K$ -Sparse. (Here and throughout the paper, the boundary condition induced by a configuration  $\omega(B^c)$  on a set  $B$  wires two vertices  $w, w' \in V(B)$  if they are in the same connected component of  $\omega(B^c)$ .) By definition of  $\Gamma_T$  and monotonicity of the FK-dynamics, we have for every  $\mathcal{G} \in \Gamma_T$ ,

$$\mathbb{P}(Y_T^0(B_v^c) \notin \mathcal{H}_v) \leq \mathbb{P}(Y_T^1(B_v^c) \notin \mathcal{H}_v) \leq n^{-5}.$$

Therefore,  $\mathbb{P}(Y_t^1(e) = 1) - \mathbb{P}(Y_t^0(e) = 1)$  is bounded by

$$\max_{\phi^1, \phi^0 \in \mathcal{H}_v} \left[ \mathbb{P}(Y_t^1(e) = 1 \mid Y_T^1(B_v^c) = \phi^1) - \mathbb{P}(Y_t^0(e) = 1 \mid Y_T^0(B_v^c) = \phi^0) \right] + 2n^{-5}$$

Now fix any  $\phi^1, \phi^0 \in \mathcal{H}_v$ . From the triangle inequality, we have

$$\begin{aligned} \mathbb{P}(Y_{T+s}^1(e) = 1 \mid Y_T^1(B_v^c) = \phi^1) - \mathbb{P}(Y_{T+s}^0(e) = 1 \mid Y_T^0(B_v^c) = \phi^0) \\ \leq \left| \mathbb{P}(Y_{T+s}^1(e) = 1 \mid Y_T^1(B_v^c) = \phi^1) - \pi_{\mathcal{G}}(\omega(e) = 1 \mid \omega(B_v^c) = \phi^1) \right| \end{aligned} \quad (7.2)$$

$$+ \left| \pi_{\mathcal{G}}(\omega(e) = 1 \mid \omega(B_v^c) = \phi^1) - \pi_{\mathcal{G}}(\omega(e) = 1 \mid \omega(B_v^c) = \phi^0) \right| \quad (7.3)$$

$$+ \left| \mathbb{P}(Y_{T+s}^0(e) = 1 \mid Y_T^0(B_v^c) = \phi^0) - \pi_{\mathcal{G}}(\omega(e) = 1 \mid \omega(B_v^c) = \phi^0) \right|. \quad (7.4)$$

Observe that the chain  $(Y_{T+s}^1)_{s \geq 0}$  may be viewed as an FK-dynamics on  $B_v$  with the boundary condition induced by  $\phi^1$ , initialized from the (random) configuration  $Y_T^1(B_v)$  and with stationary distribution  $\pi_{\mathcal{G}}(\omega(B_v) \in \cdot \mid \omega(B_v^c) = \phi^1) = \pi_{B_v}^{\phi^1}$ ; the analogous statement is true for  $(Y_{T+s}^0)_{s \geq 0}$  and  $\pi_{B_v}^{\phi^0}$ .

Setting  $\hat{T} = T + \hat{S}_n$  where  $\hat{S}_n = \hat{C} \log n$  for a constant  $\hat{C}(p, q, \gamma, L, K)$  sufficiently large, since  $B_v$  is  $L$ -Treelike and  $\phi^1$  is  $K$ -Sparse, we obtain from Lemma 6.7 that

$$\left| \mathbb{P}(Y_{\hat{T}}^1(e) = 1 \mid Y_T^1(B_v^c) = \phi^1) - \pi_{\mathcal{G}}(\omega(e) = 1 \mid \omega(B_v^c) = \phi^1) \right| \leq n^{-5};$$

the same bound holds for (7.4).

Finally, since both  $\phi^1$  and  $\phi^0$  induce  $K$ -Sparse boundary conditions on  $B_v$  and  $\mathcal{G}$  is  $(L, R)$ -Treelike with  $(\gamma, \varepsilon)$ -volume growth, by Theorem 2.9 there exists  $C = C(p, q, L, K, \gamma) > 0$  such that (7.3) is at most

$$\left\| \pi_{B_v}^{\phi^1}(\omega(E_v) \in \cdot) - \pi_{B_v}^{\phi^0}(\omega(E_v) \in \cdot) \right\|_{\text{TV}} \leq C \hat{p}^{2(1-C\sqrt{\varepsilon})R} \leq C \hat{p}^{(1-2\delta)(1-C\sqrt{\varepsilon}) \log_{\gamma} n},$$

where  $E_v$  is the set of edges incident to  $v$ , and we used  $R = (\frac{1}{2} - \delta) \log_{\gamma} n$ . Setting  $\theta = (1 - 2\delta)(1 - C\sqrt{\varepsilon})$ ,

$$\left\| \pi_{B_v}^{\phi^1}(\omega(E_v) \in \cdot) - \pi_{B_v}^{\phi^0}(\omega(E_v) \in \cdot) \right\|_{\text{TV}} \leq C \hat{p}^{\theta \log_{\gamma} n} = C n^{-\theta(1 - \frac{1}{\log_{\hat{p}} \gamma})}. \quad (7.5)$$

Since  $\hat{p} < 1/\gamma$ ,  $\log_{\hat{p}} \gamma < 0$ , there is some  $c_{p, \gamma} > 0$  such that the right-hand side is  $C n^{-\theta(1+c_{p, \gamma})}$ . By taking  $\varepsilon, \delta$  sufficiently small,  $\theta$  can be made arbitrarily close to 1, so that (7.5) is  $o(1/n)$ .

