# Random-cluster dynamics

*Antonio Blanca-Pimentel*

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2016-141
http://www.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-141.html

August 11, 2016

**Random-cluster Dynamics**

by

Antonio Blanca-Pimentel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alistair Sinclair, Chair
Associate Professor Prasad Raghavendra
Associate Professor Allan Sly

Summer 2016

**Random-cluster Dynamics**

## Abstract

Random-cluster Dynamics

by

Antonio Blanca-Pimentel

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Alistair Sinclair, Chair

The *random-cluster model* has been widely studied as a unifying framework for spin systems, random graphs and electrical networks. The random-cluster model on a graph $G = (V, E)$ with parameters $p \in (0, 1)$ and $q > 0$ assigns to each subgraph $(V, A \subseteq E)$ a probability proportional to $p^{|A|}(1-p)^{|E|-|A|}q^{c(A)}$, where $c(A)$ is the number of connected components in $(V, A)$. When $q = 1$ this model corresponds to the standard *bond percolation model.* For integer $q \geq 2$ the random-cluster model is closely related to the classical ferromagnetic *q-state Ising/Potts model.* When $q \to 0$, the set of weak limits that arise contains the uniform measures over the spanning trees, spanning forests and connected subgraphs of $G$.

In this thesis we investigate the dynamics of the random-cluster model. While dynamics for the Ising/Potts model have been widely studied, random-cluster dynamics have so far largely resisted analysis. We focus on two canonical cases: the case when $G$ is the complete graph on $n$ vertices, known as the *mean-field model*, and the case when $G$ is the infinite 2-dimensional lattice graph $\mathbb{Z}^2$. Mean-field models have historically proven to be a useful starting point in understanding dynamics on more general graphs. In statistical mechanics, however, probabilistic models are most frequently studied in the setting of infinite lattice graphs; understanding random-cluster dynamics in $\mathbb{Z}^2$ is thus of foremost importance.

In the first part of this thesis we establish the mixing time of Chayes-Machta dynamics in the the mean-field case. For $q \in (1, 2]$ we prove that the mixing time is $\Theta(\log n)$ for all $p \neq p_c(q)$, where $p = p_c(q)$ is the critical value corresponding to the emergence of a "giant" component. For $q > 2$, we identify a critical window $(p_s, p_S)$ of the parameter $p$ around $p_c(q)$ in which the dynamics undergoes an exponential slowdown. Namely, we prove that the mixing time is $\Theta(\log n)$ when $p \notin [p_s, p_S]$ and $\exp(\Omega(\sqrt{n}))$ when $p \in (p_s, p_S)$. We also show that the mixing time is $\Theta(n^{1/3})$ for $p = p_s$ and $\Theta(\log n)$ for $p = p_S$. In addition, we prove that the Glauber dynamics undergoes a similar exponential slowdown in $(p_s, p_S)$.

The second part of this thesis focuses on the analysis of the Glauber dynamics of the random-cluster model in the case where the underlying graph is an $n \times n$ box in the Cartesian lattice $\mathbb{Z}^2$. Our main result is a tight $\Theta(n^2 \log n)$ bound for the mixing time at all values of the model parameter $p$ except the critical point $p = p_c(q)$, and for all values of $q \geq 1$.

*To Lisdy...*

# Contents

# List of Figures

## Acknowledgments

First and foremost I would like to thank my advisor Alistair Sinclair. During the past five years, Alistair has been a reliable source of ideas, guidance and knowledge. Our many conversations, technical or otherwise, have been most illuminating.

I was very fortunate to split my “desk-time” at Berkeley between the sixth floor of Soda Hall and the Simons Institute. I am very grateful to all the students, postdocs and faculty in the Theory group for creating a relaxed and stimulating atmosphere for research. For this, I must thank Nima Anari, Frank Ban, Anand Bhaskar, Brielin Brown, Jonah Brown-Cohen, Siu Man Chan, Siu On Chan, Paul Christiano, Lynn Chua, James Cook, Anindya De, Rafael Frongillo, Rishi Gupta, Fotis Iliopoulos, Jingcheng Liu, Pasin Manurangsi, Peihan Miao, George Pierrakos, Anupam Prakash, Alex Psomas, Aviad Rubinstein, Manuel Sabin, Aaron Schild, Tselil Schramm, Jarett Schwartz, Akshayaram Srinivasan, Isabelle Stanton, Greg Valiant, Thomas Vidick, Di Wang, Tom Watson, Benjamin Weitz, Chris Wilkens and Sam Wong (whose initials appear so many times in this thesis). I would like to specially thank Urmila Mahadev and Piyush Srivastava, for many thought-provoking conservations, and Seung Woo Shin for making my time at Berkeley such much more enjoyable.

Naturally, the theory faculty also contributes their share to the great atmosphere in the Theory group. Thanks to Sanjam Garg, Elchanan Mossel, Christos Papadimitriou, Prasad Raghavendra, Satish Rao, Alistair Sinclair, Nikhil Srivastava, Allan Sly and Umesh Vazirani, for many interesting courses, seminars and memorable talks. I would also like to thank Prasad, Satish, Alistair, Allan and Umesh for serving on my prelim, qual and/or dissertation committees.

I also had the opportunity to collaborate with several visitors to the Simons Insitute, especially during the Counting Complexity and Phase Transitions program. I would like to thank Raimundo Briceño, Prieto Caputo, Andreas Galanis, Tyler Helmuth, Fabio Martinelli, Yuval Peres, Will Perkins, Eric Vigoda and Daniel Štefankovič for many enlightening discussions.

This thesis would have not been possible without the unconditional love and support from family. My lovely wife Lisdy helped me keep the right balance between research and all the other interesting things happening in the world. My parents' courageous decision to leave Cuba provided my sister and I with fantastic opportunities, educational and otherwise; mami, papi, for this and so much more I am very grateful. For as far as I can remember I have been constantly learning from my grandparents; Bertica, Llilli, Abuelo, many years later and you are still going strong in that regard. I would also like to thank my sister Laura, and my acquired brother Roly, for being my alter egos, and Jett, Tio, Miya, Aimee, David and Tony for their encouragement and affection during the course of my studies.

Unconditional love and support I also receive from my friends. Thanks to Gabriel, Hari, Edileo, Ari, Nelson, Yeny, Ernesto, Lois and Lucian for making this city Cuban enough. Special thanks to Gabriel for providing “back-saving” furniture that allow me to finish up writing this thesis. I would also like to thank my longtime friends Ale, Artur, Daniel, David, Ere, Karel, Katy, Lini, Pablo, Pachy, Peter and Reniel, who spread through three continents still find a way of being close. Andrés and Javier, thanks for being amazing role models and for all I have learned from you.

# Chapter 1

# Introduction

The *random-cluster model* was introduced in the late 1960s by Fortuin and Kasteleyn [20] as a unifying framework for studying random graphs, spin systems in physics and electrical networks. The random-cluster model on a finite graph $G = (V, E)$ with parameters $p \in (0, 1)$ and $q > 0$ assigns to each subgraph $(V, A \subseteq E)$ a probability

$$\mu_{G,p,q}(A) = \frac{p^{|A|}(1-p)^{|E|-|A|}q^{c(A)}}{Z_{G,p,q}}, \tag{1.1}$$

where $c(A)$ is the number of connected components in $(V, A)$ and $Z_{G,p,q}$ is the normalizing constant called the *partition function.*

When $q = 1$ this model corresponds to the standard *bond percolation model* on subgraphs of $G$, where each edge appears independently with probability $p$. For integer $q \geq 2$ the random-cluster model is, in a precise sense, dual to the classical ferromagnetic *q-state Potts model.* Configurations in the $q$-state Potts model are assignments of spin values $\{1, \ldots, q\}$ to the vertices of $G$. Each configuration $\sigma \in \{1, \ldots, q\}^{|V|}$ is assigned a probability

$$\pi_{G,\beta,q}(\sigma) = \frac{e^{\beta \sum_{(u,v)\in E} \mathbb{1}(\sigma(u)=\sigma(v))}}{Z_{G,\beta,q}},$$

where $\beta$ is a model parameter associated with the inverse temperature of the system and $Z_{G,\beta,q}$ is the appropriate normalizing constant. The $q = 2$ case corresponds to the *Ising model.* If $\beta$ is set equal to $-\ln(1-p) > 0$, then it is straightforward to check that correlations in the Ising/Potts model correspond to paths in the random-cluster setting. Consequently, the random-cluster model has allowed the use of sophisticated methods from stochastic geometry, developed in the study of bond percolation, in the context of spin systems. This has illuminated much of the physical theory of the ferromagnetic Ising/Potts model. However, the random-cluster model is not a "spin system" in the usual sense: in particular, the probability that an edge $e$ belongs to $A$ does not depend only on the dispositions of its neighboring edges but on the entire configuration $A$, since connectivity is a global property.

At the other extreme, when $q \to 0$, the set of (weak) limits that arise for various choices of $p$ contains fundamental distributions on $G$, including the uniform measures over the spanning trees, spanning forests and connected subgraphs of $G$.

The random-cluster model is also well defined for infinite graphs. If $G$ is an infinite graph and $\{G_n\}$ is a sequence of finite subgraphs of $G$ such that $\{G_n\} \to G$, then the random-cluster measure on $G$ is given by the limit of the sequence of random-cluster measures on $\{G_n\}$. In this setting, a key feature of the model is the presence of the following *phase transition*: there is a critical value $p = p_c(q)$ such that for $p < p_c(q)$ all connected components are finite w.h.p.[1], while for $p > p_c(q)$ there is at least one infinite component w.h.p. The former regime is called the *disordered phase*, and the latter is the *ordered phase*. In finite graphs $p_c(q)$ corresponds to the critical value for the emergence of a "giant" component of linear size.

Random-cluster *dynamics*, i.e., Markov chains on random-cluster configurations that converge to the random-cluster measure, are of major interest. There are a variety of natural dynamics for this model, including the standard *Glauber* dynamics and the more sophisticated *Swendsen-Wang* [48] and *Chayes-Machta* [10] processes, which have received much attention. The primary object of study is the *mixing time*, i.e., the number of steps until the dynamics is close to its stationary distribution, starting from any initial configuration. A fundamental question is how the mixing time of these dynamics grows as the size of the graph $G$ increases, and in particular how it relates to the phase transition.

This thesis study these questions in two canonical cases: the case when $G$ is the complete graph on $n$ vertices, known as the *mean-field model*, and the case when $G$ is the infinite 2-dimensional lattice graph $\mathbb{Z}^2$. The mean-field random-cluster model may be viewed as the standard Erdős-Rényi random graph model $\mathcal{G}_{n,p}$, enriched by a factor that depends on the component structure. As we shall see, this case is already quite non-trivial; moreover, it has historically proven to be a useful starting point in understanding the dynamics on more general graphs. In statistical mechanics, however, probabilistic models are most frequently studied in the setting of infinite lattice graphs; understanding random-cluster dynamics in $\mathbb{Z}^2$ is thus of foremost importance.

## 1.1 Markov chains

Dynamics for spin systems have been widely studied in both statistical physics and computer science. On the one hand, they provide a Markov chain Monte Carlo algorithm for sampling configurations of the system from the Gibbs distribution; on the other hand, they are in many cases a plausible model for the evolution of the underlying physical system.

There has been much activity over the past two decades in analyzing dynamics for spin systems such as the Ising/Potts model, and deep connections have emerged between the mixing time and the phase structure of the physical model. In contrast, dynamics for the random-cluster model remain poorly understood. The main reason for this appears to be the fact mentioned

[1] We say that an event occurs *with high probability (w.h.p.)* if it occurs with probability approaching 1 as $n \to \infty$.

above that connectivity is a global property; this has led to the lack of a precise understanding of the phase transition, as well as the failure of existing Markov chain analysis tools.

### 1.1.1 Glauber dynamics

A *Glauber dynamics* for the random-cluster model is any *local* Markov chain on configurations that is ergodic and reversible with respect to the measure (1.1), and hence converges to it. For definiteness we consider the *heat-bath dynamics*, which at each step updates one edge of the current random-cluster configuration $A$ as follows:

(i) pick an edge $e \in E$ uniformly at random (u.a.r.);

(ii) replace $A$ by $A \cup \{e\}$ with probability
$$\frac{\mu_{G,p,q}(A \cup \{e\})}{\mu_{G,p,q}(A \cup \{e\}) + \mu_{G,p,q}(A \setminus \{e\})};$$

(iii) else replace $A$ by $A \setminus \{e\}$.

These transition probabilities can be easily computed:
$$\frac{\mu_{G,p,q}(A \cup \{e\})}{\mu_{G,p,q}(A \cup \{e\}) + \mu_{G,p,q}(A \setminus \{e\})} = \begin{cases} \frac{p}{p+q(1-p)} & \text{if } e \text{ is a "cut edge" in } (V,A); \\ p & \text{otherwise.} \end{cases}$$
We say $e$ is a *cut edge* in $(V,A)$ iff changing the current configuration of $e$ changes the number of connected components of $(V,A)$.

### 1.1.2 Swendsen-Wang dynamics

The *Swendsen-Wang (SW) dynamics* [48] is primarily a dynamics for the Ising/Potts model, but it may alternatively be viewed as a Markov chain for the random-cluster model using a coupling of these measures due to Edwards and Sokal [17]. The SW dynamics mixes rapidly (i.e., in polynomial time) in some cases where Glauber dynamics are known to mix exponentially slowly. Consequently, this dynamics has been well studied and is widely used. Given a random-cluster configuration $(V,A)$, a new configuration is obtained as follows:

(i) assign to each connected component of $(V,A)$ a color from $\{1,\dots,q\}$ u.a.r.;

(ii) remove all edges;

(iii) add each *monochromatic* edge independently with probability $p$.

Note that the SW dynamics is highly non-local as it modifies the entire configuration in one step. Using the Edwards-Sokal coupling, it is straightforward to check that the SW dynamics is reversible with respect to the random-cluster measure. If the starting configuration is instead an Ising/Potts configuration $\sigma$, steps (iii)-(i)-(ii) above, in that order, give the SW dynamics for the Ising/Potts model.

### 1.1.3 Chayes-Machta dynamics

Another non-local Markov chain is the *Chayes-Machta (CM) dynamics* [10]. Given a random-cluster configuration $(V, A)$, one step of this dynamics is defined as follows:

(i) *activate* each connected component of $(V, A)$ independently with probability $1/q$;

(ii) remove all edges connecting active vertices;

(iii) add each edge connecting active vertices independently with probability $p$, leaving the rest of the configuration unchanged.

It is easy to check that this dynamics is reversible with respect to (1.1) [10]. For integer $q$, the CM dynamics is a close cousin of the Swendsen-Wang dynamics. However, the SW dynamics is only well-defined for integer $q$, while the random-cluster model makes perfect sense for all $q > 0$. The CM dynamics, which is feasible for any *real* $q \geq 1$, was introduced precisely in order to allow for this generalization.

## 1.2 Phase transition

### 1.2.1 Mean-field phase transition

The phase transition for the mean-field random-cluster model is already well understood [7, 40]. It is natural here to re-parameterize by setting $p = \lambda/n$; the phase transition then occurs at the critical value $\lambda = \lambda_c(q)$ given by

$$\lambda_c(q) = \begin{cases} q & \text{for } 0 < q \leq 2; \\ 2\left(\frac{q-1}{q-2}\right)\log(q-1) & \text{for } q > 2. \end{cases}$$

For $\lambda < \lambda_c(q)$ all components are of size $O(\log n)$ w.h.p., while for $\lambda > \lambda_c(q)$ there is a unique giant component of size $\theta_r n$, where $\theta_r = \theta_r(\lambda, q)$ is the largest $x > 0$ satisfying the equation

$$e^{-\lambda x} = 1 - \frac{qx}{1 + (q-1)x}. \tag{1.2}$$

This phase transition is analogous to that in $\mathcal{G}_{n,p}$ corresponding to the appearance of a "giant" component of linear size. Indeed, when $q = 1$ equation (1.2) becomes $e^{-\lambda x} = 1 - x$ whose positive solution specifies to the size of the giant component in $\mathcal{G}_{n,p}$.

### 1.2.2 Phase transition in $\mathbb{Z}^2$

As mentioned earlier, the random-cluster measure is well defined for the infinite 2-dimensional lattice graph $\mathbb{Z}^2$ as the limit of the sequence of random-cluster measures on $n \times n$ square regions

$\Lambda_n$ of $\mathbb{Z}^2$ as $n$ goes to infinity. Recent breakthrough work of Beffara and Duminil-Copin [3] for the infinite measure in $\mathbb{Z}^2$ established that, for all $q \geq 1$, the phase transition occurs at

$$p = p_c(q) = \frac{\sqrt{q}}{\sqrt{q}+1};$$

hence, for $p < p_c(q)$ all components are finite w.h.p., and for $p > p_c(q)$ there is an infinite component w.h.p. It was also established in [3] that for $p < p_c(q)$ the model exhibits "decay of connectivities", i.e., the probability that two vertices lie in the same connected component decays to zero exponentially with the distance between them. This property is analogous to the classical "decay of correlations" that has long been known for the Ising model (see, e.g., [41]).

## 1.3 Results

### 1.3.1 Mean-field dynamics

In Chapter 3 of this thesis we analyze the mixing time of the CM and Glauber dynamics for the mean-field random-cluster model; the results in Chapter 3 appeared in [5]. Our first result shows that the CM dynamics reaches equilibrium very rapidly for all non-critical values of $\lambda$ and all $q \in (1, 2]$.