Now notice that  $|E(\mathcal{G})| = O(n)$ . To see this, observe that by Jensen's inequality  $(\frac{1}{n} \sum_v d_v)^2 \leq \frac{1}{n} \sum_v d_v^2$ , and since  $(\mathbf{d}_n) \in \mathcal{D}_{\gamma, \kappa}$ , we also have  $\sum_v d_v^2 \leq (1 + \gamma) \sum_v d_v$ . Combining these two inequalities we find that  $|E(\mathcal{G})| \leq \frac{(1+\gamma)n}{2}$ . Therefore, each of (7.2)–(7.4) are  $o(1/|E(\mathcal{G})|)$ , implying (7.1) as desired.  $\square$

**7.2. Lower bound on the mixing time of FK-dynamics.** We now turn to proving the mixing time lower bound of Theorem 1.6. Though the argument is a straightforward adaptation of the proof of the lower bound in [4], given our results on  $(\gamma, \varepsilon)$ -growth of the random graph, and the exponential decay rate on random trees from Lemma 2.7, we include the proof for completeness, demonstrating that our new results give the requisite inputs to adapt the proof of [4].

**Claim 7.1.** Fix  $\varepsilon$  small. Suppose  $\kappa$  is sufficiently large and  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ . With  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ -probability  $1 - o(1)$ ,  $\mathcal{G}$  satisfies  $(\gamma, \varepsilon)$ -volume growth, and there exist  $n^{1/5}$  vertices whose balls of radius  $\frac{1}{5} \log_\gamma n$  are disjoint, and are trees.

*Proof.* On the one hand, by Lemma 2.4, with probability  $1 - o(1)$ ,  $\mathcal{G}$  satisfies  $(\gamma, \varepsilon)$ -volume growth, as long as  $\kappa$  is sufficiently large (depending on  $\varepsilon$ ). We prove the rest of the events have probability  $1 - o(1)$  by repeated application of the breadth-first revealing of Process 4.6. Namely, consider the procedure where we repeatedly take an arbitrary vertex  $v$  that has not been discovered yet, and reveal its ball of radius  $R = \frac{1}{5} \log_\gamma n$  via Process 4.6. Let  $v_i$  be the  $i$ 'th vertex to be selected in this procedure, and let  $\mathcal{A}_i$  be  $\bigcup_{j \leq i} E(B_R(v_j))$ . Then, for integer  $m \leq n$  the probability that one of  $(B_R(v_1), \dots, B_R(v_m))$  is not disjoint trees, is at most

$$\mathbb{P}_{\text{CM}(\mathbf{d}_n)} \left( \bigcup_{i=1}^m \{B_R(v_i) \cap \mathcal{A}_{i-1} = \emptyset \text{ or } B_R(v_i) \text{ is not a tree}\}, \mathcal{G} \in (\gamma, \varepsilon)\text{-volume growth} \mid \mathcal{A}_{i-1} \right).$$

Using the fact that  $\mathcal{G}$  is of  $\gamma, \varepsilon$ -volume growth that we are intersecting with, the event can be rewritten as in its first  $\gamma^R$  many matching attempts, none match with anything in  $\mathcal{A}_i$  or any half-edge belonging to a newly discovered half-edge of  $B_R(v_i)$ . In any one edge matching, uniformly over what has already been revealed, this probability is bounded by

$$\frac{\|\mathbf{d}_n\|_\infty m n^{1/5}}{\|\mathbf{d}_n\|_1 - \|\mathbf{d}_n\|_\infty m n^{1/5}},$$

which, for  $m = n^{1/5}$ , is at most  $n^{-1/2}$  as long as  $\kappa$  is sufficiently large, so that  $\varepsilon_*(\kappa) < 1/10$ . As there are at most  $n^{2/5}$  edges to match, the probability that no edge gets matched to an already discovered vertex, and thus all the revealed balls form disjoint trees, is at most  $\mathbb{P}(\text{Bin}(n^{2/5}, n^{-1/2}) > 0)$  which is  $o(1)$  simply by a Markov inequality.  $\square$

Fix  $\eta \in (0, 1/5)$  to be taken sufficiently small later. For every  $\mathcal{G}$  having  $n^{1/5}$  many vertices whose balls of radius  $\frac{1}{5} \log_\gamma n$  are disjoint trees, choose arbitrarily some  $n^\eta$  vertices amongst the  $n^{1/5}$  of Claim 7.1, and for each vertex collect a representative edge incident to it to form the set  $\mathcal{C} = \mathcal{C}_\eta(\mathcal{G})$ . Our proof will rely on a coupling of the restrictions of  $X_{t, \mathcal{G}}$  and  $\pi_{\mathcal{G}}$  to  $\mathcal{C}$  to  $\text{Ber}(\hat{p})$  product chains. For this, let:

- $X_t = X_{t, \mathcal{G}}$  be a realization of the FK-dynamics;
- $Y_t = Y_{t, \mathcal{G}}$  be a realization of the FK-dynamics that censors all updates in  $E(\mathcal{G}) \setminus \mathcal{C}$ ;
- $\nu$  as the product measure over  $|\mathcal{C}|$  many  $\text{Ber}(\hat{p})$  random variables.

As before, let  $Y_t^0$  be the chain  $Y_t$  initialized from the all-0 configuration.