**Theorem 1.1.** *For any $q \in (1, 2]$, the mixing time of the mean-field CM dynamics is $\Theta(\log n)$ for all $\lambda \neq \lambda_c(q)$.*

To state our results for $q > 2$, we identify two further critical points, $\lambda_s(q)$ and $\lambda_S(q)$, with the property that $\lambda_s(q) < \lambda_c(q) < \lambda_S(q)$. (The definitions of these points are somewhat technical and can be found in Chapter 2.) We show that the CM dynamics mixes rapidly for $\lambda$ outside the "critical" window $(\lambda_s, \lambda_S)$, and exponentially slowly inside this window. We also establish the mixing time at the critical points $\lambda_s$ and $\lambda_S$.

**Theorem 1.2.** *For any $q > 2$, the mixing time $\tau_{\text{mix}}^{\text{CM}}$ of the mean-field CM dynamics satisfies:*

$$\tau_{\text{mix}}^{\text{CM}} = \begin{cases} \Theta(\log n) & \textit{if } \lambda \in (\lambda_s, \lambda_S]; \\ \Theta(n^{1/3}) & \textit{if } \lambda = \lambda_s; \\ e^{\Omega(\sqrt{n})} & \textit{if } \lambda \notin (\lambda_s, \lambda_S). \end{cases}$$

As a byproduct of the results above we deduce new bounds for the mixing time of the mean-field heat-bath dynamics for all $q > 1$.

**Theorem 1.3.** *For any $q \in (1,2]$, the mixing time $\tau_{\mathrm{mix}}^{\mathrm{HB}}$ of the mean-field heat-bath dynamics is $\tilde{O}(n^4)$ for all $\lambda \neq \lambda_c(q)$. Moreover, for $q > 2$ we have*

$$\tau_{\mathrm{mix}}^{\mathrm{HB}} = \begin{cases} \tilde{O}(n^4) & \text{if } \lambda \in (\lambda_s, \lambda_S]; \\ \tilde{O}(n^{4+1/3}) & \text{if } \lambda = \lambda_s; \\ e^{\Omega(\sqrt{n})} & \text{if } \lambda \notin (\lambda_s, \lambda_S). \end{cases}$$

(The $\tilde{O}$ notation hides polylogarithmic factors.)

We now provide an interpretation of these results. When $q > 2$ the mean-field random-cluster model exhibits a *first-order* phase transition, which means that at criticality ($\lambda = \lambda_c$) the ordered and disordered phases mentioned earlier *coexist* [40], i.e., each contributes about half of the probability mass. Phase coexistence suggests exponentially slow mixing for most natural dynamics, because of the difficulty of moving between the phases. Moreover, by continuity we should expect that, within a constant-width interval around $\lambda_c$, the effect of the non-dominant phase (ordered below $\lambda_c$, disordered above $\lambda_c$) will still be felt, as it will form a second mode (local maximum) for the random-cluster measure. This leads to so-called *metastable* states near that local maximum from which it is very hard to escape, so slow mixing should persist throughout this interval. Intuitively, the values $\lambda_s, \lambda_S$ mark the points at which the local maxima disappear. A similar phenomenon was captured in the case of the Potts model by Cuff *et al.* [13]. Our results make the above picture for the dynamics rigorous for the random-cluster model for all $q > 2$; notably, in contrast to the Potts model, in the random-cluster model metastability affects the mixing time on *both* sides of $\lambda_c$. For $q \leq 2$, the model exhibits a *second-order* phase transition and there is no phase existence; hence, metastable states are not present and there is no slow mixing window.

We provide next some brief remarks about our mean-field techniques. Both our upper and lower bounds on the mixing time of the CM dynamics focus on the evolution of the one dimensional random process given by the size of the largest component. A key ingredient in our analysis is a function that describes the expected change, or "drift", of this random process at each step; the critical points $\lambda_s$ and $\lambda_S$ discussed above arise naturally from consideration of the zeros of this drift function.

For our upper bounds, we construct a multiple-phase coupling of the evolution of two arbitrary configurations; this coupling is similar in flavor to that used by Long *et al.* [38] for the SW dynamics for $q = 2$, but there are significant additional complexities in that our analysis has to identify the "slow mixing" window $(\lambda_s, \lambda_S)$ for $q > 2$, and also has to contend with the fact that in the CM dynamics only a subset of the vertices (rather than the whole graph, as in SW) are active at each step.

For our exponential lower bounds we use the drift function to identify the metastable states mentioned earlier from which the dynamics cannot easily escape. For both upper and lower bounds, we have to handle the sub-critical and super-critical cases, $\lambda < \lambda_c$ and $\lambda > \lambda_c$, separately because the structure of typical configurations differs in the two cases.

For the heat-bath dynamics we use a recent surprising development of Ullrich [50, 51], who showed that the mixing time of the heat-bath dynamics on any graph differs from that of the SW dynamics by at most a poly($n$) factor. Thus the previously known bounds for SW translate to bounds for the heat-bath dynamics for integer $q$. By adapting Ullrich's technology to the CM setting, we are able to obtain a similar translation of our results and establish Theorem 1.3.

**Remark.** In Theorem 1.1 we leave open the mixing time of the CM dynamics for $q \in (1,2]$ and $\lambda = \lambda_c$. However, in the final version of [5], which is currently in preparation, we show that the mixing time when $q \in (1,2)$ and $\lambda = \lambda_c$ is also $\Theta(\log n)$. This confirms a conjecture of Machta [22]. Moreover, it shows that in this regime the CM dynamics does not suffer from a critical slowdown, a very atypical behavior among dynamics of physical systems.

### 1.3.2 Glauber dynamics in $\mathbb{Z}^2$

In Chapter 4 we explore the consequences of Beffara-Duminil-Copin's structural result for the dynamics of the model when $G$ is an $n \times n$ square region $\Lambda_n$ of $\mathbb{Z}^2$ and $q \geq 1$; the results in Chapter 4 appeared in [6]. We prove the following tight theorem:

**Theorem 1.4.** *For any $q \geq 1$, the mixing time of the Glauber dynamics for the random-cluster model on $\Lambda_n \subset \mathbb{Z}^2$ is $\Theta(n^2 \log n)$ at all values of $p \neq p_c(q)$.*

Theorem 1.4, as stated, holds for the random-cluster model with so-called "free" boundary conditions (i.e., there are no edges in $\mathbb{Z}^2 \setminus \Lambda_n$). In fact, as a consequence of our proof, it also holds for the case of "wired" boundary conditions (in which all vertices on the external face of $\Lambda_n$ are connected).

The main component of our result is the analysis of the sub-critical regime $p < p_c$; the result for the super-critical regime $p > p_c$ follows from it easily by self-duality of $\mathbb{Z}^2$ and the fact that $p_c$ is exactly the self-dual point [3]. Our sub-critical upper bound analysis makes crucial use of the exponential decay of connectivities for $p < p_c$ established recently by Beffara and Duminil-Copin [3], as discussed earlier. This analysis is reminiscent of similar results for spin systems (such as the Ising model), in which exponential decay of *correlations* has been shown to imply rapid mixing [42]. However, since the random-cluster model exhibits decay of connectivities rather than decay of correlations, we need to rework the standard tools used in these contexts. In particular, we make three innovations.

First, the classical notion of "disagreement percolation" [4], which is used to bound the speed at which influence can propagate in $\mathbb{Z}^2$ under the dynamics, has to be extended to take account of the fact that in the random-cluster model influence spreads not from vertex to vertex but from cluster to cluster. Second, we need to translate the decay of connectivities in the infinite volume $\mathbb{Z}^2$ (as proved in [3]) to a stronger "spatial mixing" property in finite volumes $\Lambda_n$, with suitable *boundary conditions* around the external face; in doing this we use the machinery developed by Alexander in [1], but adapted to hold for arbitrary (not just integer) $q$ and for a suitable class of boundary conditions that we call "side-homogeneous" (see Section 4.1 for a definition). Finally, while we follow standard recursive arguments in relating the mixing time in $\Lambda_n$ to that in smaller regions $\Lambda_{n'}$ for $n' \ll n$, our approach differs in its sensitivity to the boundary conditions on the smaller regions: previous applications for spin systems have typically required rapid mixing to hold in $\Lambda_{n'}$ for *arbitrary* boundary conditions, while in our case we require it to hold only for side-homogeneous conditions. This aspect of our proof is actually essential because the random-cluster model does *not* exhibit spatial mixing for arbitrary boundary conditions (see Section 4.3); our definition of side-homogeneous conditions is motivated by the fact that they are both restricted enough to allow spatial mixing to hold, and general enough to make the recursion go through. Our lower bound proof uses technology from analogous results for spin systems of Hayes and Sinclair [28], again adapted to the random-cluster setting.

## 1.4 Related work

The random-cluster model has been the subject of extensive research in both the applied probability and statistical physics communities, which is summarized in the book by Grimmett [26].

A central open problem was to rigorously establish the phase transition in $\mathbb{Z}^2$ at $p_c(q) = \sqrt{q}/(\sqrt{q}+1)$, though this was not achieved until 2012 by Beffara and Duminil-Copin [3]. The continuity (or "order") of this phase transition is still not fully understood: it is conjectured to be continuous of second order for $q \leq 4$ and discontinuous of first order for $q > 4$ [26]. This conjecture has only been verified for large $q \geq 25.72$ by Laanait et al. [35] and in a recent development by Duminil-Copin, Sidoravicius and Tassion for $1 \leq q \leq 4$ [14]. The much simpler mean-field phase transition was established by Bollobás, Grimmett and Janson [7]. A more detailed description of this phase transition was later provided by Luczak and Łuczak [40].

Thanks to decades of research, Ising/Potts model dynamics are well understood in many settings. For example, the Glauber dynamics for the Ising model on $\mathbb{Z}^2$ is essentially completely understood: at all parameter values below the critical point $\beta_c$, the mixing time in $\Lambda_n$ is $O(n^2 \log n)$, while above $\beta_c$ it is $\exp(\Omega(n))$ (see [41] for a comprehensive treatment, and also [39] for the behavior at $\beta_c$).

Analogous results for the $q$-state Potts model for $\beta < \beta_c$ follow from the random-cluster results of Beffara and Duminil-Copin [3] and Alexander [2], combined with the earlier work of Martinelli, Olivieri and Schonmann [42] relating spatial mixing to mixing times. Very recently, Gheissari and Lubetzky [24] use the results of Duminil-Copin, Sidoravicius and Tassion [14] to establish the mixing time at the critical point $\beta_c$. They show that the mixing time is at most polynomial in $n$ for $q = 3$, at most quasi-polynomial for $q = 4$ and $\exp(\Omega(n))$ for $q > 4$, the latter assuming the expected (but not yet established) discontinuity of the phase transition for $q > 4$. Gheissari and Lubetzky also prove an exponential lower bound when $q > 1$ and $\beta > \beta_c$.

All the mixing time bounds above for the Ising/Potts model *indirectly* provide bounds for random-cluster dynamics. The comparison technology developed by Ullrich [51, 50, 49] allows bounds for the Glauber dynamics of the Ising/Potts models to be translated to the SW dynamics, and then again to the random-cluster dynamics. This leads, for example, to an upper bound of $O(n^6 \log^2 n)$ on the mixing time of the random-cluster Glauber dynamics in $\Lambda_n \subset \mathbb{Z}^2$, at all values $p \neq p_c(q)$ for all integer $q \geq 2$.

This approach has several limitations. First, the comparison method invokes linear algebra, and hence suffers an inherent penalty of at least $\Omega(n^4)$ in the mixing time bound. Second, the comparison method also yields no insight into the actual behavior of the random-cluster dynamics, so, e.g., it is unlikely to illuminate the connections with phase transitions. Finally, since it relies on comparison with the Ising/Potts models, this analysis applies only for *integer* values of $q$, while the random-cluster model is defined for all positive values of $q$.

Glauber dynamics for the mean-field Ising/Potts model has been studied thoroughly (see, e.g., [36, 13]). The mean-field SW dynamics has also received much attention, and its mixing time is now fully understood. For $q = 2$, Long, Nachmias, Ning and Peres [38], building on earlier work of Cooper, Dyer, Frieze and Rue [12], showed that the mixing time is $\Theta(1)$ for $\lambda < \lambda_c$, $\Theta(\log n)$ for $\lambda > \lambda_c$, and $\Theta(n^{1/4})$ for $\lambda = \lambda_c$. Recent work of Galanis, Štefankovič and Vigoda [21], independent of our work and published at around the same time, provides a similarly comprehensive treatment of the SW mixing time for $q \geq 3$. Indeed, they proved the equivalent of our Theorem 1.2 for the SW dynamics. The first studies of the $q \geq 3$ case were due to Huber [30] and Gore and Jerrum [25]. All these bounds were established for the SW dynamics in the framework of the Ising/Potts model, but they apply also to the random-cluster version of the dynamics. However, as mentioned earlier, the relevance of the SW dynamics in the random-cluster setting is limited to the special case of integer $q$.

Finally, we mention relevant results on the dynamics in other graphs. For the $d$-dimensional torus, Borgs et al. [8, 9] proved exponential lower bounds for the mixing time of the SW dynamics for $p = p_c(q)$ and $q$ sufficiently large. Ge and Štefankovič [23] provide a polynomial bound on the mixing time of the Glauber dynamics on graphs with bounded tree-width. Very recently, Guo and Jerrum [27] proved that both the random-cluster Glauber dynamics and the SW dynamics mix in polynomial time on *any* graph in the special case $q = 2$.

## 1.5 Bibliographic Notes

The results in this thesis were derived in collaboration with my advisor Alistair Sinclair, who I thank for allowing the inclusion of our coauthored work. Most of these results have already been published in [5] and [6].

# Chapter 2

# Probabilistic Preliminaries

In this chapter we gather a number of standard definitions and background results in probability theory that we will refer to repeatedly in Chapters 3 and 4. Proofs are provided for those results that are not readily available elsewhere.

## 2.1 Concentration bounds

**Theorem 2.1** (Chernoff Bounds). *Let $X_1, \ldots, X_k$ be independent Bernoulli random variables. Let $X = \sum_i X_i$ and $\mu = \mathrm{E}[X]$; then for any $\delta \in (0, 1)$,*

$$\Pr[|X - \mu| > \delta\mu] \le 2 \exp\left(-\frac{\delta^2 \mu}{4}\right).$$

**Theorem 2.2** (Hoeffding's Inequality). *Let $X_1, \ldots, X_k$ be independent random variables such that $\Pr[X_i \in [a_i, b_i]] = 1$. Let $X = \sum_i X_i$ and $\mu = \mathrm{E}[X]$; then for any $\delta > 0$,*

$$\Pr[|X - \mu| > \delta] \le 2 \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{k}(b_i - a_i)^2}\right).$$

**Theorem 2.3** ([43], Theorem 2.7). *Let $X_1, \ldots, X_k$ be independent random variables such that $|X_i| \le B$ for all $i$. Let $X = \sum_i X_i$, $\mu = \mathrm{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$; then for any $\delta > 0$,*

$$\Pr[X > \mu + \delta] \le \exp\left(-\frac{\delta^2}{2\sigma^2 + \delta B}\right).$$

## 2.2 Mixing time

Let $P$ be the transition matrix of a finite, ergodic Markov chain $M$ with state space $\Omega$ and stationary distribution $\pi$. Let

$$\tau_{\mathrm{mix}}(\varepsilon) := \max_{z \in \Omega} \min_{t} \left\{ ||P^t(z, \cdot) - \pi(\cdot)||_{\mathrm{TV}} \le \varepsilon \right\}$$

where $||\cdot||_{\mathrm{TV}}$ denotes total variation distance. The *mixing time* of $M$ is given by $\tau_{\mathrm{mix}} := \tau_{\mathrm{mix}}(1/4)$. It is well-known that $\tau_{\mathrm{mix}}(\varepsilon) \le \lceil \log_2 \varepsilon^{-1} \rceil\, \tau_{\mathrm{mix}}$ for any positive $\varepsilon < 1/2$ (see, e.g., [37, Ch. 4.5]).

A *(one step) coupling* of the Markov chain $M$ specifies for every pair of states $(X_t, Y_t) \in \Omega^2$ a probability distribution over $(X_{t+1}, Y_{t+1})$ such that the processes $\{X_t\}$ and $\{Y_t\}$, viewed in isolation, are faithful copies of $M$, and if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$. The *coupling time* is defined by

$$T_{\mathrm{coup}} = \max_{x,y\in\Omega} \min_t \{X_t = Y_t | X_0 = x, Y_0 = y\}.$$

For any $\delta \in (0,1)$, the following standard inequality (see, e.g., [37]) provides a bound on the mixing time:

$$\tau_{\mathrm{mix}} \le \min_t \left\{ \Pr[T_{\mathrm{coup}} > t] \le 1/4 \right\} \le O\left(\delta^{-1}\right) \cdot \min_t \left\{ \Pr[T_{\mathrm{coup}} > t] \le 1 - \delta \right\}. \tag{2.1}$$

## 2.3 Random graphs

Let $G_d$ be distributed as a $G(n, p = d/n)$ random graph where $d > 0$. We say that $d$ is bounded away from 1 if there exists a constant $\xi$ such that $|d-1| \ge \xi$. Let $\mathcal{L}(G_d)$ denote the largest component of $G_d$ and let $L_i(G_d)$ denote the *size* of the $i$-th largest component of $G_d$. (Thus, $L_1(G_d) = |\mathcal{L}(G_d)|$.) In our proofs we will use several facts about the random variables $L_i(G_d)$, which we gather here for convenience. We provide proofs for those results that are not available in the random graph literature.

**Lemma 2.4** ([38], Lemma 5.7)**.** *Let $I(G_d)$ denote the number of isolated vertices in $G_d$. If $d = O(1)$, then there exists a constant $C > 0$ such that* $\Pr[I(G_d) > Cn] = 1 - O\left(n^{-1}\right)$.

**Lemma 2.5.** *If $d = O(1)$, then $L_2(G_d) < 2n^{11/12}$ with probability $1 - O\left(n^{-1/12}\right)$ for sufficiently large $n$.*

*Proof.* If $d \le 1 + n^{-1/12}$, then by Theorem 5.9 in [38] (with $A^2 = c^{-1}\log n$ and $\epsilon = n^{-1/12}/2$), $L_1(G_d) < 2n^{11/12}$ with probability $1 - O(n^{-1})$. When $d > 1 + n^{-1/12}$ we bound $L_2(G_d)$ using Theorem 5.12 in [33]. Observe that this result applies to the random graph model $G(n, M)$ where an instance $G_M$ is chosen u.a.r. from the set of graphs with $n$ vertices and $M$ edges. The $G(n,p)$ and $G(n,M)$ models are known to be essentially equivalent when $M \approx \binom{n}{2}p$ and we can easily transfer this result to our setting.

Let $M_d$ be the number of edges in $G_d$ and $I := \left[\binom{n}{2}p - \sqrt{8dn\log n}, \binom{n}{2}p + \sqrt{8dn\log n}\right]$; a Chernoff bound implies

$$\Pr[L_2(G_d) > n^{2/3}] \quad \le \quad \sum_{m\in I} \Pr[L_2(G_m) > n^{2/3}] \Pr[M_d = m] + O(n^{-1}).$$

Let $s = m - n/2$ as in [33]; since $d > 1 + n^{-1/12}$, then $s \ge \frac{n^{11/12}}{4}$ for $m \in I$ and $n$ sufficiently large. Theorem 5.12 in [33] implies that $\Pr[L_2(G_m) > n^{2/3}] = O(n^{-1/12})$; thus, $L_2(G_d) < 2n^{11/12}$ with probability $1 - O(n^{-1/12})$. □

**Lemma 2.6** ([12], Lemma 7)**.** *If $d < 1$ is bounded away from 1, then $L_1(G_d) = O(\log n)$ with probability $1 - O\left(n^{-1}\right)$.*

For $d > 1$, let $\beta = \beta(d)$ be the unique positive root of the equation

$$e^{-dx} = 1 - x. \tag{2.2}$$

(Note that this equation has a positive root iff $d > 1$; see, e.g., [33].)

**Lemma 2.7.** *Let $\widetilde{G}_{d_n}$ be distributed as a $G(n+m, d_n/n)$ random graph where $|m| = o(n)$ and $\lim\limits_{n\to\infty} d_n = d$. Assume $1 < d_n = O(1)$ and $d_n$ is bounded away from 1 for all $n \in \mathbb{N}$. Then,*

*(i) $L_2(\widetilde{G}_{d_n}) = O(\log n)$ with probability $1 - O\left(n^{-1}\right)$.*

*(ii) For $A = o(\log n)$ and sufficiently large $n$, there exists a constant $c > 0$ such that*

$$\Pr[|L_1(\widetilde{G}_{d_n}) - \beta(d)n| > |m| + A\sqrt{n}] \le e^{-cA^2}. \tag{2.3}$$

*Proof.* Part (i) follows immediately from Lemma 7 in [12]. For Part (ii), let $M = n + m$ and $d_M = d_n M/n$. By Lemma 11 in [12], there exists a constant $c > 0$ such that

$$\begin{aligned} e^{-cA^2} &\ge \Pr[|L_1(\widetilde{G}_{d_n}) - \beta(d_M)M| > A\sqrt{M}] \\ &\ge \Pr[|L_1(\widetilde{G}_{d_n}) - \beta(d_M)n| > |\beta(d_M)m| + A\sqrt{M}] \\ &\ge \Pr[|L_1(\widetilde{G}_{d_n}) - \beta(d)n| > |\beta(d)n - \beta(d_M)n| + |\beta(d_M)m| + A\sqrt{M}]. \end{aligned}$$

Now, since $d_M \to d$, by continuity$\beta(d_M) \to \beta(d)$ as $n \to \infty$. Therefore, for a sufficiently large $n$,

$$\Pr[|L_1(\widetilde{G}_{d_n}) - \beta(d)n| > |\beta(d_M)m| + 3A\sqrt{n}] \le e^{-cA^2}$$

and the result follows since $\beta(d_M) \le 1$. □

**Lemma 2.8.** *Let $\widetilde{G}_{d_n}$ be distributed as a $G(n+m, d_n/n)$ random graph where $\lim_{n\to\infty} d_n = d$ and $|m| = o(n)$. Assume $1 < d_n = O(1)$ and $d_n$ is bounded away from 1 for all $n \in \mathbb{N}$. Then, $\mathrm{Var}(L_1(\widetilde{G}_{d_n})) = \Theta(n)$, and for any desired constant $\alpha > 0$ and sufficiently large $n$, we have*

$$\beta(d)(n+m) - 2n^\alpha \le \mathrm{E}[L_1(\widetilde{G}_{d_n})] \le \beta(d)(n+m) + 2n^\alpha.$$

*Proof.* Let $M = n + m$ and $d_M = d_n M/n$. Observe that $G(n+m, d_n/n) = G(M, d_M/M)$ and thus Theorem 5 from [11] implies that $\mathrm{Var}(L_1(\widetilde{G}_{d_n})) = \Theta(M) = \Theta(n)$. Also by Theorem 5 from [11], we have

$$\beta(d_M)M - n^\alpha \le \mathrm{E}[L_1(\widetilde{G}_{d_n})] \le \beta(d_M)M + n^\alpha$$

for any desired constant $\alpha$. Since $d_M \to d$, by continuity $\beta(d_M) \to \beta(d)$ as $n \to \infty$. Therefore, for a sufficiently large $n$, $\beta(d)(n+m) - 2n^\alpha \le \mathrm{E}[L_1(\widetilde{G}_{d_n})] \le \beta(d)(n+m) + 2n^\alpha$, as desired. □

**Lemma 2.9.** *Consider a $G_{d_n}$ random graph where $\lim_{n\to\infty} d_n = d$. Assume $1 < d_n = O(1)$ and $d_n$ is bounded away from 1 for all $n \in \mathbb{N}$. Then, for any constant $\varepsilon \in (0,1)$ there exists a constant $c(\varepsilon) > 0$ such that, for sufficiently large n,*

$$\Pr[|L_1(G_{d_n}) - \beta(d_n)n| > \varepsilon n] \leq e^{-c(\varepsilon)n}.$$

*Proof.* This result follows easily from Lemma 5.4 in [38]. Let $a_1$ and $a_2$ be constants such that $d \in (\gamma_1, \gamma_2)$. Since $\{d_n\} \to d$, there exists $N \in \mathbb{N}$ such that $d_n \in (\gamma_1, \gamma_2)$ for all $n > N$.

By Lemma 5.4 in [38] (with $A = \varepsilon\sqrt{n}$), there exist constants $c_1(\varepsilon)$, $c_2(\varepsilon) > 0$ such that $\Pr[L_1(G_{\gamma_1}) < \beta(\gamma_1)n - \varepsilon n] \leq \exp(-c_1(\varepsilon)n)$ and $\Pr[L_1(G_{\gamma_2}) > \beta(\gamma_2)n + \varepsilon n] \leq \exp(-c_2(\varepsilon)n)$. By monotonicity $\beta(\gamma_2) > \beta(d_n) > \beta(\gamma_1)$, and by continuity we can choose $\gamma_1$ and $\gamma_2$ sufficiently close to each other such that $|\beta(\gamma_2) - \beta(\gamma_1)| < \varepsilon$. Observe also that $L_1(G_{\gamma_2}) \succeq L_1(G_{d_n}) \succeq L_1(G_{\gamma_1})$, where $\succeq$ indicates stochastic domination[1]. Thus,

$$\begin{aligned} e^{-c_1(\varepsilon)n} &\geq \Pr[L_1(G_{\gamma_1}) < \beta(\gamma_1)n - \varepsilon n] \\ &\geq \Pr[L_1(G_{d_n}) < \beta(\gamma_1)n - \varepsilon n] \\ &\geq \Pr[L_1(G_{d_n}) < \beta(d_n)n - 2\varepsilon n] \end{aligned}$$

and similarly,

$$\begin{aligned} e^{-c_2(\varepsilon)n} &\geq \Pr[L_1(G_{\gamma_2}) > \beta(\gamma_2)n + \varepsilon n] \\ &\geq \Pr[L_1(G_{d_n}) > \beta(\gamma_2)n + \varepsilon n] \\ &\geq \Pr[L_1(G_{d_n}) > \beta(d_n)n + 2\varepsilon n]. \end{aligned}$$

Hence, there exist a constant $c(\varepsilon)$ such that $\Pr[|L_1(G_{d_n}) - \beta(d_n)n| > \varepsilon n] \leq e^{-c(\varepsilon)n}$. □

**Lemma 2.10.** *Assume d is bounded away from 1. If $d < 1$, then $L_1(G_d) = O(\sqrt{n})$ with probability $1 - e^{-\Omega(\sqrt{n})}$. If $d > 1$, then $L_2(G_d) = O(\sqrt{n})$ with probability $1 - e^{-\Omega(\sqrt{n})}$.*

*Proof.* When $d < 1$ the result follows immediately from Lemma 6 in [25]. When $d > 1$, by Lemma 2.9, $L_1(G_d) \in I = [(\beta(d) - \varepsilon)n, (\beta(d) + \varepsilon)n]$ with probability $1 - e^{-\Omega(n)}$. Conditioning on $L_1(G_d) = m$, by the *discrete duality principle* (see, e.g., [29]) the remaining subgraph is distributed as a $G(n-m, d/n)$ random graph which is sub-critical for $m \in I$ and $\varepsilon$ sufficiently small. Therefore as for $d < 1$, $L_2(G_d) = O(\sqrt{n})$ with probability $1 - e^{-\Omega(\sqrt{n})}$ as desired. □

**Lemma 2.11.** *Assume d is bounded away from 1. If $d < 1$, then $\sum_{i\geq 1} L_i(G_d)^2 = O(n)$ with probability $1 - O\left(n^{-1}\right)$. If $d > 1$, then $\sum_{i\geq 2} L_i(G_d)^2 = O(n)$ with probability $1 - O\left(n^{-1}\right)$.*

*Proof.* When $d < 1$ the result follows by Chebyshev's inequality from Theorem 1.1 in [32]. When $d > 1$ the result follows from the discrete duality principle as in Lemma 2.10. □

[1] For distributions $\mu$ and $\nu$ over a partially ordered set $\Gamma$, we say that $\mu$ stochastically dominates $\nu$ if $\int g\, d\nu \leq \int g\, d\mu$ for all increasing functions $g : \Gamma \to \mathbb{R}$.

### 2.3.1 Random graphs near the critical regime

**Lemma 2.12** ([38], Theorem 5.9). *Let $G$ be distributed as a $G(n, \frac{1+\varepsilon}{n})$ random graph. For any positive constant $\rho \le 1/10$, there exist constants $C \ge 1$ and $c > 0$ such that if $\varepsilon^3 n \ge C$, then*

$$\Pr[|L_1(G) - 2\varepsilon n| > \rho\varepsilon n] = O(e^{-c\varepsilon^3 n}).$$

**Lemma 2.13** ([38], Theorem 5.12). *Let $G$ be distributed as a $G(n, \frac{1-\varepsilon}{n})$ random graph with $\varepsilon > 0$. Then, $\mathrm{E}[\sum_{i\ge1} L_i(G)^2] = O(\frac{n}{\varepsilon})$.*

**Lemma 2.14** ([38], Theorem 5.13). *Let $G$ be distributed as a $G(n, \frac{1+\varepsilon}{n})$ random graph with $\varepsilon > 0$ and $\varepsilon^3 n \ge 1$ for large $n$. Then, $\mathrm{E}[\sum_{i\ge2} L_i(G)^2] = O(\frac{n}{\varepsilon})$.*

**Lemma 2.15.** *Let $G$ be distributed as a $G(n, \frac{1+\varepsilon}{n})$ random graph with $\varepsilon > 0$ and $\varepsilon^3 n \ge 1$ for large $n$. Then, $\mathrm{E}[L_1(G)^2] = O(\frac{n}{\varepsilon}) + O(\varepsilon^2 n^2)$.*

*Proof.* Follows from Corollary 5.10 in [38]. □

**Lemma 2.16.** *For any $d > 0$, $\mathrm{E}[\sum_{i\ge2} L_i(G_d)^2] = O(n^{4/3})$.*

*Proof.* If $d \le 1-n^{-1/3}$ (resp., $d \ge 1+n^{-1/3}$) the result follows from Lemma 2.13 (resp., Lemma 2.14). If $d \in (1-n^{-1/3}, 1+n^{-1/3})$, by monotonicity it is enough to show that $\mathrm{E}[L_1(G_d) + \sum_{i\ge2} L_i(G_d)^2] = O(n^{4/3})$ when $d = 1 + n^{-1/3}$. This follows from Lemmas 2.14 and 2.15. □

**Lemma 2.17.** *Let $G$ be distributed as a $G(n, \frac{1+\varepsilon}{n})$ random graph with $\varepsilon \in (-cn^{-1/3}, cn^{-1/3})$, where $c > 0$ is a constant independent of $n$. Then, for any constant $a > 0$, there exists $r = r(a, c)$ such that $\Pr[L_1(G) \ge an^{2/3}] \ge r$.*

*Proof.* This is a direct corollary of Theorem 5.20 in [33], which establishes this fact for the random graph model $G(n, M)$. □

## 2.4 Binomial coupling

In our coupling constructions we will use the following fact about the coupling of two binomial random variables.

**Lemma 2.18.** *Let $X$ and $Y$ be binomial random variables with parameters $m$ and $r$, where $r \in (0, 1)$ is a constant. Then, for any integer $y > 0$, there exists a coupling $(X, Y)$ such that for a suitable constant $\gamma = \gamma(r) > 0$,*

$$\Pr[X - Y = y] \ge 1 - \frac{\gamma y}{\sqrt{m}}.$$

*Moreover if $y = a\sqrt{m}$ for a fixed constant $a$, then $\gamma a < 1$.*

*Proof.* This lemma is a slight generalization of Lemma 6.7 in [38] and, like that lemma, follows from a standard fact about symmetric random walks. When $y = \Theta(\sqrt{m})$ the result follows directly from Lemma 6.7 in [38], so we assume $y < \sqrt{m}$ which will simplify our calculations.

Let $X_1, \ldots, X_m, Y_1, \ldots, Y_m$ be Bernoulli i.i.d's with parameter $r$. Let $X = \sum_{i=1}^m X_i$, $Y = \sum_{i=1}^m Y_i$, and $D_k = \sum_{i=1}^k (X_i - Y_i)$. We construct a coupling for $(X, Y)$ by coupling each $(X_k, Y_k)$ as follows:

1. If $D_k \neq y$, sample $X_{k+1}$ and $Y_{k+1}$ independently.
2. If $D_k = y$, set $X_{k+1} = Y_{k+1}$.

Clearly this is a valid coupling since $X$ and $Y$ are both binomially distributed.

If $D_k = y$ for any $k \leq m$, then $X - Y = y$. Therefore, $\Pr[X - Y = y] \geq \Pr[M_m \geq y]$ where $M_m = \max\{D_0, \ldots, D_m\}$. Observe that while $D_k \neq y$, $\{D_k\}$ behaves like a (lazy) symmetric random walk. The result then follows from the following fact:

**Fact 2.19.** *Let $\xi_1, \ldots, \xi_m$ be i.i.d such that $\Pr[\xi_i = 1] = \Pr[\xi_i = -1] = w$ and $\Pr[\xi_i = 0] = 1 - 2w$. Let $S_k = \sum_{i=1}^k \xi_i$ and $M_k = \max\{S_1, \ldots, S_k\}$. Then, for any positive integer $y < \sqrt{m}$, there exists a constant $\gamma = \gamma(w) > 0$ such that*

$$\Pr[M_m \geq y] \geq 1 - \frac{\gamma y}{\sqrt{m}}.$$

*Proof.* This is a well-known fact about symmetric random walks, so we just sketch one way of proving it. By the *reflection principle*, $\Pr[M_m \geq y] \geq 2\Pr[S_m > y]$ (see, e.g., [23]) and by the Berry-Esséen inequality, $|\Pr[S_m > k\sqrt{2wm}] - \Pr[N > k]| = O(m^{-1/2})$ where $N$ is a standard normal random variable (see, e.g., [18]). The result follows from the fact that $2\Pr[N > k] \geq 1 - \sqrt{\frac{2}{\pi}}k$. ◻

To complete the proof the lemma, apply Fact 2.19 with $w = r(1 - r)$. ◻

## 2.5 Hitting time estimates

We will require the following hitting time estimates.