**Lemma 7.2.** Let  $\mathcal{G}$  be any graph satisfying  $(\gamma, \varepsilon)$ -volume growth for  $\varepsilon < 1/6$ , and having at least  $n^{1/5}$  vertices whose balls of radius  $\frac{1}{5} \log_\gamma n$  are disjoint trees. For every  $q > 1$ , and  $p < p_u(q, \gamma)$ , there exists  $\eta > 0$  sufficiently small such that we have the following for  $\mathcal{C} = \mathcal{C}_\eta(\mathcal{G})$ :

(1) For all  $T = O(\log n)$ , for all  $t \leq T$ ,

$$\|P(X_t^0(\mathcal{C}) \in \cdot) - P(Y_t^0(\mathcal{C}) \in \cdot)\|_{\text{TV}} \leq o(1).$$

(2)  $\|\pi_{\mathcal{G}}(\omega(\mathcal{C}) \in \cdot) - \nu\|_{\text{TV}} \leq o(1)$ .

*Proof.* We start with part (1). Our aim is to show that under the grand coupling of  $X_t^0$  and  $Y_t^0$ , for every  $t \leq T = O(\log n)$ , we have  $\mathbb{P}(X_t^0 \neq Y_t^0) \leq o(1)$ . Under the grand coupling, let  $\mathcal{T}_T = (t_1, t_2, \dots, t_{s(T)})$  denote the sequence of times on which the updated edge is in  $\mathcal{C}$ , so that  $s(T)$  counts the number of updates in  $\mathcal{C}$  by time  $T$ . We can then bound

$$\mathbb{P}(X_t^0 \neq Y_t^0) \leq \mathbb{P}(s(T) > n^{2\eta}) + \mathbb{P}(X_t^0 \neq Y_t^0, s(T) \leq n^{2\eta}).$$

The first term on the right-hand side is at most the probability that  $\text{Pois}(T|\mathcal{C}|) \geq n^{2\eta}$  which is  $o(1)$  by standard tail estimates for Poisson variables. It thus suffices to work on the event  $s(T) \leq n^{2\eta}$ .

Let  $R := \frac{1}{6} \log_\gamma n$  and let  $Z_t$  be the FK-dynamics chain (coupled to  $X_t, Y_t$  through the grand coupling) that freezes the configuration on  $\mathcal{C} \cup (E(\mathcal{G}) \setminus \bigcup_{e \in \mathcal{C}} E(B_R(e)))$  to be all-1. Let  $Z_t^0$  be the chain  $Z_t$  initialized from the configuration that is all-0 on  $\bigcup_{e \in \mathcal{C}} E(B_R(e)) \setminus \{e\}$  (but all-1 on the frozen edges). Observe, trivially, that  $X_t^0 \leq Z_t^0$  for all  $t \geq 0$ . Also, observe that the updates of  $Z_t^0$  are stochastically dominated by Glauber updates on the union of  $2|\mathcal{C}|$  many  $d$ -ary trees  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  of depth  $R$ , rooted at the endpoints of the edges of  $\mathcal{C}$ , and each having  $(1, \mathcal{C})$  boundary conditions. By monotonicity of the FK-dynamics, for every  $t \geq 0$ ,

$$\mathbb{P}\left(Z_t^0\left(\bigcup_{e \in \mathcal{C}} \{E(B_R(e)) \setminus \{e\}\}\right) \in \cdot\right) \leq \bigotimes_{e \in \mathcal{C}} \bigotimes_{i \in \{1,2\}} \pi_{\mathcal{T}_{e,i}}^{(1,\mathcal{C})}. \quad (7.6)$$

For each time  $t_i \in \mathcal{T}_T$ , when an edge  $e_{t_i} \in \mathcal{C}$  is updated,  $Y_{t_i}^0(e_{t_i})$  is drawn from an independent  $\text{Ber}(\hat{p})$ . At the same time,  $X_{t_i}^0(e_{t_i})$  is drawn from  $\text{Ber}(\hat{p})$  if the endpoints of  $e_{t_i}$  are not connected in  $Z_{t_i}^0$ , which in turn must occur if none of  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  have an open root-to-leaf path in  $Z_{t_i}^0$ . We thus consider the probability of this event.

Since  $\mathcal{G}$  has  $(\gamma, \varepsilon)$ -volume growth for  $\varepsilon < 1/6$ , every tree among  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  has at most  $\gamma^R$  many leaves. Thus, by the stochastic domination of (7.6), and Lemma 2.7, the probability that the endpoints of  $e_{t_i}$  are connected in  $Z_{t_i}^0$  is at most  $2C(\hat{p}\gamma)^R$ , which for  $\eta$  sufficiently small is  $O(n^{-3\eta})$ . On the event that  $\{s(T) \leq n^{2\eta}\}$ , we can union bound the above probability over the  $s(T)$  times in  $\mathcal{T}_T$ , to find that  $\mathbb{P}(X_t^0 \neq Y_t^0, s(T) \leq n^{2\eta})$  is at most  $O(n^{-\eta}) = o(1)$  as desired.

For part (2), consider the  $2|\mathcal{C}|$  many  $d$ -ary trees  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  emanating from the endpoints of the edges of  $\mathcal{C}$ . Notice that if none of  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  have an open root-to-leaf path, then the values  $\omega(\mathcal{C})$  are conditionally distributed as a product of  $\text{Ber}(\hat{p})$  random variables, i.e.,  $\omega(\mathcal{C})$  would conditionally be distributed as  $\nu(A)$ .