**Lemma 2.20.** *Consider a stochastic process $\{Z_t\}$ such that $Z_t \in [-n, n]$ for all $t \geq 0$. Assume $Z_0 > a$ for some $a \in [-n, n]$ and let $T = \min\{t > 0 : Z_t \leq a\}$. Suppose that if $Z_t > a$, then $\mathrm{E}[Z_{t+1} - Z_t | \mathcal{F}_t] \leq -A$, where $A > 0$ and $\mathcal{F}_t$ is the history of the first $t$ steps. If also $\mathrm{E}[T] < \infty$, then $\mathrm{E}[T] \leq 4n/A$.*

*Proof.* Let $Y_t = Z_t^2 - 4nZ_t - 2nAt$. A standard calculation reveals that $\mathrm{E}[Y_{t+1} - Y_t | \mathcal{F}_t] \geq 0$ for all $t < T$; i.e., $\{Y_{t \wedge T}\}$ with $t \wedge T := \min\{t, T\}$ is a submartingale. Moreover, $T$ is a stopping time, $\mathrm{E}[T] < \infty$, and $|Y_{t+1} - Y_t|$ is bounded for all $t \geq 0$. Thus, the optional stopping theorem (see, e.g., [15]) implies

$$5n^2 - 2nA\,\mathrm{E}[T] \geq \mathrm{E}[Y_T] \geq \mathrm{E}[Y_0] \geq -3n^2.$$

Hence, $\mathrm{E}[T] \leq 4n/A$, as desired. ◻

**Corollary 2.21.** *Consider a stochastic process $\{Z_t\}$ such that $Z_t \in [0, n]$ for all $t \geq 0$. Assume $Z_0 \in [a, b]$ for some $a, b \in [0, n]$ and let $T = \min\{t > 0 : Z_t \notin [a, b]\}$. Suppose that if $Z_t \in (a, b)$, then $\mathrm{E}[Z_{t+1} - Z_t \mid \mathcal{F}_t] \leq -A$, where $A > 0$ and $\mathcal{F}_t$ is the history of the first $t$ steps. If also $\mathrm{E}[T] < \infty$, then $T \leq 4\kappa n/A$ and $Z_T \leq a$ with probability at least $1 - \frac{Z_0}{b} - \frac{1}{\kappa}$ for any $\kappa > 0$.*

*Proof.* Proceeding exactly as in the proof Lemma 2.20 we can show that $E[T] \leq 4n/A$. Hence, Markov's inequality implies that $\Pr[T \geq 4\kappa n/A] \leq 1/\kappa$. Moreover, $\{Z_{t\wedge T}\}$ with $t \wedge T := \min\{t, T\}$ is a supermartingale, and thus by the optional stopping theorem we have

$$Z_0 \geq \mathrm{E}[Z_T] \geq b \Pr[Z_T > b].$$

Hence, $\Pr[Z_T < a] \geq 1 - Z_0/b$. The result follows by a union bound. □

**Lemma 2.22.** *Let $\{Z_t\}$ be a stochastic process such that $Z_0 \in [0, \alpha M]$ where $M > 0$ and $\alpha \in (0, 1)$. Suppose that if $Z_t \in [0, M]$, then $Z_{t+1} \leq Z_t - D + Y_{t+1}$ where $D > 0$ and $Y_1, Y_2, \ldots$ are independent random variables satisfying:*

*(i)* $\mathrm{E}[Y_t] \leq \frac{D}{\kappa}$;

*(ii)* $\mathrm{Var}(Y_t) \leq \frac{MD}{C}$; *and*

*(iii)* $|Y_t| \leq \frac{M}{C}$,

*with $\kappa > 1$ and $C > 0$. Then, there exists $T \leq \frac{2\kappa}{\kappa-1}\frac{\alpha M}{D}$ such that $Z_T < 0$ with probability at least $1 - 2\exp\left(-\frac{\alpha\kappa}{5\kappa-1}C\right)$.*

*Proof.* Let $\tilde{T} := \min\{t \geq 0 : Z_t \notin [0, M]\}$, $\hat{T} := \frac{\beta M}{D}$ and $T := \min\{\tilde{T}, \hat{T}\}$, where $\beta := \frac{2\kappa\alpha}{\kappa-1}$. Since

$$\Pr[Z_T \geq 0] = \Pr[Z_T > M] + \Pr[Z_T \in [0, M]], \tag{2.4}$$

it is sufficient to bound from above each term in the right hand side of (2.4). Observe that $Z_t \in [0, M]$ for all $t < T$, and so

$$Z_T \leq Z_0 - DT + S_T \leq (\alpha - \beta)M + S_T,$$

where $S_T := \sum_{i=1}^{T} Y_i$. Hence,

$$\begin{aligned}
\Pr[Z_T > M] \leq \Pr[S_T > (1 - \alpha + \beta)M] &\leq \max_{t \leq \hat{T}} \Pr[S_t > (1 - \alpha + \beta)M] \\
&= \max_{t \leq \hat{T}} \Pr\left[S_t > \frac{\beta M}{\kappa} + (1 + \alpha)M\right] \\
&\leq \max_{t \leq \hat{T}} \Pr[S_t > \mathrm{E}[S_t] + (1 + \alpha)M] \\
&\leq \exp\left(-\frac{(1 + \alpha)^2 M^2}{2\hat{T}\frac{MD}{C} + (1 + \alpha)\frac{M^2}{C}}\right) \\
&\leq \exp\left(-\frac{(1 + \alpha)^2}{2\beta + 1 + \alpha}C\right),
\end{aligned} \tag{2.5}$$

where in the fourth inequality we used the fact that $\mathrm{E}[S_t] \le \frac{Dt}{\kappa} \le \frac{\beta M}{\kappa}$, and the fifth follows from Theorem 2.3.

We next bound $\Pr[Z_T \in [0, M]]$. Note that $Z_T \in [0, M]$ only if $Z_t \in [0, M]$ for all $t \in [0, \hat{T}]$. But if $Z_t \in [0, M]$ for all $t \in [0, \hat{T})$, we have

$$Z_{\hat{T}} \le Z_0 - D\hat{T} + S_{\hat{T}} \le (\alpha - \beta)M + S_{\hat{T}}.$$

Thus, let $\mathcal{A}_T$ be the event that $Z_t \in [0, M]$ for all $t \in [0, \hat{T})$. Then,

$$\begin{aligned}
\Pr[Z_T \in [0, M]] &= \Pr[Z_{\hat{T}} \in [0, M], \mathcal{A}_T] \\
&\le \Pr\left[S_{\hat{T}} \ge (\beta - \alpha)M, \mathcal{A}_T\right] \\
&\le \Pr\left[S_{\hat{T}} \ge (\beta - \alpha)M\right] \\
&= \Pr\left[S_{\hat{T}} \ge \frac{\beta M}{\kappa} + \alpha M\right] \\
&\le \Pr\left[S_{\hat{T}} \ge E[S_{\hat{T}}] + \alpha M\right] \\
&\le \exp\left(-\frac{(\alpha M)^2}{2\hat{T}\frac{MD}{C} + \alpha\frac{M^2}{C}}\right) \\
&\le \exp\left(-\frac{\alpha^2}{2\beta + \alpha}C\right),
\end{aligned} \tag{2.6}$$

where the second to last inequality follows from Theorem 2.3.

Plugging (2.5) and (2.6) into (2.4), we get

$$\begin{aligned}
\Pr[Z_T < 0] \ge 1 - \exp\left(-\frac{\alpha^2}{2\beta + \alpha}C\right) - \exp\left(-\frac{(1+\alpha)^2}{2\beta + 1 + \alpha}C\right) &\ge 1 - 2\exp\left(-\frac{\alpha^2}{2\beta + \alpha}C\right) \\
&= 1 - 2\exp\left(-\frac{\alpha\kappa}{5\kappa - 1}C\right).
\end{aligned}$$

Hence, there exists $\tau \le \frac{2\kappa}{\kappa - 1}\frac{\alpha M}{D}$ such that $Z_\tau < 0$ with probability $1 - 2\exp\left(-\frac{\alpha\kappa}{5\kappa - 1}C\right)$. ☐

**Lemma 2.23.** *Let $\{Z_t\}$ be a stochastic process in the interval $[-M, M]$ such that $Z_0 \in [-A, A]$. Suppose that for $Z_t \in [-A, A]$, we have:*

*(i)* $-C \le \mathrm{E}[Z_{t+1} - Z_t \,|\, \mathcal{F}_t] \le C$,

*(ii)* $\mathrm{Var}(Z_{t+1} \,|\, \mathcal{F}_t) \ge \sigma^2$,

*(iii)* $\Pr[|Z_{t+1} - Z_t| > L \mid \mathcal{F}_t] \le \varepsilon$,

*where $\mathcal{F}_t$ is the history of the first $t$ steps, $A, C, L, M > 0$ and $\varepsilon \in (0, 1)$. Then, there exits $\tau < \frac{4\kappa(A+L)^2}{\sigma^2}$ such that, for any $\kappa > 0$,*

$$\Pr[Z_\tau < -A/2] \ge \frac{A - Z_0}{2A + L} - \frac{4(A + L)}{M} - \frac{1}{\kappa},$$

*provided $\sigma^2 \ge 8\kappa \cdot \max\{\varepsilon M^2, C(A + L)^2/A\}$.*

*Proof.* Let $Y_t = Z_t - A_t$ where $A_t := \sum_{k=1}^{t} \mathrm{E}[Z_k - Z_{k-1} \mid \mathcal{F}_{k-1}]$ for all $t \geq 1$ and $A_0 = 0$; $Y_t + A_t$ is called the Doob decomposition of $\{Z_t\}$, and it is a standard fact that $\{Y_t\}$ is a martingale.

Let $\tilde{T} := \min\{t \geq 0 : Y_t \notin [-A, A]\}$, $\hat{T} := \frac{4\kappa(A+L)^2}{\sigma^2}$ and $T := \tilde{T} \wedge \hat{T} := \min\{\tilde{T}, \hat{T}\}$. Let $W_t = Y_t^2 - \sigma^2 t$. Given $\mathcal{F}_t$, $A_{t+1}$ is deterministic, so $\mathrm{Var}(Y_{t+1} \mid \mathcal{F}_t) = \mathrm{Var}(Z_{t+1} \mid \mathcal{F}_t) \geq \sigma^2$. Hence, $\mathrm{E}[W_{t+1} \mid \mathcal{F}_t] \geq W_t$ whenever $t < T$; i.e., $W_{t \wedge T}$ is a submartingale. Since the stopping time $T$ is bounded (by $\hat{T}$), the optional stopping theorem implies

$$Z_0^2 = W_0 \leq \mathrm{E}[W_T] = \mathrm{E}[Y_T^2] - \sigma^2\,\mathrm{E}[T],$$

and thus $\mathrm{E}[T] \leq \frac{\mathrm{E}[Y_T^2]}{\sigma^2}$. It follows from (i) that $Z_T - C\hat{T} \leq Y_T \leq Z_T + C\hat{T}$, so

$$\mathrm{E}[T] \leq \frac{\mathrm{E}[Z_T^2] + 2C\hat{T}\,\mathrm{E}[|Z_T|] + (C\hat{T})^2}{\sigma^2}. \tag{2.7}$$

Let $\mathcal{H}_{t+1}$ be the event that, given $\mathcal{F}_t$, $|Z_{t+1} - Z_t| \leq L$ and let $\mathcal{H} = \bigcap_{k=1}^{T} \mathcal{H}_k$. By assumption, if $Z_t \in [-A, A]$ then $\Pr[\neg\mathcal{H}_{t+1}] \leq \varepsilon$. Hence,

$$\Pr[\neg\mathcal{H}] = \Pr\left[\bigcup\nolimits_{k=1}^{T} (\neg\mathcal{H}_k)\right] = \Pr\left[\bigcup\nolimits_{k=1}^{T} \left(\neg\mathcal{H}_{k\wedge\tilde{T}}\right)\right] \leq \Pr\left[\bigcup\nolimits_{k=1}^{\hat{T}} \left(\neg\mathcal{H}_{k\wedge\tilde{T}}\right)\right] \leq \varepsilon\hat{T},$$

where the last inequality follows from a union bound. This inequality implies:

$$\begin{aligned} \mathrm{E}[Z_T^2] &\leq \mathrm{E}[Z_T^2 \mid \mathcal{H}] + \varepsilon\hat{T}M^2 \leq (A+L)^2 + \varepsilon\hat{T}M^2, \text{ and} \\ \mathrm{E}[|Z_T|] &\leq \mathrm{E}[|Z_T| \mid \mathcal{H}] + \varepsilon\hat{T}M \leq A + L + \varepsilon\hat{T}M. \end{aligned}$$

Plugging these bounds into (2.7), we get

$$\begin{aligned} \mathrm{E}[T] &\leq \frac{(A+L)^2 + \varepsilon\hat{T}M^2 + 2C\hat{T}(A+L+\varepsilon\hat{T}M) + (C\hat{T})^2}{\sigma^2}. \\ &= \frac{(A+L)^2}{\sigma^2} + \frac{2C\hat{T}(A+L)}{\sigma^2} + \frac{(C\hat{T})^2}{\sigma^2} + \frac{\varepsilon\hat{T}M(M+2C\hat{T})}{\sigma^2} \\ &\leq \frac{(A+L)^2}{\sigma^2} + \frac{(A+L)^2}{\sigma^2} + \frac{(A+L)^2}{4\sigma^2} + \frac{(A+L)^2}{\sigma^2}, \\ &\leq \frac{13(A+L)^2}{4\sigma^2}, \end{aligned} \tag{2.8}$$

where the last three terms in the right-hand side of (2.8) are bounded using the fact that $\sigma^2 \geq 8\kappa \max\{\varepsilon M^2, C(A+L)^2/A\}$. For ease of notation we set $R := \frac{13(A+L)^2}{4\sigma^2}$. Then, Markov's inequality implies

$$\Pr[T \geq \kappa R] \leq \frac{1}{\kappa}. \tag{2.9}$$

We next bound the probability that $Y_T \leq A$. Since the stopping time $T$ is bounded, by optional stopping $\mathrm{E}[Y_T] = Y_0 = Z_0$. Hence,

$$Z_0 = \mathrm{E}[Y_T \mid \mathcal{H}]\Pr[\mathcal{H}] + \mathrm{E}[Y_T \mid \neg\mathcal{H}]\Pr[\neg\mathcal{H}],$$

and

$$\begin{aligned}\mathrm{E}[Y_T \mid \mathcal{H}] &= Z_0 + \Pr[\neg\mathcal{H}](\mathrm{E}[Y_T \mid \mathcal{H}] - \mathrm{E}[Y_T \mid \neg\mathcal{H}]) \\ &\le Z_0 + 2\varepsilon\hat{T}(M + A + L),\end{aligned}$$

since $M + A \ge Y_T \ge -(M + A)$ and $\Pr[\neg\mathcal{H}] \le \varepsilon\hat{T}$. By Markov's inequality, we have

$$\Pr[Y_T \ge A \mid \mathcal{H}] = \Pr[Y_T + A + L \ge 2A + L \mid \mathcal{H}] \le \frac{Z_0 + 2\varepsilon\hat{T}(M + A) + A + L}{2A + L}. \tag{2.10}$$

Observe that

$$\Pr[Y_T \le A] \ge \Pr[Y_T \le A \mid \mathcal{H}]\Pr[\mathcal{H}] \ge \Pr[Y_T < A \mid \mathcal{H}](1 - \varepsilon\hat{T}),$$

so (2.10) together with the facts that $\sigma^2 \ge 8\kappa\varepsilon M^2$ and $L \le 2M$ imply

$$\Pr[Y_T \le A] \ge \frac{A - Z_0}{2A + L} - \frac{2\varepsilon\hat{T}(M + A)}{2A + L} - \varepsilon\hat{T} \ge \frac{A - Z_0}{2A + L} - \frac{4(A + L)}{M}. \tag{2.11}$$

Now, (2.9), (2.11) and a union bound imply that

$$\Pr\left[T < \kappa R,\ Y_T \le A\right] \ge \frac{A - Z_0}{2A + L} - \frac{4(A + L)}{M} - \frac{1}{\kappa}.$$

Since $\hat{T} = \frac{4\kappa(A+L)^2}{\sigma^2} > \kappa R$, we get that $T < \kappa R$ only if $T = \tilde{T}$. Hence,

$$\Pr\left[T < \kappa R,\ Y_T \le A\right] = \Pr\left[\tilde{T} < \kappa R,\ Y_{\tilde{T}} \le A\right] = \Pr\left[\tilde{T} < \kappa R,\ Y_{\tilde{T}} < -A\right].$$

Finally, observe that if $Y_{\tilde{T}} < -A$, then $Z_{\tilde{T}} < -A + C\tilde{T}$ and since $\sigma^2 \ge \frac{8\kappa C(A+L)^2}{A}$, we have that $Z_{\tilde{T}} < -A/2$ as desired. □

**Lemma 2.24.** *Let $\{Z_t\}$ be a stochastic process in the interval $[0, M]$. Suppose that for some $\alpha \in (0, 1)$ and $D > 0$, $\mathrm{E}[Z_{t+1}|\mathcal{F}_t] \le (1 - \alpha)Z_t + D$, where $\mathcal{F}_t$ is the history of the first $t$ steps. If $M > 2D/\alpha$, then there exists $c = c(\alpha)$ and $T \le c\log(\alpha M/2D)$ such that $Z_T < 4D/\alpha$ with probability at least 1/2.*

*Proof.* Let $T = \lceil \log_b(\alpha M/2D) \rceil$, where $b = 2/(2 - \alpha)$. (Note that $T > 0$ since $M > 2D/\alpha$.) If $Z_t \le 2D/\alpha$ for any $t \le T$, we are done. Otherwise,

$$\mathrm{E}[Z_T] \le (1 - \alpha/2)^T M \le 2D/\alpha,$$

and by Markov's inequality $Z_T < 4D/\alpha$ with probability at least $1/2$. Hence, there exists $t \le T$ such that $Z_t < 4D/\alpha$ with probability at least $1/2$, as desired. □

# Chapter 3

# Dynamics for the mean-field random-cluster model

## 3.1 Preliminaries

### 3.1.1 The random-cluster model

Recall from the Introduction that the mean-field random-cluster model exhibits a phase transition at $\lambda = \lambda_c(q)$ (see [7]): in the sub-critical regime $\lambda < \lambda_c$ the largest component is of size $O(\log n)$, while in the super-critical regime $\lambda > \lambda_c$ there is a unique giant component of size $\sim\theta_r n$, where $\theta_r = \theta_r(\lambda, q)$ is the largest $x > 0$ satisfying the equation

$$e^{-\lambda x} = 1 - \frac{qx}{1 + (q-1)x}. \tag{3.1}$$

(Note that, as expected, this equation is identical to (2.2) when $q = 1$, and $\theta_r(\lambda, q) < \beta(\lambda)$ for all $q > 1$.) The following is a more precise statement of this fact.

**Lemma 3.1** ([7]). *Let $G$ be distributed as a mean-field random-cluster configuration where $\lambda > 0$ and $q > 1$ are constants independent of $n$. If $\lambda < \lambda_c$, then $L_1(G) = O(\log n)$ w.h.p. Moreover, if $\lambda > \lambda_c$, then w.h.p. $L_2(G) = O(\log n)$ and $|L_1(G) - \theta_r n| = O(n\omega^{-1}(n))$ for some sequence $\omega(n)$ satisfying $\omega(n) \to \infty$.*

More accurate versions of this result can readily be obtained by combining the techniques from [7] with stronger error bounds for random graph properties [31]. We will use the following version in our proofs which we defer to Section 3.1.2.

**Corollary 3.2.** *If $\lambda > q$, then $|L_1(G) - \theta_r n| = O(n^{8/9})$ w.h.p.*

### 3.1.2 Drift function

As indicated in the Introduction, our analysis relies heavily on understanding the evolution of the size of the largest component under the CM dynamics. To this end, for fixed $\lambda$ and $q$ let $\phi(\theta)$

be the largest $x > 0$ satisfying the equation

$$e^{-\lambda x} = 1 - \frac{qx}{1+(q-1)\theta}. \tag{3.2}$$

Note this equation corresponds to (2.2) for a $G\left(\left(\theta + \frac{1-\theta}{q}\right)n, \lambda/n\right)$ random graph, so

$$\phi(\theta) = \beta\left(\frac{\lambda(1+(q-1)\theta)}{q}\right). \tag{3.3}$$

Thus, $\phi$ is well-defined when $\lambda(1+(q-1)\theta) > q$. In particular, $\phi$ is well-defined in the interval $(\theta_{\min}, 1]$, where $\theta_{\min} = \max\{(q-\lambda)/\lambda(q-1), 0\}$.

We will see in Section 3.2 that for a configuration with a unique "large" component of size $\theta n$, the expected "drift" in the size of the largest component will be determined by the sign of the function

$$f(\theta) = \theta - \phi(\theta). \tag{3.4}$$

When $f(\theta) > 0$ the drift is negative and $f(\theta) < 0$ corresponds to a positive drift. Thus, let

$$\lambda_s = \max\{\lambda \le \lambda_c : f(\theta) > 0 \ \ \forall \theta \in (\theta_{\min}, 1]\} \text{ and,}$$

$$\lambda_S = \min\{\lambda \ge \lambda_c : f(\theta)(\theta - \theta_r) > 0 \ \ \forall \theta \in (\theta_{\min}, 1]\}.$$

Intuitively, $\lambda_s$ and $\lambda_S$ are the maximum and minimum values, respectively, of $\lambda$ for which the drift in the size of the largest component is always in the required direction (i.e., towards 0 in the sub-critical case and towards $\theta_r n$ in the super-critical case). The following lemma, which we will prove shortly, reveals basic information about the quantities $\lambda_s$ and $\lambda_S$.

**Lemma 3.3.** *For $q \le 2$, $\lambda_s = \lambda_c = \lambda_S = q$; and for $q > 2$, $\lambda_s < \lambda_c < \lambda_S = q$.*

For integer $q \ge 3$, $\lambda_s$ corresponds to the threshold $\beta_s$ in the mean-field $q$-state Potts model at which the local (Glauber) dynamics undergoes an exponential slowdown [13]. In fact, a change of variables reveals that $\lambda_s = 2\beta_s$ for the specific mean-field Potts model normalization in [13].

In Figure 3.1 we sketch $f$ in its only two qualitatively different regimes: $q \le 2$ and $q > 2$. The following are useful facts about the functions $\phi$ and $f$ which in most cases follow easily from their definitions.

**Fact 3.4.**

*(i)* $\hat{\theta} \in (\theta_{min}, 1]$ *is a fixed point of* $\phi$ *if and only if* $\hat{\theta}$ *is a solution of (3.1).*

*(ii)* $\phi$ *is continuous, differentiable, strictly increasing and strictly concave in* $(\theta_{min}, 1]$.

*(iii)* $\frac{2(q-1)}{q} > \phi'(\theta) > \frac{q-1}{q}$ *for all* $\theta \in (\theta_{min}, 1]$.

**Fact 3.5.**

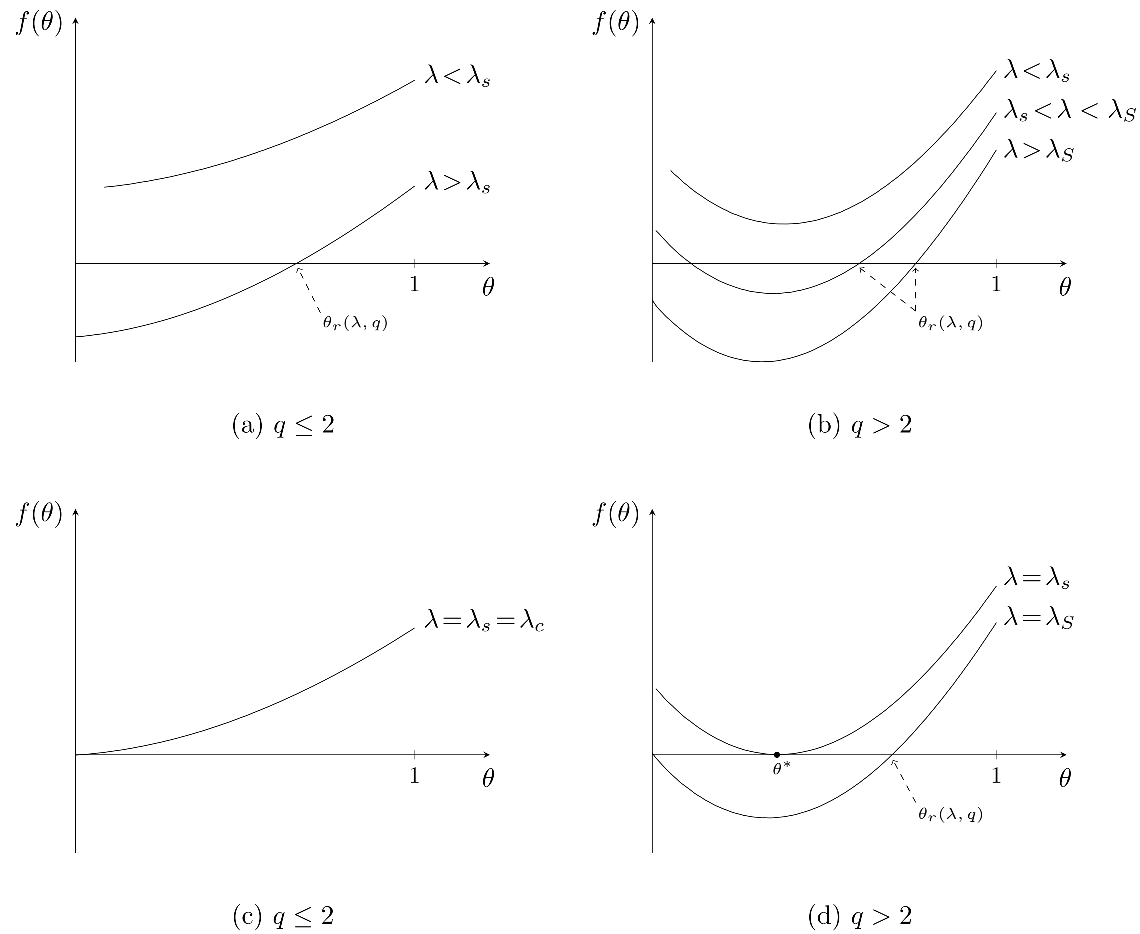


Figure 3.1: Sketches of the function $f$.

*(i)* $f$ *is continuous, differentiable and strictly convex in* $(\theta_{min}, 1]$.

*(ii)* $f(\theta_r) = 0$, $f(1) > 0$ *and* $f'(\theta) < 1/q$ *for all* $\theta \in (\theta_{min}, 1]$.

*(iii) Let* $f(\theta^+_{min}) = \lim_{\theta \to \theta_{min}} f(\theta)$; *then* $\mathrm{sgn}(f(\theta^+_{min})) = \mathrm{sgn}(q - \lambda)$.

Observe that if $\hat{\theta}$ is a zero of $f$, then $\hat{\theta}$ is a fixed point of $\phi$ and consequently a root of equation (3.1). Lemma 2.5 from [7] dissects the roots of equation (3.1) and hence identifies the roots of $f$ in $(\theta_{\min}, 1]$.

**Fact 3.6.** *The roots of the function* $f$ *in* $(\theta_{min}, 1]$ *are given as follows:*

- *When* $q \le 2$*:*

   - *if $\lambda \le \lambda_c$, $f$ has no positive roots; and*
   - *if $\lambda > \lambda_c$, $f$ has a unique positive root.*

- *When $q > 2$, there exists $\lambda_{min} < \lambda_c$ such that:*
   - *if $\lambda < \lambda_{min}$, $f$ has no positive roots;*
   - *if $\lambda = \lambda_{min}$, $f$ has a unique positive root;*
   - *if $\lambda_{min} < \lambda < q$, $f$ has exactly two positive roots; and*
   - *if $\lambda \ge q$, $f$ has a unique positive root.*

The following three lemmas provide bounds for the drift of the size of the largest component under CM steps and are thus crucial for establishing its mixing time.

**Lemma 3.7.** *If $\lambda < \lambda_s$, then for all $\theta \in (\theta_{min}, 1]$ there exists a constant $\delta > 0$ such that $f(\theta) \ge \delta$.*

**Lemma 3.8.** *For $\lambda = \lambda_s$, let $\theta^*$ be the unique zero of the function $f$. Then:*

*(i) There exists positive constants $\delta_1, \delta_2, \delta_3$ such that for all $\theta \in (\theta_{min}, 1]$, we have*

$$\delta_1(\theta - \theta^*)^2 - \delta_2(\theta - \theta^*)^3 \le f(\theta) \le \delta_3(\theta - \theta^*)^2 + \delta_2(\theta - \theta^*)^3.$$

*(ii) The function $f$ is decreasing in the interval $(\theta_{min}, \theta^*)$ and increasing in $(\theta^*, 1)$.*

*(iii) If $|\theta - \theta^*| \ge \xi_1$ for some positive constant $\xi_1$, then there exists a positive $\xi_2$ such that $f(\theta) > \xi_2$.*

**Lemma 3.9.** *Let $\lambda > \lambda_S = q$. Then,*

*(i) For all and $\theta \in (\theta_{min}, 1]$, if $\theta > \theta_r$, then $\theta \ge \phi(\theta) \ge \theta_r$ and if $\theta < \theta_r$, then $\theta \le \phi(\theta) \le \theta_r$.*

*(ii) For all fixed $\varepsilon$ such that $\theta_{min} < \varepsilon < \theta_r$, there exists a constant $\delta = \delta(\varepsilon) \in (0, 1)$ such that $\delta|\theta - \theta_r| \le |\phi(\theta) - \theta|$ for all $\theta \in [\varepsilon, 1]$.*

*Parts (i) and (ii) also hold for $\lambda = \lambda_S = q$ when $q > 2$.*

Finally, the following fact will also be helpful.

**Fact 3.10.** *If $\lambda > q$, then $\theta_r > 1 - q/\lambda$.*

We provide next the proofs of all the facts and lemmas in this section.

*Proof of Lemma 3.3.* Since $f(1) > 0$, by continuity $f$ is strictly positive in $(\theta_{\min}, 1]$ if and only if $f$ has no roots in $(\theta_{\min}, 1]$. When $q \le 2$, by Fact 3.6, if $\lambda \le \lambda_c$ then $f$ has no roots in $(\theta_{\min}, 1]$, and if $\lambda > \lambda_c$ then $f$ has a unique root in $(\theta_{\min}, 1]$; thus, $\lambda_s = \lambda_c = q$. When $q > 2$, by Fact 3.6, $\lambda_s = \lambda_{\min} < \lambda_c$.

If $\lambda > q$, then $f(\theta_{\min}^+) < 0$ and Fact 3.6 implies that $f$ has a unique root in $(\theta_{\min}, 1]$. Hence $f$ is negative in $(\theta_{\min}, \theta_r)$ and positive in $(\theta_r, 1]$ and then $\lambda_S \le q$. For $q \le 2$ this readily implies $\lambda_s = \lambda_c = \lambda_S = q$. For $q > 2$, if $q > \lambda > \lambda_c$, Fact 3.6 implies that $f$ has exactly two positive roots in $(\theta_{\min}, 1]$. Recall that $f(\theta_r) = 0$ and let $\theta^*$ be the other root of $f$ in $(\theta_{\min}, 1]$; by the definition of $\theta_r$, $\theta^* < \theta_r$. Moreover, $f(1) > 0$ and $f(\theta_{\min}^+) > 0$ since $q > \lambda$. Therefore, $f$ is positive in $(0, \theta^*) \cup (\theta_r, 1]$ and negative in $(\theta^*, \theta_r)$. If $\theta < \theta^*$, then $f(\theta)(\theta - \theta_r) < 0$; thus, $\lambda_S = q$. □

*Proof of Fact 3.4.* Obviously any fixed point of $\phi$ is also a solution of (3.1). For the other direction, consider the injective function $h(x) = \frac{x}{1-e^{-\lambda x}}$; if $\hat{\theta}$ is a root of equation (3.1), then $h(\hat{\theta}) = h(\phi(\hat{\theta}))$ and $\phi(\hat{\theta}) = \hat{\theta}$.

By differentiating both sides of (3.2),

$$\phi'(\theta) = \frac{q-1}{q} \cdot \frac{(1 - e^{-\lambda\phi(\theta)})^2}{1 - e^{-\lambda\phi(\theta)} - \lambda\phi(\theta)e^{-\lambda\phi(\theta)}}$$

which implies that $\phi$ is differentiable and continuous. Since $e^{-\lambda\phi(\theta)} > 1 - \lambda\phi(\theta)$, then $\phi'(\theta) > \frac{q-1}{q}$, which establishes the lower bound in part (iii), and also implies that $\phi$ is strictly increasing. Moreover, it is straightforward to check that $\frac{(1-e^{-x})^2}{1-e^{-x}-xe^{-x}} < 2$ for all $x > 0$; hence $\phi'(\theta) < \frac{2(q-1)}{q}$, which completes the proof of part (iii).

Finally to show that $f$ is concave, consider the function

$$g(x) = \frac{qx}{(q-1)(1-e^{-\lambda x})} - \frac{1}{q-1} - x.$$

By solving for $\theta$ in (3.2), observe that $g(\phi(\theta)) = \theta - \phi(\theta)$ for $\theta \in (\theta_{\min}, 1]$. Therefore,

$$\phi''(\theta) = -g''(\phi(\theta))\phi'(\theta)(1 + g'(\phi(\theta)))^{-2}$$

and a straightforward calculation shows that $g'' > 0$ in $(0, 1]$. Consequently, $\phi$ is strictly concave in $(\theta_{\min}, 1]$. □

*Proof of Fact 3.5.* Parts (i) and (ii) follow immediately from Fact 3.4. For Part (iii), observe that when $\lambda > q$, $\theta_{\min} = 0$ and the function $\phi$ is defined at 0; thus, $f(\theta_{\min}^+) = -\phi(0) < 0$. When $\lambda < q$, $\theta_{\min} = (q - \lambda)/\lambda(q-1)$ and by continuity, $\lim_{\theta \to \theta_{\min}} \phi(\theta) = 0$; hence, $f(\theta_{\min}^+) = \theta_{\min} > 0$. □

*Proof of Lemma 3.7.* If $\lambda < \lambda_s$, then $f(\theta) > 0$ for all $\theta \in (\theta_{\min}, 1]$ by definition. Also, $f$ is continuous in $(\theta_{\min}, 1]$ and $f(\theta_{\min}^+) > 0$; thus, $f$ must attain a minimum value $\delta > 0$ in $(\theta_{\min}, 1]$ which implies the result. □

*Proof of Lemma 3.8.* Note that $f(\theta^*) = 0$ and $f'(\theta^*) = 0$, since $f$ is differentiable and $\theta^*$ is the unique global minimum of $f$. Hence, using Taylor's expansion

$$\frac{f''(\theta^*)}{2}(\theta-\theta^*)^2 + \frac{\sup_{\hat{\theta}} |f'''(\hat{\theta})|}{6}(\theta-\theta^*)^3 \ \ge\ f(\theta) \ \ge\ \frac{f''(\theta^*)}{2}(\theta-\theta^*)^2 - \frac{\sup_{\hat{\theta}} |f'''(\hat{\theta})|}{6}(\theta-\theta^*)^3. \quad (3.5)$$

To obtain the bounds for $f$ in part (i), we provide absolute bounds for $f''$ and $f'''$. For this let

$$h(x) = \frac{qx}{\lambda(q-1)(1-e^{-x})} - \frac{x}{\lambda} - \frac{1}{q-1}.$$

By solving for $\theta$ in (3.2), we get $h(\lambda\phi(\theta)) = f(\theta) = \theta - \phi(\theta)$ and $\phi'(\theta) = \frac{1}{1+\lambda h'(\lambda\phi(\theta))}$. Hence,

$$f''(\theta) = -\phi''(\theta) = \frac{\lambda^2\phi'(\theta)h''(\lambda\phi(\theta))}{(1+\lambda h'(\lambda\phi(\theta)))^2}.$$

A direct calculation of the derivatives of the function $\frac{x}{1-e^{-x}}$ reveals the following facts about the derivatives of $h$:

(i) $h'$ is strictly increasing in $(0,+\infty)$ and $h' > \frac{q}{2\lambda(q-1)} - \frac{1}{\lambda}$.