As such, the total-variation distance  $\|\pi_{\mathcal{G}}(\omega(\mathcal{C}) \in \cdot) - \nu\|_{\text{TV}}$  is bounded by the  $\pi_{\mathcal{G}}$ -probability that one of  $(\mathcal{T}_{e,1}, \mathcal{T}_{e,2})_{e \in \mathcal{C}}$  has an open root-to-leaf path. By the stochastic domination

$$\pi_{\mathcal{G}}\left(\omega\left(\bigcup_{e \in \mathcal{C}} \mathcal{T}_{e,1} \cup \mathcal{T}_{e,2}\right) \in \cdot\right) \leq \bigotimes_{e \in \mathcal{C}} \bigotimes_{i \in \{1,2\}} \pi_{\mathcal{T}_{e,i}}^{(1,\mathcal{C})}.$$

By a union bound, the left-hand side above is then at most

$$\sum_{e \in \mathcal{C}} \sum_{i \in \{1,2\}} \pi_{\mathcal{T}_{e,i}}^{(1,\mathcal{C})}(e \leftrightarrow \partial \mathcal{T}_{e,i}),$$

which the  $(\gamma, \varepsilon)$ -volume growth condition and Lemma 2.7 together show is at most  $2n^\eta \cdot C(\hat{p}\gamma)^R$ . For  $\varepsilon$  sufficiently small (depending on  $p, q, \gamma$ ) this is  $o(1)$ .  $\square$

**Proof of Theorem 1.6: lower bound.** Take any  $n$ -vertex graph  $\mathcal{G}$  having  $(\gamma, \varepsilon)$ -volume growth for  $\varepsilon < 1/6$  and with  $n^{1/5}$  many vertices whose balls of radius  $\frac{1}{5} \log_\gamma n$  are disjoint trees. Note that by Claim 7.1, such graphs have  $\mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ -probability  $1 - o(1)$ . Take  $\eta$  sufficiently small per Lemma 7.2. Consider the event  $A^+ \subset \{0, 1\}^{\mathcal{C}}$  that at least  $\hat{p}n^\eta - n^{2\eta/3}$  of the edges in  $\mathcal{C}$  are open. Let  $(\bar{Y}_s)$  be the (discrete-time) product Markov chain over  $|\mathcal{C}| = n^\eta$  many i.i.d.  $\text{Ber}(\hat{p})$  random variables, coupled to  $Y_t(\mathcal{C})$  via  $\bar{Y}_{s(t)} = Y_t(\mathcal{C})$  for all  $t$ , where  $s(t)$  counts the number of updates in  $\mathcal{C}$  by time  $t$ . By item (1) of Lemma 7.2, for every  $T = O(\log n)$ ,

$$\begin{aligned} \mathbb{P}(X_T^0(\mathcal{C}) \in A^+) &\leq \mathbb{P}(s(T) > cn^\eta \log n) + \mathbb{P}(Y_T^0 \in A^+, s(T) \leq cn^\eta \log n) + o(1) \\ &\leq \mathbb{P}(s(T) > cn^\eta \log n) + \max_{s \leq cn^\eta \log n} \mathbb{P}(\bar{Y}_s^0 \in A^+) + o(1). \end{aligned}$$

(In the latter equation, we used the fact that the law of  $\bar{Y}_s^0$  only depends on the sequence of times  $(t_1, \dots, t_{s(T)})$  through the number of total updates  $s(T)$ .) Taking  $T := c^2 \log n$  for  $c > 0$  sufficiently small, the probability that  $s(T)$  is more than  $cn^\eta \log n$  is  $o(1)$  by tail bounds of a Poisson random variable with rate

$T|\mathcal{C}| = c^2 n^\eta \log n$ . Turning to the middle term above, by the standard coupon collector bound, for every  $c > 0$  sufficiently small,  $\sup_{s \leq cn^\eta \log n} \mathbb{P}(\bar{Y}_s^0 \in A^+) \leq o(1)$ .

Combining the above, we obtain

$$\mathbb{P}(X_T^0(\mathcal{C}) \in A^+) = o(1).$$

At the same time, by a Chernoff bound,  $\nu(A^+) = 1 - o(1)$  and by item (2) of Lemma 7.2, then,  $\pi_{\mathcal{G}}(A^+) = 1 - o(1)$ . These two together imply that the (continuous-time) mixing time is at least  $T = \Omega(\log n)$  as claimed.  $\square$

## 8. HIGH-DEGREE VERTICES SLOW DOWN MIXING FOR POTTS GLAUBER DYNAMICS

Our lower bound on the mixing time of the Glauber dynamics for the Potts model in a random graph is derived from a bottleneck argument. For the special case of the Erdős–Rényi random graph, the slow down can be attributed to *isolated* stars whose central vertex has degree  $\Theta(\frac{\log n}{\log \log n})$ . Such a star appears in the random graph with high probability, and since it is disconnected from the rest of  $\mathcal{G}$ , the mixing time on the star serves as a lower bound for the mixing time on the full graph. This straightforwardly gives a lower bound of  $n^{1+\Theta(\frac{1}{\log \log n})}$  on the discrete-time mixing time of the Glauber dynamics; see [53, Proposition 1.8].

For more general degree sequences, especially when there exist vertices of degree  $\omega(\log n)$ , the neighborhoods of the high-degree vertices will *not* be isolated from the remainder of the graph, and in fact will correspond to the denser parts of the random graph. We use the exponential decay of random-cluster connectivities when  $p < p_u(q, \gamma)$  to still leverage this star structure to give a lower bound on the mixing time of the Potts Glauber dynamics on a random graph that are exponential in its largest degree.