(ii) $h''$ is strictly decreasing in $(0,+\infty)$ and $\frac{q}{6\lambda(q-1)} > h'' > 0$; and

(iii) $h'''$ is negative and bounded in $(0,+\infty)$;

As a result, $h''(\lambda\phi(\theta)) > h''(\lambda)$, $h'(\lambda\phi(\theta)) < h'(\lambda)$ and $1 + \lambda h'(\lambda\phi(\theta)) > 0$. Moreover, by Fact 3.4, we have $\frac{2(q-1)}{q} > \phi'(\theta) > \frac{q-1}{q}$. Thus,

$$\frac{4\lambda(q-1)^2}{3q^2} \geq f''(\theta) \geq \frac{\lambda^2(q-1)h''(\lambda)}{q(1+\lambda h'(\lambda))^2} > 0.$$

Hence, $\delta_1 \geq f'' \geq \delta_3 > 0$ for a suitable constants $\delta_1$ and $\delta_3$. A similar argument reveals that $|f'''| \leq \delta_2$ for some constant $\delta_2 > 0$. Plugging these bounds into (3.5) and adjusting the constants we obtain part (i).

For part (ii), observe that $f'' > 0$ and thus $f'$ is strictly increasing. Since $f'(\theta^*) = 0$, we have $f'(\theta) < 0$ when $\theta \in (\theta_{\min}, \theta^*)$ and $f'(\theta) > 0$ when $\theta \in (\theta^*, 1)$. Finally for part (iii), observe that if $\theta > \theta^*$, then $f(\theta) > f(\theta^* + \xi_1)$ since $f$ is increasing in this interval; hence, part (iii) follows by taking $\xi_2 = f(\theta^* + \xi_1)$. The same argument works when $\theta < \theta^*$. □

*Proof of Lemma 3.9.* Part (i) follows from the definition of $\lambda_S$ and the fact that $\phi$ is increasing in $(\theta_{\min}, 1]$. For the second part, note that he function $f$ is continuous, differentiable and convex in $(\theta_{\min}, 1]$, so it lies above all of its tangents. From part (i) we know that $\varepsilon \leq \phi(\varepsilon)$, since $\varepsilon < \theta_r$ and, by Fact 3.6, $f$ has a unique positive root in $(\theta_{\min}, 1]$ when either $\lambda > q$ or when $\lambda = q > 2$. Hence, $\varepsilon < \phi(\varepsilon)$ and $f(\varepsilon) < 0$.

Let $T$ be the line tangent to $f$ at $\theta_r$. Observe that $f'(\theta_r) > 0$ since $f$ is convex in $(\theta_{\min}, 1]$ and $f(\varepsilon) < 0$. Let $M := \min\{f'(\theta_r), -f(\varepsilon)/(\theta_r - \varepsilon)\}$; by Fact 3.5, $f' < 1/q$ and so $M \in (0, 1/q]$. Consider the line $S(\theta) = \frac{M}{2}(\theta - \theta_r)$ and the line $R$ going through the points $(0, f(\varepsilon))$ and $(\theta_r, 0)$. The slope of $R$ is $-f(\varepsilon)/(\theta_r - \varepsilon)$, and the lines $S$, $R$ and $T$ intersect at $(\theta_r, 0)$. Therefore, $S$ lies above $R$ in $(0, \theta_r)$ and below $T$ in $(\theta_r, 1]$. By convexity, $f$ lies below $R$ in $(0, \theta_r)$ and above $T$ in $(\theta_r, 1]$. Thus, $S$ lies above $f$ in $(0, \theta_r)$ and below $f$ in $(\theta_r, 1]$. Therefore, if $\theta < \theta_r$ then $\frac{M}{2}(\theta - \theta_r) > \theta - \phi(\theta)$ and if $\theta > \theta_r$ then $\frac{M}{2}(\theta - \theta_r) < \theta - \phi(\theta)$. Part (ii) then follows by taking $\delta = M/2$. □

*Proof of Fact 3.10.* By solving for $\lambda$ in (3.1), it is sufficient to show that

$$q > \frac{1-x}{x} \ln \left( \frac{1+(q-1)x}{1-x} \right) = h(x)$$

for $x \in [0,1]$. A straightforward calculation shows that $h$ is decreasing in $(0,+\infty)$ and that $\lim_{x \to 0} h(x) = q$. □

Finally, we can use the results in this subsection to prove Corollary 3.1 stated in the previous subsection.

*Proof of Corollary 3.2.* By Lemma 3.2 in [7], $L_2(G) = o(n^{3/4})$ w.h.p. Conditioning on this event, independently color each component of $G$ red with probability $1/q$. Let $L_r$ denote the size of the largest red component and $n_r$ the total number of red vertices.

Let $\Gamma_\theta$ be the intersection of the events that $\mathcal{L}(G)$ is colored red and $L_1(G) = \theta n$ where $\theta n \in \mathbb{N}$. Observe that $\Pr[L_r = \theta n \mid \Gamma_\theta] = 1$, and by Hoeffding's inequality $\Pr[n_r \in J \mid \Gamma_\theta] = 1 - O(n^{-2})$ where $J := \left[ \left(\theta + \frac{1-\theta}{q}\right) n - \xi, \left(\theta + \frac{1-\theta}{q}\right) n + \xi \right]$ with $\xi = \sqrt{n^{7/4} \log n}$. Putting these two facts together,

$$\frac{1}{2q} \Pr[L_1(G) = \theta n] \leq \Pr[n_r \in J \mid \Gamma_\theta] \Pr[\Gamma_\theta] \leq \Pr[L_r = \theta n, n_r \in J].$$

By Lemma 3.1 in [7], conditioned on the red vertex set, the red subgraph is distributed as a $G(n_r, p)$ random graph, so

$$\frac{1}{2q} \Pr[L_1(G) = \theta n] \leq \sum_{m \in J} \Pr[L_r = \theta n | n_r = m] \Pr[n_r = m] \leq \max_{m \in J} \Pr[\ell(m) = \theta n]$$

where $\ell(m)$ is distributed as the size of the largest component of a $G(m, p)$ random graph. Note that for $m \in J$ the random graph $G(m, p)$ is super-critical because $\lambda > q$. Since $\xi = \sqrt{n^{7/4} \log n}$, by (3.3) and Lemma 2.7 with $A = \sqrt{n^{3/4} \log n}$, $\Pr[|\ell(m) - \phi(\theta)n| > 2\xi] = O(n^{-2})$. Since $\lambda > q = \lambda_S$, Lemma 3.9(ii) implies that there exists a constant $\delta \in (0,1)$ such that $|\theta - \phi(\theta)| > \delta |\theta - \theta_r|$. Thus, if $|\theta - \theta_r| n > n^{8/9}$, then $\Pr[L_1(G) = \theta n] = O(n^{-2})$. The result follows by a union bound over all the positive integer values of $\theta n$ such that $|\theta - \theta_r| n > n^{8/9}$ and $\theta n \leq n$. □

## 3.2 Mixing time bounds: proof organization and notation

The following theorem gives our bounds for the mixing time of the CM dynamics in the mean-field random-cluster model.

**Theorem 3.11.** *Consider the CM dynamics for the mean-field random-cluster model with parameters $p = \lambda/n$ and $q$ where $\lambda > 0$ and $q > 1$ are constants independent of $n$. Then,*

*(i) If $\lambda < \lambda_s$, then $\tau_{\text{mix}} = \Theta(\log n)$.*

(ii) *If $\lambda = \lambda_s$ and $q > 2$, then $\tau_{\rm mix} = \Theta(n^{1/3})$.*

(iii) *If $\lambda \in (\lambda_s, \lambda_S)$ and $q > 2$, then $\tau_{\rm mix} = e^{\Omega(\sqrt{n})}$.*

(iv) *If $\lambda = \lambda_S$ and $q > 2$, then $\tau_{\rm mix} = \Theta(\log n)$.*

(v) *If $\lambda > \lambda_S$, then $\tau_{\rm mix} = \Theta(\log n)$.*

Since $\lambda_s = \lambda_c = \lambda_S$ for $1 < q \le 2$ and $\lambda_s < \lambda_c < \lambda_S$ for $q > 2$ (see Lemma 3.3), Theorem 3.11 implies Theorems 1.1 and 1.2 from the Introduction.

The mixing time upper bounds in Theorem 3.11 are proved in Sections 3.3, 3.4, 3.5 and 3.6. All the lower bounds, including those in part (iii) are derived in Section 3.7. The proof of the upper bounds is organized as follows. Section 3.3 deals with the $\lambda < \lambda_s$ case; i.e., the sub-critical regime and part (i) of the theorem. There we also develop a number of tools that will be reused in the proofs in Sections 3.4, 3.5 and 3.6. In Section 3.4 we establish the upper bounds for the super-critical regime $\lambda > \lambda_S = q$; this corresponds to part (v) of the theorem. Finally, in Sections 3.5 and 3.6 we treat the boundary points of the critical window $(\lambda_s, \lambda_S)$ for $q > 2$; this establishes parts (ii) and (iv) of Theorem 3.11.

The bounds for the mixing time of the mean-field heat-bath dynamics in Theorem 1.3 from the Introduction are derived in Section 3.8 as a by product of Theorem 3.11.

We now introduce some notation that will be used throughout the rest of the chapter. We call a component *large* if it contains at least $2n^{11/12}$ vertices; otherwise it is *small*. Following Section 2.3, we will use $I(X_t)$ for the number of isolated vertices in $X_t$, $\mathcal{L}(X_t)$ for the largest component in $X_t$ and $L_j(X_t)$ for the *size* of the $j$-th largest component of $X_t$. (Thus, $L_1(X_t) = |\mathcal{L}(X_t)|$.) For convenience, we will sometimes write $\theta_t n$ for $L_1(X_t)$. Also, we will use $\mathcal{E}_t$ for the event that $\mathcal{L}(X_t)$ is activated, and $A_t$ for the number of activated vertices at time $t$.

## 3.3 Mixing time upper bounds: the sub-critical case

In this section we prove our mixing time upper bound for the CM dynamics when $0 < \lambda < \lambda_s$ and $q > 1$. This regime comprises the entire sub-critical regime for $1 < q \le 2$. This is not true for $q > 2$, since $\lambda_s < \lambda_c$ in this case (see Lemma 3.3); the rest of the sub-critical regime when $q > 2$, i.e., $\lambda_s \le \lambda < \lambda_c$, is analyzed in Sections 3.5 and 3.7.

We prove the following theorem.

**Theorem 3.12.** *Let $q > 1$ and $0 < \lambda < \lambda_s$. Then, the mixing time of the CM dynamics is $O(\log n)$.*

We start by describing the main ideas in the proof of Theorem 3.12. The mixing time is bounded by constructing a coupling of the CM dynamics and bounding its coupling time (see (2.1)). The coupling will have a "burn-in" period where both copies of the chain evolve independently; the "burn-in" period consists of two phases. The goal of the first phase will be to reach a configuration with at most one large component. Using the $\mathcal{G}_{n,p}$ fact we proved in Lemma 2.5, it is straightforward to check that this only takes $O(\log n)$ steps. The second phase of the burn-in tracks the evolution of the largest component in the configuration. Using the fact that the function $f$, as defined in (3.4), is *strictly positive* for $\lambda < \lambda_s$ (see Figure 3.1), we can show that the size of the large component has a *strictly negative drift.* Moreover, this drift is $\Omega(n)$ in magnitude; thus, after only $O(1)$ steps, it is unlikely that the large component is still present in the configuration; in fact, we will show that every component have size $O(\log n)$ with probability $\Omega(1)$.

Once the burn-in period concludes, the coupling starts off from two configurations whose largest components have size $O(\log n)$. We also assume that the number of isolated vertices in each starting configuration is $\Omega(n)$, a property that is satisfied after the burn-in period w.h.p. Our coupling will be a composition of two couplings. The first one is designed to quickly reach a pair of configurations with the same component structure. This coupling attempts to activate the same number of vertices from each configuration in each step. If it succeeds for $O(\log n)$ consecutive steps, then the two final configurations will have the same component structure with probability $\Omega(1)$. In order to activate the same number of vertices from each configuration, the idea is to first couple the activation of the components of size two or more in a way that minimizes the discrepancy in the number of active vertices; then we attempt to correct this discrepancy by coupling the activation of the isolated vertices using the binomial coupling from Section 2.4. Since, to some extent, we can control the size of this discrepancy because all the components are small, we are also able to control the probability of success of the binomial coupling. The final step is a straightforward coupling that starts from two configurations with the same component structure and takes $O(\log n)$ steps to couple them. The reminder of this section fleshes out the above proof sketch.

Theorem 3.12 is a consequence of the following lemmas.

**Lemma 3.13.** *Let $q > 1$ and $\lambda > 0$. For any starting random-cluster configuration $X_0$, there exists $T = O(\log n)$ such that $X_T$ has at most one large component with probability $\Omega(1)$.*

**Lemma 3.14.** *Let $q > 1$ and $0 < \lambda < \lambda_s$. If $X_0$ has at most one large component, then there exists $T = O(\log n)$ such that $L_1(X_T) = O(\log n)$ and $I(X_T) = \Omega(n)$ with probability $\Omega(1)$.*

**Lemma 3.15.** *Let $q > 1$ and $0 < \lambda < q$. Let $X_0$ be a random-cluster configuration such that $L_1(X_0) = O(\log n)$ and $I(X_0) = \Omega(n)$. Suppose the same holds for $Y_0$. Then, there exists a coupling of the CM steps such that $X_T$ and $Y_T$ have the same component structure after $T = O(\log n)$ steps with probability $\Omega(1)$.*

**Lemma 3.16.** *Let $q > 1$ and $\lambda > 0$. Let $X_0$ and $Y_0$ be two random-cluster configurations with the same component structure. Then, there exists a coupling of the CM steps such that after $T = O(\log n)$ steps $X_T = Y_T$ w.h.p.*

*Proof of Theorem 3.12.* Consider two copies $\{X_t\}$ and $\{Y_t\}$ of the CM dynamics starting from two arbitrary configurations $X_0$ and $Y_0$. We design a coupling $(X_t, Y_t)$ of the CM steps and show that $\Pr[X_T = Y_T] = \Omega(1)$ for some $T = O(\log n)$; the result then follows from (2.1). The coupling consists of four phases, and each phase is analyzed in one of the lemmas above.

In the first phase $\{X_t\}$ and $\{Y_t\}$ are run independently. Lemma 3.13 establishes that after $O(\log n)$ steps $\{X_t\}$ and $\{Y_t\}$ each have at most one large component with probability $\Omega(1)$. In the second phase, $\{X_t\}$ and $\{Y_t\}$ also evolve independently. Conditioned on the success of the first phase, Lemma 3.14 implies that after $O(\log n)$ additional steps, with probability $\Omega(1)$, the largest components in $\{X_t\}$ and $\{Y_t\}$ have sizes $O(\log n)$ and each have $\Omega(n)$ isolated vertices.

In the third phase, $\{X_t\}$ and $\{Y_t\}$ are coupled to obtain two configurations with the same component structure. By Lemma 3.15, there is a coupling that, conditioned on a successful conclusion of the second phase, succeeds in reaching two configurations with the same component structure with probability $\Omega(1)$ after $O(\log n)$ steps. The last phase uses the coupling provided by Lemma 3.16. Putting all this together, we have that there exists a coupling $(X_t, Y_t)$ such that, after $T = O(\log n)$ steps, $X_T = Y_T$ with probability $\Omega(1)$. ◻

Lemma 3.13 is proved in Section 3.3.1, Lemma 3.14 in Section 3.3.2, Lemma 3.15 in Section 3.3.4 and Lemma 3.16 in Section 3.3.5. Note that Lemmas 3.13, 3.15 and 3.16 are true for more general values of $\lambda$; i.e., not only for $\lambda < \lambda_s$. This will be useful later when we analyze the mixing time in other regimes.

### 3.3.1 Proof of Lemma 3.13

*Proof of Lemma 3.13.* Let $B_t$ be the number of *new* large components created at time $t$. If $A_t < 2n^{11/12}$, then $B_t = 0$. Together with Lemma 2.5 this implies that for all $a \in [0, n]$, $\Pr[B_t > 1 | X_t, A_t = a] = O(a^{-1/12})$. Thus,

$$\begin{aligned} \mathrm{E}[B_t|X_t] &= \sum_{a=0}^{n} \mathrm{E}[B_t|X_t, A_t = a] \Pr[A_t = a|X_t] \\ &\le \sum_{a=0}^{n} \left( \Pr[B_t \le 1|X_t, A_t = a] + \frac{a}{2n^{11/12}} \Pr[B_t > 1|X_t, A_t = a] \right) \Pr[A_t = a|X_t] \\ &\le \kappa, \end{aligned}$$

where $\kappa > 0$ is a constant. Let $K_t$ be the number of large components in $X_t$ and let $C_t$ be the number of activated large components at time $t$. Then,

$$\mathrm{E}[K_{t+1}|X_t] = K_t - \mathrm{E}[C_t|X_t] + \mathrm{E}[B_t|X_t] \le K_t - \frac{K_t}{q} + \kappa.$$

Hence, Lemma 2.24 implies that $K_T \le 4\kappa q$ with probability $\Omega(1)$ for some $T = O(\log n)$. If at time $T$ the remaining $K_T$ large components become active, then $K_{T+1} \le 1$ w.h.p. by Lemma 2.5. All $K_T$ components become active simultaneously with probability at least $q^{-4\kappa q}$ and thus $K_{T+1} \le 1$ with probability $\Omega(1)$, as desired. ◻

### 3.3.2 Proof of Lemma 3.14

For ease of notation we set $\Theta_s := \theta_{\min}$. (Recall from Section 3.1.2 that $\theta_{\min} = (q - \lambda)/\lambda(q - 1)$ when $\lambda < q$.) Also, let $\varepsilon > 0$ be a small constant (independent of $n$) to be chosen later and let

$\xi(r) = \sqrt{2nr \log n}$. The following preliminary facts, which we prove in Section 3.3.3, are used in the proof of Lemma 3.14.

**Fact 3.17.** *If* $0 < \lambda < q$ *and* $L_2(X_t) \le r < 2n^{11/12}$*, then for sufficiently large* $n$*:*

*(i) If* $\mathcal{L}(X_t)$ *is inactive, then with probability* $1 - O(n^{-2})$

$$A_t \in \hat{J}_t := \left[\frac{n - L_1(X_t)}{q} - \xi(r), \frac{n - L_1(X_t)}{q} + \xi(r)\right].$$

*Moreover, if* $A_t \in \hat{J}_t$ *then* $G(A_t, p)$ *is sub-critical and all new components in* $X_{t+1}$ *have size* $O(\log n)$ *(resp.,* $O(\sqrt{n})$*) with probability* $1 - O(n^{-1})$ *(resp.,* $1 - O(n^{-2})$*).*

*(ii) If* $\mathcal{L}(X_t)$ *is active, then with probability* $1 - O(n^{-2})$

$$A_t \in J_t := \left[L_1(X_t) + \frac{n - L_1(X_t)}{q} - \xi(r), L_1(X_t) + \frac{n - L_1(X_t)}{q} + \xi(r).\right] \tag{3.6}$$

*Moreover, if* $A_t \in J_t$ *and* $L_1(X_t) \ge (\Theta_s + \varepsilon)n$*, then* $G(A_t, p)$ *is super-critical and the second largest new component has size* $O(\log n)$ *with probability* $1 - O(n^{-1})$*.*

*(iii) If* $\mathcal{L}(X_t)$ *is active and* $L_1(X_t) \le (\Theta_s - \varepsilon)n$*, then the largest new component has size* $O(\log n)$ *with probability* $1 - O(n^{-1})$*.*

*(iv) If there is no large component in* $X_t$*, then* $L_1(X_{t+1}) = O(\log n)$ *with probability* $1 - O(n^{-1})$*.*

**Fact 3.18.** *If* $0 < \lambda < q$ *and* $X_0$ *has a unique large component, then* $L_2(X_t) < 2n^{11/12}$ *for all* $0 \le t \le T$ *w.h.p. for any* $T = O(\log n)$*.*

**Fact 3.19.** *Suppose that* $0 < \lambda < q$ *and that* $X_t$ *has at most one large component. Then for sufficiently large* $n$*:*

*(i) If* $L_1(X_t) \ge (\Theta_s + \varepsilon)n$*, then* $\mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t] \le \phi(\theta_t)n + 3n^{1/4}$ *and*

$$-\frac{f(\theta_t)n}{q} - 2n^{1/4} \le \mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \le -\frac{f(\theta_t)n}{q} + 2n^{1/4}.$$

*(The function* $f$ *was defined in Section 3.1.2; see (3.4)).*

*(ii) If* $X_t$ *is such that* $\theta_t \in (\Theta_s - \varepsilon, \Theta_s + \varepsilon)$*,*

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \le -\frac{f(\Theta_s + \varepsilon)n}{q} + \frac{2\varepsilon n}{q} + 2n^{1/4}.$$

**Fact 3.20.** *If* $0 < \lambda < q$ *and* $X_0$ *has a unique large component such that* $L_1(X_0) \le (\Theta_s - \varepsilon)n$*, then there exists* $T = O(\log n)$ *such that* $L_1(X_T) = O(\log n)$ *with probability* $\Omega(1)$*.*

We are now ready to prove Lemma 3.14.

*Proof of Lemma 3.14.* By Fact 3.18 we may assume throughout the proof that there is at most one large component. If $\theta_t \geq \Theta_s + \varepsilon$, by Fact 3.19(i), we have

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \leq -\frac{f(\theta_t)n}{q} + O(n^{1/4})$$

Since $\lambda < \lambda_s$, Lemma 3.7 implies that there exists a constant $\delta > 0$ such that $f(\theta_t) \geq \delta$. Hence,

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \leq -\frac{\delta n}{q} + O(n^{1/4}). \tag{3.7}$$

(Note that Lemma 3.7 only holds when $\lambda < \lambda_s$; this is why Lemma 3.14 only holds for $\lambda$ in this regime.)

Similarly, if $\theta_t \in (\Theta_s - \varepsilon, \Theta_s + \varepsilon)$, Fact 3.19(ii) and Lemma 3.7 imply

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \leq -\frac{f(\Theta_s + \varepsilon)n}{q} + \frac{2\varepsilon n}{q} + O(n^{1/4}) \leq -\frac{\delta n}{q} + \frac{2\varepsilon n}{q} + O(n^{1/4}).$$

By choosing $\varepsilon$ sufficiently small, we see that there exists a constant $\gamma > 0$ such that, if $\theta_t > \Theta_s - \varepsilon$, then

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \leq -\gamma n.$$

Let $\tau = \min\{t > 0 : L_1(X_t) \leq (\Theta_s - \varepsilon)n\}$. Note that $\mathrm{E}[\tau] < \infty$, and thus Lemma 2.20 implies that $\mathrm{E}[\tau] \leq 4/\gamma$. By Markov's inequality $\Pr[\tau > 8/\gamma] \leq 1/2$, so $L_1(X_T) \leq (\Theta_s - \varepsilon)n$ for some $T = O(1)$ with probability $\Omega(1)$. Fact 3.20 then implies that after $O(\log n)$ additional steps the largest component in the configuration has size $O(\log n)$.

Finally, note that if $L_1(X_t) = O(\log n)$ then Facts 3.17(i) and 3.17(ii) imply that the number of active vertices is $\Omega(n)$ with probability $1 - O(n^{-1})$. Hence, $I(X_{t+1}) = \Omega(n)$ w.h.p. by Lemma 2.4. Since $L_1(X_{t+1}) = O(\log n)$ w.h.p. by Fact 3.17(iv), the result follows. □

### 3.3.3 Proofs of preliminary facts

Here we provide the proofs of the auxiliary facts used in the previous section.

*Proof of Fact 3.17.* Observe that $\mathrm{E}[A_t | X_t, \neg\mathcal{E}_t] = \frac{n - L_1(X_t)}{q} =: \mu$, and $\sum_{j \geq 2} L_j(X_t)^2 \leq rn$ since $L_2(X_t) \leq r$. Then, Hoeffding's inequality implies

$$\Pr\left[\, |A_t - \mu| > \xi(r) \mid X_t, \neg\mathcal{E}_t\right] \leq 2\exp\left(-\frac{2rn\log n}{rn}\right) \leq \frac{2}{n^2}.$$

Thus, $A_t \in \hat{J}_t$ with probability at least $1 - O(n^{-2})$. Also, $(\mu + \xi(r))\frac{\lambda}{n} < \frac{\lambda}{q} + o(1) < 1$ for sufficiently large $n$ since $\lambda < q$; hence, the random graph $G(A_t, p)$ is sub-critical and part (i) follows from Lemmas 2.6 and 2.10.

Parts (ii), (iii) and (iv) follow in similar fashion. If $\mathcal{L}(X_t)$ is active, then $A_t \in J_t$ with probability $1 - O(n^{-2})$ by Hoeffding's inequality. If $A_t \in J_t$ and $L_1(X_t) \geq (\Theta_s + \varepsilon)n$, then the random graph $G(A_t, p)$ is super-critical, since

$$\left(L_1(X_t) + \frac{n - L_1(X_t)}{q} - \xi(r)\right)\frac{\lambda}{n} \geq \left(\Theta_s + \varepsilon + \frac{1 - \Theta_s - \varepsilon}{q}\right)\lambda - o(1)$$
$$= 1 + \left(1 - \frac{1}{q}\right)\varepsilon\lambda - o(1) > 1,$$

where in last equality we used the fact that $\Theta_s = \frac{q-\lambda}{\lambda(q-1)}$. Part (ii) then follows from Lemma 2.7.

If $\mathcal{L}(X_t)$ is active, then $A_t \in J_t$ with probability $1 - O(n^{-2})$ by part (ii). Since also $L_1(X_t) \leq (\Theta_s - \varepsilon)n$, we have

$$\left(L_1(X_t) + \frac{n - L_1(X_t)}{q} + \xi(r)\right)\frac{\lambda}{n} \leq 1 - \left(1 - \frac{1}{q}\right)\varepsilon\lambda + o(1) < 1.$$

Hence, $G(A_t, p)$ is sub-critical and part (ii) follows from Lemma 2.6.

Finally, note that $\mathrm{E}[A_t|X_t] = n/q$ and since in this case $\mathcal{L}(X_t)$ is also small, Hoeffding's inequality implies that $A_t \in [n/q - \xi(r), n/q + \xi(r)]$ with probability $1 - O(n^{-2})$. If $A_t \in [n/q - \xi(r), n/q + \xi(r)]$ and $\lambda < q$, then the random graph $G(A_t, p)$ is sub-critical and part (iv) follows from Lemma 2.6. □

*Proof of Fact 3.18.* If $X_t$ has a unique large component and $\mathcal{L}(X_t)$ is activated, then Lemma 2.5 implies that $X_{t+1}$ has at most one large component with probability $1 - O(n^{-1/12})$. Otherwise, if $\mathcal{L}(X_t)$ is not activated, $X_{t+1}$ will have a unique large component with probability $1 - O(n^{-1})$ by Fact 3.17(i). The result then follows by a union bound over the $T = O(\log n)$ steps. □

*Proof of 3.19.* Let $N_t$ be the size of the largest new component created at time $t$. First observe that when $\mathcal{L}(X_t)$ is inactive, Fact 3.17(i) implies that $N_t = O(\log n)$ with probability $1 - O(n^{-1})$. Since by assumption $L_1(X_t) \geq (\Theta_s + \varepsilon)n$, $L_1(X_{t+1}) = L_1(X_t)$ with probability $1 - O(n^{-1})$ and thus

$$-O(1) \leq \mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t, \neg\mathcal{E}_t] \leq O(1). \tag{3.8}$$

To bound $\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t, \mathcal{E}_t]$, let $\mu_t = \theta_t n + \frac{(1-\theta_t)n}{q}$, $M_t = A_t - \mu_t$ and let $\ell_m(\theta_t)$ denote the size of the largest component of a $G(\mu_t + m, p)$ random graph. Note that if $M_t = m$, then $N_t$ and $\ell_m(\theta_t)$ have the same distribution. Also, if $A_t \in J_t$ (see (3.6)) then $M_t \in J_t' := [-\xi(r), \xi(r)]$. Hence, by Fact 3.17(ii)

$$\mathrm{E}[N_t \mid X_t, \mathcal{E}_t] \;\leq\; \sum_{m \in J_t'} \mathrm{E}[\ell_m(\theta_t)]\ \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1). \tag{3.9}$$

When $\theta_t \geq \Theta_s + \varepsilon$ and $m \in J_t'$, Fact 3.17(ii) implies that $G(\mu_t + m, p)$ is a super-critical random graph and by Lemma 2.8

$$\mathrm{E}[\ell_m(\theta_t)] \leq \phi(\theta_t)(n + m) + 2n^{1/4}. \tag{3.10}$$

(The function $\phi$ was defined in Section 3.1.2; see (3.3)). Hence,

$$\mathrm{E}[N_t \mid X_t, \mathcal{E}_t] \le \phi(\theta_t)n + \phi(\theta_t)\,\mathrm{E}[M_t \mid X_t, \mathcal{E}_t] + 2n^{1/4} + O(1) = \phi(\theta_t)n + 2n^{1/4} + O(1),$$

since $\mathrm{E}[M_t \mid X_t, \mathcal{E}_t] = 0$. Similarly, we obtain

$$\mathrm{E}[N_t \mid X_t, \mathcal{E}_t] \ge \phi(\theta_t)n - 2n^{1/4} - O(1).$$

Now, when $\mathcal{L}(X_t)$ is active, $L_1(X_{t+1}) \ne N_t$ only if the size of the largest inactive component is at least $N_t$. But since the second largest component is small (i.e., $L_2(X_t) < 2n^{11/12}$) and the percolation step is super-critical, Lemma 2.7 implies that $\Pr[L_1(X_{t+1}) \ne N_t \mid X_t, \mathcal{E}_t] = \exp(-\Omega(n))$. Consequently,

$$\phi(\theta_t)n - 2n^{1/4} - O(1) \le \mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t] \le \phi(\theta_t)n + 2n^{1/4} + O(1). \tag{3.11}$$

Putting together (3.8) and (3.11), we get

$$-\frac{f(\theta_t)n}{q} - \frac{2n^{1/4}}{q} - O(1) \le \mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \le -\frac{f(\theta_t)n}{q} + \frac{2n^{1/4}}{q} + O(1); \tag{3.12}$$

part (i) then follows from (3.11) and (3.12).

For part (ii), note that for any $\theta_t \in (\Theta_s - \varepsilon, \Theta_s + \varepsilon)$ and any $m \in J_t'$, we have $\mathrm{E}[\ell_m(\theta_t)] \le \mathrm{E}[\ell_m(\Theta_s + \varepsilon)]$ by monotonicity. Also, if $\hat{\mu}_t = (\Theta_s + \varepsilon)n + \frac{(1-\Theta_s-\varepsilon)n}{q}$, then by Fact 3.17(ii) the random graph $G(\hat{\mu}_t + m, p)$ is super-critical for any $m \in J_t'$. Hence,

$$\mathrm{E}[\ell_m(\theta_t)] \le \mathrm{E}[\ell_m(\Theta_s + \varepsilon)] \le \phi(\Theta_s + \varepsilon)(n + m) + O(n^{1/4}), \tag{3.13}$$

by Lemma 2.8. The inequality in (3.9) also holds for $\theta_t \in (\Theta_s - \varepsilon, \Theta_s + \varepsilon)$. Since again $L_1(X_{t+1}) = N_t$ with probability $1 - \exp(-\Omega(n))$, we get from (3.9) and (3.13) that

$$\mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t] \le \phi(\Theta_s + \varepsilon)n + O(n^{1/4}).$$

Together with (3.8), which also holds in this setting, this implies that

$$\begin{aligned}\mathrm{E}[L_1(X_{t+1}) \mid X_t] &\le \left(1 - \frac{1}{q}\right)\theta_t n + \frac{\phi(\Theta_s + \varepsilon)n}{q} + O(n^{1/4}) \\ &\le L_1(X_t) - \frac{f(\Theta_s + \varepsilon)n}{q} + \frac{2\varepsilon n}{q} + O(n^{1/4}),\end{aligned}$$

from which part (ii) follows. □

*Proof of Fact 3.20.* If $L_1(X_0) \le (\Theta_s - \varepsilon)n$ and $\mathcal{L}(X_0)$ is activated, then by Fact 3.17(iii) the largest new component has size $O(\log n)$ with probability $1 - O(n^{-1})$. Hence, $X_1$ has no large component with probability $\Omega(1)$. Now, Fact 3.17(iv) and a union bound imply that all the new components created during the $O(\log n)$ steps immediately after have size $O(\log n)$ w.h.p. Another union bound over components shows that during these $O(\log n)$ steps, every component in $X_1$ is activated w.h.p. Thus, after $O(\log n)$ steps the largest component in the configuration has size $O(\log n)$ with probability $\Omega(1)$, as desired. □

### 3.3.4 Coupling to the same component structure: Proof of Lemma 3.15

In this section we design a coupling of the CM steps which, starting from two configurations with certain properties (i.e., those established in Lemma 3.14), quickly converges to a pair of configurations with the same component structure. (We say that two random-cluster configurations $X$ and $Y$ have the same component structure if $L_j(X) = L_j(Y)$ for all $j \geq 1$.) In particular, we establish Lemma 3.15.

The following corollary of Lemma 3.14 will be used in the proof.

**Corollary 3.21.** *Let $q > 1$, $0 < \lambda < q$ and suppose $X_0$ is a random-cluster configuration such that $L_1(X_0) = O(\log n)$ and $I(X_0) = \Omega(n)$. Then, these two properties are maintained for $T$ steps of the CM dynamics w.h.p. provided $T = O(\log n)$.*

*Proof.* If $L_1(X_t) = O(\log n)$, then $L_1(X_{t+1}) = O(\log n)$ with probability $1 - O(n^{-1})$ by Fact 3.17(iv). Moreover, $A_t = \Omega(n)$ with probability $1-O(n^{-1})$ by Facts 3.17(i) and 3.17(ii). Thus, $I(X_{t+1}) = \Omega(n)$ with probability $1 - O(n^{-1})$ by Lemma 2.4. The result then follows by a union bound over the $T$ steps. □

*Proof of Lemma 3.15.* Since $L_1(X_0) = O(\log n)$ and $I(X_0) = \Omega(n)$, by Corollary 3.21 $L_1(X_t) = O(\log n)$ and $I(X_t) = \Omega(n)$ for all $t \in [0, T]$ w.h.p. provided $T = O(\log n)$; the same holds for $\{Y_t\}$. Hence, it is safe to assume that these properties are maintained throughout the $O(\log n)$ steps of the coupling.

Our coupling will be a composition of two couplings. Coupling I contracts a certain notion of distance between $\{X_t\}$ and $\{Y_t\}$. This contraction will boost the probability of success of Coupling II. Coupling II uses the binomial coupling from Lemma 2.18 to achieve two configurations with the same component structure with probability $\Omega(1)$.

**Coupling I:** Consider a maximal matching $W_t$ between the components of $X_t$ and $Y_t$ with the restriction that only components of equal size are matched to each other. Let $M(X_t)$ and $M(Y_t)$ be the components in the matching from $X_t$ and $Y_t$ respectively. Let $D(X_t)$ and $D(Y_t)$ be the complements of $M(X_t)$ and $M(Y_t)$ respectively, and let $d_t = |D(X_t)| + |D(Y_t)|$ where $|\cdot|$ denotes the total number of vertices in the respective components.