We will work with the discrete-time Potts Glauber dynamics, which at each step selects a vertex  $v \in V$  uniformly at random, and resamples its spin  $\sigma_v$  according to the following conditional distribution:

$$\mu_{G, \beta, q}(\sigma_v = i \mid \sigma(V \setminus \{v\})) = \frac{e^{\beta \sum_{(v,w) \in E} \mathbf{1}_{\{\sigma_w = i\}}}}{\sum_{i=1}^q e^{\beta \sum_{(v,w) \in E} \mathbf{1}_{\{\sigma_w = i\}}}}, \quad \text{for } i = 1, \dots, q.$$

**Proof of Theorem 1.7.** Let  $v_\star$  be a vertex in  $\mathcal{G}$  of maximum degree, and let  $m_i(\sigma)$  denote the number of vertices adjacent to  $v_\star$  that are assigned spin  $i$  in configuration  $\sigma$ . Define the following bottleneck set:

$$\mathcal{A}_\varepsilon := \left\{ \sigma : \sigma_{v_\star} = 1, m_1(\sigma) - \max_{j \neq 1} m_j(\sigma) \geq \lfloor \varepsilon d_{v_\star} \rfloor \right\}.$$

Our aim is to show that  $\mathcal{A}_\varepsilon$  is a set of small conductance. Namely, we wish to show that there exists  $\varepsilon > 0$  such that

$$\Phi(\mathcal{A}_\varepsilon) = \frac{Q(\mathcal{A}_\varepsilon, \mathcal{A}_\varepsilon^c)}{\mu(\mathcal{A}_\varepsilon)\mu(\mathcal{A}_\varepsilon^c)} \leq e^{-\Omega(d_{v_\star})},$$

where  $Q(\mathcal{A}_\varepsilon, \mathcal{A}_\varepsilon^c) = \sum_{\sigma \in \mathcal{A}_\varepsilon, \sigma' \in \mathcal{A}_\varepsilon^c} \mu(\sigma) P(\sigma, \sigma')$  with  $P$  denoting the transition matrix of the discrete-time Glauber dynamics.

For this, notice that we can expand  $Q(\mathcal{A}_\varepsilon, \mathcal{A}_\varepsilon^c)$  into its contribution from transitions that exit  $\mathcal{A}_\varepsilon$  by flipping the spin of  $\sigma_{v_\star}$ , and those that exit  $\mathcal{A}_\varepsilon$  by flipping the spin of a neighbor of  $v_\star$  in the configuration. Hence, let

$$\hat{\mathcal{A}}_\varepsilon := \left\{ \sigma \in \mathcal{A}_\varepsilon : m_1(\sigma) - \max_{j \neq 1} m_j(\sigma) = \lfloor \varepsilon d_{v_\star} \rfloor \right\}.$$

Namely, we can bound

$$\begin{aligned} \Phi(\mathcal{A}_\varepsilon) &\leq \frac{\sum_{\sigma \in \mathcal{A}_\varepsilon} \sum_{j=2}^q \mu(\sigma) P(\sigma, \sigma^{v_\star \rightarrow j})}{\mu(\mathcal{A}_\varepsilon)\mu(\mathcal{A}_\varepsilon^c)} + \frac{\sum_{\sigma \in \hat{\mathcal{A}}_\varepsilon, \sigma' \in \mathcal{A}_\varepsilon^c} \mu(\sigma) P(\sigma, \sigma')}{\mu(\mathcal{A}_\varepsilon)\mu(\mathcal{A}_\varepsilon^c)} \\ &\leq \frac{\max_{\sigma \in \mathcal{A}_\varepsilon, j \neq 1} P(\sigma, \sigma^{v_\star \rightarrow j})}{\mu(\mathcal{A}_\varepsilon^c)} + \frac{\mu(\hat{\mathcal{A}}_\varepsilon) \max_{\sigma \in \hat{\mathcal{A}}_\varepsilon} P(\sigma, \mathcal{A}_\varepsilon^c)}{\mu(\mathcal{A}_\varepsilon)\mu(\mathcal{A}_\varepsilon^c)}, \end{aligned} \quad (8.1)$$

where  $\sigma^{v_\star \rightarrow j}$  is the configuration which agrees with  $\sigma$  everywhere except on  $v_\star$  where it takes spin  $j$ . Observe first of all, that by the spin symmetry of the model Potts model,  $\mu(\mathcal{A}_\varepsilon) \leq 1/q$  and thus  $\mu(\mathcal{A}_\varepsilon^c) \geq \frac{q-1}{q} \geq \frac{1}{2}$ . Moreover, by the definition of the Glauber dynamics, the transition matrix  $P$  satisfies

$$\max_{\sigma \in \mathcal{A}_\varepsilon} P(\sigma, \sigma^{v_\star \rightarrow j}) = \frac{1}{n} \cdot \frac{e^{\beta m_j(\sigma)}}{\sum_j e^{\beta m_j(\sigma)}} \leq \frac{e^{\beta(m_j(\sigma) - m_1(\sigma))}}{n} \leq \frac{e^{-\beta \varepsilon d_{v_\star}}}{n}.$$

Also, for every  $\sigma \in \hat{\mathcal{A}}_\varepsilon$ , it satisfies

$$\max_{\sigma \in \hat{\mathcal{A}}_\varepsilon} P(\sigma, \mathcal{A}_\varepsilon^c) \leq \frac{d_{v_\star}}{n},$$

as one needs to select a neighbor of  $v_\star$  to update in order to move from  $\sigma \in \hat{\mathcal{A}}_\varepsilon$  to  $\mathcal{A}_\varepsilon^c$ . As such,

$$\Phi(\mathcal{A}_\varepsilon) \leq \frac{2}{n} e^{-\beta \varepsilon d_{v_\star}} + \frac{2d_{v_\star}}{n} \frac{\mu(\hat{\mathcal{A}}_\varepsilon)}{\mu(\mathcal{A}_\varepsilon)}. \quad (8.2)$$

It remains to bound the ratio of the probabilities of the events  $\hat{\mathcal{A}}_\varepsilon$  to  $\mathcal{A}_\varepsilon$ . It will be convenient to work with the random-cluster representation of the Potts model. Let

$$\mathcal{A}_\varepsilon^{\text{RC}} := \{\omega \in \{0, 1\}^{E(\mathcal{G})} : |\{e \in E_{v_\star} : \omega(e) = 1\}| \geq \varepsilon d_{v_\star} \text{ and } |\mathfrak{V}_{E_{v_\star}}(\omega)| \leq \varepsilon d_{v_\star}/2\},$$

where we recall that  $E_{v_\star}$  is the set of edges incident to  $v_\star$  and  $\mathfrak{V}_{E_{v_\star}}(\omega)$  is the set of neighbors of  $v_\star$  in non-trivial connected components in the configuration induced by  $\omega(E(\mathcal{G}) \setminus E_{v_\star})$ . In words this is the event that an  $\varepsilon$  fraction of the edges incident to  $v_\star$  are open, and at most  $\varepsilon d_{v_\star}/2$  of the neighbors of  $v_\star$  are connected to one another in the configuration outside the immediate neighborhood of  $v_\star$ .

We first note that for some  $\varepsilon(p, q, \gamma) > 0$ , with high probability under the random graph, the event  $\mathcal{A}_\varepsilon^{\text{RC}}$  has high probability under the random-cluster measure  $\pi$ . For this, observe that since  $\pi$  stochastically dominates the independent edge percolation measure with edge probability  $\hat{p}$ , and by a Chernoff bound, for any  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$

$$\pi_{\mathcal{G}}(|\{e \in E_{v_\star} : \omega(e) = 1\}| < \varepsilon d_{v_\star}) \leq \mathbb{P}(\text{Bin}(d_{v_\star}, \hat{p}) < \varepsilon d_{v_\star}) \leq e^{-\Omega(\hat{p} d_{v_\star})},$$

for  $\varepsilon$  sufficiently small (say, less than  $\hat{p}/2$ ). By Lemma 5.22, if  $\kappa$  is sufficiently large and  $(\mathbf{d}_n)_n \in \mathcal{D}_{\gamma, \kappa}$ , for every  $\varepsilon > 0$ , we have with probability  $1 - o(1)$  over the graph  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$ ,

$$\pi_{\mathcal{G}}(|\mathfrak{V}_{E_{v_\star}}(\omega)| > \varepsilon d_{v_\star}/2) \leq e^{-\Omega(\varepsilon d_{v_\star})}.$$

Hence, it follows from a union bound that there exists  $\varepsilon(p, q, \gamma)$  small, such that with probability  $1 - o(1)$ ,  $\mathcal{G} \sim \mathbb{P}_{\text{CM}(\mathbf{d}_n)}$  is such that

$$\pi_{\mathcal{G}}(\mathcal{A}_\varepsilon^{\text{RC}}) \geq 1 - e^{-\Omega(\varepsilon d_{v_\star})}.$$

As such, as long as  $\varepsilon > 0$  is sufficiently small, we can bound

$$\frac{\mu(\hat{\mathcal{A}}_\varepsilon)}{\mu(\mathcal{A}_\varepsilon)} \leq \frac{\mathbb{P}_{(\mu, \pi)}(\hat{\mathcal{A}}_\varepsilon \mid \mathcal{A}_{4\varepsilon}^{\text{RC}}) + e^{-\Omega(\varepsilon d_{v_\star})}}{\mathbb{P}_{(\mu, \pi)}(\mathcal{A}_\varepsilon \mid \mathcal{A}_{4\varepsilon}^{\text{RC}})(1 - e^{-\Omega(\varepsilon d_{v_\star})})},$$

where  $\mathbb{P}_{(\mu, \pi)}$  denotes the joint Edwards–Sokal distribution over spin-edge configurations; see [24, 38].

Now, consider a random-cluster configuration in  $\mathcal{A}_{4\varepsilon}^{\text{RC}}$ . Fixing a random-cluster configuration  $\omega$  in  $\mathcal{A}_{4\varepsilon}^{\text{RC}}$ , we claim that the probability of  $\mathcal{A}_\varepsilon$  given  $\omega$  is at least the probability of the following event  $\Gamma_\varepsilon$ , that

- (1) the component  $\mathcal{C}_{v_\star}(\omega)$  is given state 1; and
- (2) amongst the vertices of

$$\mathfrak{V}_\star^c := V(E_{v_\star}) \setminus (\mathcal{C}_{v_\star}(\omega) \cup \mathfrak{V}_{E_{v_\star}}(\omega)),$$

the number of vertices in each state in  $[q]$  is within  $\varepsilon d_{v_\star}/2$  of  $|\mathfrak{V}_\star^c|/q$ .



To see this, note that on  $\Gamma_\varepsilon$ , since  $\mathcal{C}_{v_\star}(\omega)$  has size at least  $4\varepsilon d_{v_\star}$  and  $|\mathfrak{V}_{E_{v_\star}}(\omega)| \leq 2\varepsilon d_{v_\star}$ , no matter which state the vertices of  $\mathfrak{V}_{E_{v_\star}}(\omega)$  take,  $\sigma$  will be such that

$$m_1(\sigma) - \max_{j \neq 1} m_j(\sigma) > (4\varepsilon - 2\varepsilon - \varepsilon)d_{v_\star} = \varepsilon d_{v_\star}.$$

(Here, the  $4\varepsilon$  comes from the sites in  $\mathcal{C}_{v_\star}(\omega)$ , the  $-2\varepsilon$  comes from a worst-possible assignment of states to sites of  $\mathfrak{V}_{E_{v_\star}}(\omega)$ , and the  $-\varepsilon$  comes from the maximal bias on the sites in  $\mathfrak{V}_\star^c$ .)

The probability of the event  $\Gamma_\varepsilon$ , when coloring the components of  $\omega$  independently, uniformly at random, is at least  $1/q$  (for the probability of coloring  $\mathcal{C}_{v_\star}(\omega)$  in state 1) times

$$1 - q\mathbb{P}\left(\left|\text{Bin}(|\mathfrak{V}_\star^c|, 1/q) - |\mathfrak{V}_\star^c|/q\right| > \varepsilon d_{v_\star}/2\right) \geq 1 - e^{-\Omega(\varepsilon d_{v_\star})}.$$

(Here, we used a union bound over the  $q$  different states, and a Chernoff bound.) In particular, we find that for  $\varepsilon(p, q, \gamma) > 0$  sufficiently small,

$$\mathbb{P}_{(\mu, \pi)}(\mathcal{A}_\varepsilon \mid \mathcal{A}_{4\varepsilon}^{\text{RC}}) \geq \min_{\omega \in \mathcal{A}_{4\varepsilon}^{\text{RC}}} \mathbb{P}(\Gamma_\varepsilon \mid \omega) \geq \frac{1}{q} \left(1 - e^{-\Omega(\varepsilon d_{v_\star})}\right).$$

On the other hand, the probability of  $\hat{\mathcal{A}}_\varepsilon$ , conditionally on  $\mathcal{A}_{4\varepsilon}^{\text{RC}}$  is bounded by the probability of the colorings of  $\mathfrak{V}_\star^c$  assigning at least  $2\varepsilon d_{v_\star} + |\mathfrak{V}_\star^c|/q$  many of its vertices to some state  $j \neq 1$ . By a union bound over the  $q$  states, and a Chernoff bound, this has probability at most

$$q\mathbb{P}\left(\left|\text{Bin}(|\mathfrak{V}_\star^c|, 1/q) - |\mathfrak{V}_\star^c|/q\right| > 2\varepsilon d_{v_\star}\right) \leq e^{-\Omega(\varepsilon d_{v_\star})}.$$

At this point, we can plug the above bounds into (8.2) to deduce that for all  $\varepsilon(p, q, \gamma) > 0$  sufficiently small,

$$\Phi(\mathcal{A}_\varepsilon) \leq \frac{1}{n} e^{-\Omega(\beta \varepsilon d_{v_\star})}.$$

(Notice that  $\varepsilon$  sufficiently small, needed to scale as  $\Theta(1/p)$ , so that this is  $n^{-1}e^{-\Omega(\beta^2 d_{v_\star})}$  for small  $\beta$ .) Relying on the classical Cheeger bound (see e.g., [48, Theorem 7.4]), the inverse of  $\Phi(\mathcal{A}_\varepsilon)$  serves as a lower bound on the mixing time of the Glauber dynamics for the Potts model.  $\square$

## REFERENCES

- [1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10. Société mathématique de France Paris, 2000.
- [2] A. Blanca, Z. Chen, D. Štefankovič, and E. Vigoda. The Swendsen-Wang dynamics on trees. In *Proceedings of the 25th International Workshop on Randomization and Computation (RANDOM)*, 2021.
- [3] A. Blanca, A. Galanis, L. Goldberg, D. Štefankovič, E. Vigoda, and K. Yang. Sampling in uniqueness from the Potts and random-cluster models on random regular graphs. In *Proceedings of the 22nd International Workshop on Randomization and Computation (RANDOM)*, 2018.
- [4] A. Blanca and R. Gheissari. Random-cluster dynamics on random regular graphs in tree uniqueness. *Communications in Mathematical Physics*, 2021.
- [5] A. Blanca, R. Gheissari, and E. Vigoda. Random-cluster dynamics in  $\mathbb{Z}^2$ : Rapid mixing with general boundary conditions. *Ann. Appl. Probab.*, 30(1):418–459, 02 2020.
- [6] A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 528–543, 2015.
- [7] A. Blanca and A. Sinclair. Random-cluster dynamics in  $\mathbb{Z}^2$ . *Probab. Theory Related Fields*, 2016. Extended abstract appeared in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), pp. 498–513.
- [8] A. Blanca, A. Sinclair, and X. Zhang. The critical mean-field Chayes-Machta dynamics. In *Proceedings of the 25th International Workshop on Randomization and Computation (RANDOM)*, 2021.
- [9] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.
- [10] B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.
- [11] M. Bordewich, C. Greenhill, and V. Patel. Mixing of the glauber dynamics for the ferromagnetic potts model. *Random Structures & Algorithms*, 48(1):21–52, 2016.

- [12] C. Borgs, J. Chayes, T. Helmuth, W. Perkins, and P. Tetali. Efficient sampling and counting algorithms for the Potts model on  $Z^d$  at all temperatures. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 738–751, New York, NY, USA, 2020. Association for Computing Machinery.
- [13] C. Borgs, J. T. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. H. Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *Proc. of the 40th Annual Symposium on Foundations of Computer Science (FOCS 1999)*, pages 218–229, 1999.
- [14] C. Borgs, J. T. Chayes, and P. Tetali. Tight bounds for mixing of the Swendsen-Wang algorithm at the Potts transition point. *Probab. Theory Related Fields*, 152(3-4):509–557, 2012.
- [15] L. Chayes and J. Machta. Graphical representations and cluster algorithms I. Discrete spin systems. *Physica A: Statistical Mechanics and its Applications*, 239(4):542–601, 1997.
- [16] X. Chen, W. Feng, Y. Yin, and X. Zhang. Rapid mixing of Glauber dynamics via spectral independence for all degrees. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 137–148. IEEE, 2022.
- [17] Y. S. Chow and H. Teicher. *Probability theory: independence, interchangeability, martingales*. Springer Science & Business Media, 2003.
- [18] A. Coja-Oghlan, A. Galanis, L. A. Goldberg, J. B. Ravelomanana, D. Štefankovič, and E. Vigoda. Metastability of the Potts Ferromagnet on Random Regular Graphs. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 45:1–45:20, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [19] P. Cuff, J. Ding, O. Loundor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012.
- [20] A. Dembo, A. Montanari, A. Sly, and N. Sun. The replica symmetric solution for Potts models on  $d$ -regular graphs. *Communications in Mathematical Physics*, 327(2):551–575, 2014.
- [21] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 08 1996.
- [22] M. Dyer, L. A. Goldberg, and M. Jerrum. Dobrushin conditions and systematic scan. *Combinatorics, Probability and Computing*, 17(6):761–779, 2008.
- [23] M. Dyer, L. A. Goldberg, and M. Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71–107, 2009.
- [24] R. G. Edwards and A. D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D* (3), 38(6):2009–2012, 1988.
- [25] G. Ellison. Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pages 1047–1071, 1993.
- [26] J. Felsenstein. *Inferring phylogenies*, volume 2. Sinauer Associates, Inc., Sunderland, MA, 2004.
- [27] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica*, 57:536–564, 1972.
- [28] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.
- [29] A. Galanis, L. A. Goldberg, and J. Stewart. Fast mixing via polymers for random graphs with unbounded degree. *Information and Computation*, page 104894, 2022.
- [30] A. Galanis, D. Štefanković, and E. Vigoda. Swendsen-Wang Algorithm on the Mean-Field Potts Model. In *Proc. of the 19th International Workshop on Randomization and Computation (RANDOM 2015)*, pages 815–828, 2015.
- [31] A. Galanis, D. Štefanković, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016.
- [32] S. Ganguly and I. Seo. Information percolation and cutoff for the random-cluster model. *Random Structures & Algorithms*, 57(3):770–822, 2020.
- [33] S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*, volume 1, pages 1496–1517. Berkeley, CA, 1986.
- [34] H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.
- [35] R. Gheissari and E. Lubetzky. Mixing times of critical two-dimensional Potts models. *Comm. Pure Appl. Math.*, 71(5):994–1046, 2018.
- [36] R. Gheissari and E. Lubetzky. Quasi-polynomial mixing of critical two-dimensional random cluster models. *Random Structures and Algorithms*, 2019.
- [37] R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean-field Swendsen–Wang dynamics. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):68 – 86, 2020.
- [38] G. Grimmett. The random-cluster model. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 73–123. Springer, Berlin, 2004.
- [39] H. Guo and M. Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1818–1827, 2017.
- [40] O. Häggström. The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104(2):231–253, 1996.

- [41] M. Harel and Y. Spinka. Finitary codings for the random-cluster model and other infinite-range monotone models. *Electronic Journal of Probability*, 27:1–32, 2022.
- [42] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 39–46. IEEE, 2006.
- [43] T. Helmuth, M. Jenssen, and W. Perkins. Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs, 2020.
- [44] J. Holm, K. De Lichtenberg, and M. Thorup. Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *Journal of the ACM (JACM)*, 48(4):723–760, 2001.
- [45] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, 18(1-2):205–225, 2009.
- [46] J. Jonasson. The random cluster model on a general graph and a phase transition characterization of nonamenability. *Stochastic Processes and their Applications*, 79(2):335–354, 1999.
- [47] J. H. Kim. Poisson cloning model for random graphs. In *International Congress of Mathematicians (ICM)*, 2006.
- [48] D. A. Levin and Y. Peres. Markov chains and mixing times (second edition). *The Mathematical Intelligencer*, 41(1):90–91, 2019.
- [49] Y. Long, A. Nachmias, W. Ning, and Y. Peres. A power law of order  $1/4$  for critical mean field Swendsen-Wang dynamics. *Mem. Amer. Math. Soc.*, 232(1092), 2014.
- [50] R. Lyons. The Ising model and percolation on trees and tree-like graphs. *Communications in Mathematical Physics*, 125(2):337 – 353, 1989.
- [51] A. Montanari and A. Saberi. The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, 107(47):20196–20201, 2010.
- [52] E. Mossel and A. Sly. Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Structures & Algorithms*, 35(2):250–270, 2009.
- [53] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *Ann. Probab.*, 41(1):294–328, 01 2013.
- [54] S. Osindero and G. Hinton. Modeling image patches with a directed hierarchy of Markov random fields. In *Advances in neural information processing systems*, pages 1121–1128, 2008.
- [55] Y. Peres and P. Winkler. Can extra updates delay mixing? *Communications in Mathematical Physics*, 323(3):1007–1016, 2013.
- [56] S. Roth and M. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 860–867, 2005.
- [57] L. Saloff-Coste. *Lectures on finite Markov chains*, pages 301–413. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [58] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, Jan 1987.
- [59] M. Thorup. Near-optimal fully-dynamic graph connectivity. In *Proceedings of the 32nd Annual ACM symposium on Theory of computing (STOC)*, pages 343–350, 2000.
- [60] M. Ullrich. Swendsen–Wang is faster than single-bond dynamics. *SIAM Journal on Discrete Mathematics*, 28(1):37–48, 2014.

A. BLANCA

DEPARTMENT OF CSE, PENNSYLVANIA STATE UNIVERSITY

*Email address:* ablanca@cse.psu.edu

R. GHEISSARI

DEPARTMENT OF STATISTICS AND EECS, UC BERKELEY

*Email address:* gheissari@berkeley.edu