The activation of the components in $M(X_t)$ and $M(Y_t)$ is coupled using the matching $W_t$. That is, $c \in M(X_t)$ and $W_t(c) \in M(Y_t)$ are activated simultaneously with probability $1/q$. The components in $D(X_t)$ and $D(Y_t)$ are activated independently.

Let $A(X_t)$ and $A(Y_t)$ denote the set of active *vertices* in $X_t$ and $Y_t$ respectively, and w.l.o.g. assume $|A(X_t)| \geq |A(Y_t)|$. Let $R_t$ be an arbitrary subset of $A(X_t)$ such that $|R_t| = |A(Y_t)|$ and let $Q_t = A(X_t) \setminus R_t$. The percolation step is coupled by establishing an arbitrary vertex bijection $b_t : R_t \to A(Y_t)$ and coupling the re-sampling of each edge $(u, v) \in R_t \times R_t$ with $(b_t(u), b_t(v)) \in A(Y_t) \times A(Y_t)$. Edges within $Q_t$ and in the cut $C_t = R_t \times Q_t$ are re-sampled independently. The following claim establishes a contraction in $d_t$.

**Claim 3.22.** *Let $\omega(n) = \frac{n}{\log^4 n}$; after $T = O(\log \log n)$ steps, $d_T \leq \omega(n)$ w.h.p.*

*Proof.* Coupling I guarantees that $R_t$ and $A(Y_t)$ will have the same component structure internally. However, the vertices in $Q_t$ will contribute to $d_{t+1}$ and each edge added to $C_t$ in the re-sampling would increase $d_{t+1}$ by at most (twice) the size of a component of $R_t$, which is $O(\log n)$ by assumption. Thus,

$$\mathrm{E}[d_{t+1} \mid X_t, Y_t] \leq d_t - \frac{d_t}{q} + \mathrm{E}[|Q_t| \mid X_t, Y_t] + 2\,\mathrm{E}[K_t \mid X_t, Y_t] \times O(\log n), \tag{3.14}$$

where $K_t$ is the number of edges added to $C_t$ during the re-sampling. Now,

$$\mathrm{E}[K_t \mid X_t, Y_t] \leq \sum_{m=0}^{n} \mathrm{E}[K_t \mid X_t, Y_t, |Q_t| = m] \leq \lambda\,\mathrm{E}[|Q_t| \mid X_t, Y_t]. \tag{3.15}$$

Moreover, since $|D(X_t)| = |D(Y_t)|$, the expected number of active vertices from $D(X_t)$ and $D(Y_t)$ is the same, and Hoeffding's inequality implies that $|Q_t| = O(\sqrt{n}\log n)$ with probability $1 - O(n^{-1})$. Hence, $\mathrm{E}[|Q_t| \mid X_t, Y_t] = O(\sqrt{n}\log n)$ and from (3.14) and (3.15) we get

$$\begin{aligned}\mathrm{E}[d_{t+1} \mid X_t, Y_t] &\leq d_t - \frac{d_t}{q} + O\left(\sqrt{n}\log^2 n\right) \\ &\leq \left(1 - \frac{1}{2q}\right) d_t\end{aligned} \tag{3.16}$$

provided $d_t > \omega(n)$. Thus, Markov's inequality implies $d_T \leq \omega(n)$ for some $T = O(\log\log n)$ w.h.p. (Note that for larger values of $T$, this argument immediately provides stronger bounds for $d_T$, but neither our analysis nor the order of the coupling time benefits from this.) □

**Coupling II:** Assume now that $d_0 \leq \omega(n)$ and let $I_{\mathrm{M}}(X_t)$ and $I_{\mathrm{M}}(Y_t)$ denote the isolated vertices in $M(X_t)$ and $M(Y_t)$ respectively. The activation in $X_t \setminus I_{\mathrm{M}}(X_t)$ and $Y_t \setminus I_{\mathrm{M}}(Y_t)$ is coupled as in Coupling I. This first part of the activation could activate a different number of vertices from each copy of the chain; let $\rho_t$ be this difference.

Since $d_t \leq \omega(n) = n/\log^4 n$ and the expected number of active vertices from $D(X_t)$ and $D(Y_t)$ is the same, Hoeffding's inequality implies $\rho_t = O\left(\sqrt{n}\log^{-1} n\right)$ w.h.p. We show next how to couple the activation in $I_{\mathrm{M}}(X_t)$ and $I_{\mathrm{M}}(Y_t)$ in a way that $|A(X_t)| = |A(Y_t)|$ w.h.p.

The number of active isolated vertices from $I_{\mathrm{M}}(X_t)$ is binomially distributed with parameters $|I_{\mathrm{M}}(X_t)|$ and $1/q$, and similarly for $I_{\mathrm{M}}(Y_t)$. Hence, the activation of the isolated vertices may be coupled using the binomial coupling from Lemma 2.18. Let $\mathcal{H}_t$ be the event that this coupling of the isolated vertices succeeds in correcting the error $\rho_t$. Since $|I_{\mathrm{M}}(X_t)| = |I_{\mathrm{M}}(Y_t)| = \Omega(n)$ and $\rho_t = O\left(\sqrt{n}\log^{-1} n\right)$, Lemma 2.18 implies that $\mathcal{H}_t$ occurs with probability $1 - O(\log^{-1} n)$. If this is the case, and we couple the edge re-sampling bijectively as in Coupling I, the updated part of both configurations will have the same component structure and also $d_{t+1} \leq d_t$. Hence, if $\mathcal{H}_t$ occurs for all $0 \leq t \leq T$, then $X_T$ and $Y_T$ fail to have the same component structure only if at least one of the initial components was never activated. For $T = O(\log n)$ this occurs with at most constant probability. Since $\mathcal{H}_t$ occurs for all $0 \leq t \leq T = O(\log n)$ with at least constant probability, then $X_T$ and $Y_T$ have the same component structure with probability $\Omega(1)$.

Couplings I and II succeed each with at least constant probability. Thus, the overall coupling succeeds with probability $\Omega(1)$, as desired. □

### 3.3.5 Coupling to the same configuration: Proof of Lemma 3.16

To conclude Section 3.3 we provide the proof of Lemma 3.16. In this lemma we construct a coupling of the CM steps that starts from two configurations with the same component structure and converges in $O(\log n)$ steps to two identical configurations.

*Proof of Lemma 3.16.* Let $B_t$ a bijection between the vertices of $X_t$ and $Y_t$. We first describe how to construct $B_0$. Consider a maximal matching between the components of $X_0$ and $Y_0$ with the restriction that only components of equal size are matched to each other. Since the two configurations have the same component structure all components are matched. Using this matching, vertices between matched components are mapped arbitrarily to obtain $B_0$.

Vertices mapped to themselves we call "fixed". At time $t$, the component activation is coupled according to $B_t$. That is, if $B_t(u) = v$ for $u \in X_t$ and $v \in Y_t$, then the components containing $u$ and $v$ are simultaneously activated with probability $1/q$. $B_{t+1}$ is adjusted such that if a vertex $w$ becomes active in both configurations then $B_{t+1}(w) = w$; the rest of the activated vertices are mapped arbitrarily in $B_{t+1}$ and the inactive vertices are mapped like in $B_t$. The percolation step at time $t$ is then coupled using $B_{t+1}$. That is, the re-sampling of the active edge $(u, v) \in X_t$ is coupled with the re-sampling of the active edge $(B_{t+1}(u), B_{t+1}(v)) \in Y_t$.

This coupling ensures that the component structures of $X_t$ and $Y_t$ remain the same for all $t \geq 0$. Moreover, once a vertex is fixed it remains fixed forever. The probability that a vertex is fixed in one step is $1/q^2$. Therefore, after $O(\log n)$ steps the probability that a vertex is not fixed is at most $1/n^2$. A union bound over all vertices implies that $X_T = Y_T$ w.h.p. after $T = O(\log n)$ steps. □

## 3.4 Mixing time upper bounds: the super-critical case

In this section we establish our mixing time upper bound for the CM dynamics when $\lambda > \lambda_S = q$ and $q > 1$. From Lemma 3.3 we know that $\lambda_c = \lambda_S = q$ for $q \leq 2$ and $\lambda_c < \lambda_S$ for $q > 2$, so here we analyze the mixing time in most of the super-critical regime. The mixing time for $q > 2$ and $\lambda \in (\lambda_c, \lambda_S]$ is analyzed in Sections 3.6 and 3.7.

We establish the following theorem.

**Theorem 3.23.** *Let $q > 1$ and $\lambda > \lambda_S = q$. Then, the mixing time of the CM dynamics is $O(\log n)$.*

The structure of the proof of this theorem does not differ significantly from that of Theorem 3.12 for the sub-critical regime. Indeed, we again construct a coupling of the CM dynamics with four phases: the first two correspond to a burn-in period; the third phase is a coupling that reaches two configurations with the same component structure; and the last phase couples two configurations that maintain the same component structure. For the first and last phases we are able to reuse

Lemmas 3.13 and 3.16 from Section 3.3. However, super-critical random-cluster configurations are likely to have a giant component of size roughly $\theta_r n$, where $\theta_r$ is the largest positive solution of equation (3.1). Thus, the goal of the second burn-in phase in the super-critical setting is to reach two configurations with their respective unique large component of size $\sim\theta_r n$. To do this, we again use the drift function $f$ (see (3.4)). For $\lambda > \lambda_S$, $\theta_r$ is the unique positive zero of this function. Moreover, $f$ is negative for $\theta < \theta_r$ and positive for $\theta > \theta_r$ (see Figure 3.1). Consequently, we can show that the drift is always "towards" $\theta_r n$. However, in contrast to the sub-critical case, we now have to deal the fact $f(\theta_r) = 0$, so the drift can be arbitrarily small near $\theta_r$; this requires a more subtle argument.

The coupling to the same component structure in the third phase also requires some new insights to deal with the two components of linear size, but it may be seen as an extension of the coupling in Lemma 3.15.

The following lemmas are used in the proof of Theorem 3.23. Let $\Theta_S := 1 - q/\lambda$ and let $\varepsilon > 0$ be a small constant independent of $n$ we choose later.

**Lemma 3.24.** *Let $\lambda > q$. If $X_0$ has at most one large component, then $L_1(X_1) = \Omega(n)$ and $L_2(X_1) < 2n^{11/12}$ with probability $\Omega(1)$.*

**Lemma 3.25.** *Let $\lambda \geq q$. If $L_1(X_0) = \Omega(n)$ and $L_2(X_0) < 2n^{11/12}$, then there exists $T = O(\log n)$ such that: $L_1(X_T) > (\Theta_S + \varepsilon)n$, $L_2(X_T) = O(\log n)$, and $\sum_{j\geq 2} L_j(X_T)^2 = O(n)$ with probability $\Omega(1)$. Moreover, once these properties are obtained they are preserved for a further $T' = O(\log n)$ CM steps w.h.p.*

**Lemma 3.26.** *Let $\lambda \geq q$. Suppose that $L_1(X_0) > (\Theta_S + \varepsilon)n$, $L_2(X_0) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_0)^2 = O(n)$. Then, there exists $T = O(\log n)$ such that with probability $\Omega(1)$: $|L_1(X_T) - \theta_r n| = O(\sqrt{n})$, $L_2(X_T) = O(\log n)$, $\sum_{j\geq 2} L_j(X_T)^2 = O(n)$ and $I(X_T) = \Omega(n)$.*

**Lemma 3.27.** *Let $\lambda \geq q$ and let $X_0$ and $Y_0$ be random-cluster configurations such that:*

*(i)* $|L_1(X_0) - \theta_r n|, |L_1(Y_0) - \theta_r n| = O(\sqrt{n})$;

*(ii)* $L_2(X_0), L_2(Y_0) = O(\log n)$;

*(iii)* $I(X_0), I(Y_0) = \Omega(n)$; *and*

*(iv)* $\sum_{j\geq 2} L_j(X_0)^2, \sum_{j\geq 2} L_j(Y_0)^2 = O(n)$.

*Then, there exists a coupling of the CM steps such that $X_T$ and $Y_T$ have the same component structure after $T = O(\log n)$ steps with probability $\Omega(1)$.*

*Proof of Theorem 3.23.* Let $X_0$ and $Y_0$ be two arbitrary random-cluster configurations. We construct a coupling of the CM dynamics and show that $\Pr[X_T = Y_T] = \Omega(1)$ for some $T = O(\log n)$; the result then follows from (2.1). By Lemmas 3.13, 3.24, 3.25 and 3.26 after $T = O(\log n)$ steps: $|L_1(X_T) - \theta_r n| = O(\sqrt{n})$, $L_2(X_T) = O(\log n)$, $I(X_T) = \Omega(n)$ and $\sum_{j\geq 2} L_j(X_T)^2 = O(n)$ with probability $\Omega(1)$, and the same holds for $Y_T$. The result then follows from Lemmas 3.27 and 3.16. ◻

We turn next to the proofs of Lemmas 3.24, 3.25, 3.26 and 3.27. We point out that Lemma 3.24 holds only for $\lambda > q$, while all other lemmas used in the proof of Theorem 3.23 hold for $\lambda \geq q$. This will be useful later in Section 3.6 where we analyze the $\lambda = q$ regime.

Lemmas 3.24 and 3.25 lay the groundwork for the proof of Lemma 3.26, which is proved first in Section 3.4.1. The proofs of Lemmas 3.24 and 3.25 can be found in Section 3.4.2. Finally, Lemma 3.27 is proved in Section 3.4.3.

### 3.4.1 Proof of Lemma 3.26

Let $\xi(r) = \sqrt{nr \log n}$, $\mu_t = L_1(X_t) + \frac{n-L_1(X_t)}{q}$ and $\hat{\mu}_t = \frac{n-L_1(X_t)}{q}$. The following fact will be useful.

**Fact 3.28.** *Let $\lambda \geq q$. Assume $X_t$ has at most one large component and that $L_2(X_t) \leq r < 2n^{11/12}$. Then for sufficiently large $n$, each of the following holds:*

*(i) If $\mathcal{L}(X_t)$ is inactive and $L_1(X_t) > (\Theta_S + \varepsilon)n$, then $A_t \in [\hat{\mu}_t - \xi(r), \hat{\mu}_t + \xi(r)]$ and the largest new component has size $O(\log n)$ with probability $1 - O\left(n^{-1}\right)$.*

*(ii) If $\mathcal{L}(X_t)$ is active, then $A_t \in J_{t,r} := [\mu_t - \xi(r), \mu_t + \xi(r)]$ with probability $1 - O\left(n^{-1}\right)$. Moreover, if $A_t \in J_{t,r}$ and $L_1(X_t) = \Omega(n)$, then $G(A_t, p)$ is a super-critical random graph.*

*Proof of Lemma 3.26.* By assumption we have that $L_1(X_0) > (\Theta_S + \varepsilon)n$, $L_2(X_0) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_0)^2 = O(n)$. Hence, Lemma 3.25 implies that $\{X_t\}$ retains these properties for $T$ steps w.h.p., provided $T = O(\log n)$. Thus, we may assume throughout the proof that these properties are maintained.

Let $\Delta_t := |L_1(X_t) - \theta_r n|$; we show that one step of the CM dynamics contracts $\Delta_t$ in expectation. First, if $L_1(X_t) > (\Theta_S + \varepsilon)n$ and $\mathcal{L}(X_t)$ is inactive, then $L_1(X_{t+1}) = L_1(X_t)$ with probability $1 - O(n^{-1})$ by Fact 3.28(i). Therefore,

$$\mathrm{E}[\Delta_{t+1} \mid X_t, \neg\mathcal{E}_t] \leq \mathrm{E}\left[\, |L_1(X_{t+1}) - L_1(X_t)| \mid X_t, \neg\mathcal{E}_t\right] + |L_1(X_t) - \theta_r n| \leq \Delta_t + O(1). \tag{3.17}$$

To bound $\mathrm{E}[\Delta_{t+1} \mid X_t, \mathcal{E}_t]$, let $M_t := A_t - \mu_t$ and $\Delta'_{t+1} := |L_1(X_{t+1}) - \phi(\theta_t)n|$. Note that if $A_t \in J_{t,r}$, then $M_t \in J'_{t,r} := [-\xi(r), \xi(r)]$. Hence, Fact 3.28(ii) implies

$$\mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t] \leq \sum_{m \in J'_{t,r}} \mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t, M_t = m]\ \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1) \tag{3.18}$$

Let $\ell_t(m)$ be the size of the largest component of a $G(\mu_t + m, p)$ random graph. Conditioned on $M_t = m$, the largest *new* component has size $\ell_t(m)$. The random graph $G(\mu_t + m, p)$ is a super-critical for $m \in J'_{t,r}$ by Fact 3.28(ii). Thus, $\ell_t(m) = \Omega(n)$ with probability $1 - O(n^{-1})$ by Lemma 2.7. Since the size of the largest inactive component is $O(\log n)$, we have that $L_1(X_{t+1}) = \ell_t(m)$ with probability $1 - O(n^{-1})$. From (3.18) we then get

$$\mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t] \quad \leq \quad \sum_{m \in J'_{t,r}} \mathrm{E}[|\ell_t(m) - \phi(\theta_t)n|]\ \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1). \tag{3.19}$$

Now,

$$\begin{aligned}\mathrm{E}[|\ell_t(m) - \phi(\theta_t)n|] &\le \mathrm{E}[|\ell_t(m) - \mathrm{E}[\ell_t(m)]|] + |\,\mathrm{E}[\ell_t(m)] - \phi(\theta_t)n| \\ &\le \sqrt{\mathrm{Var}(\ell_t(m))} + |\,\mathrm{E}[\ell_t(m)] - \phi(\theta_t)n|.\end{aligned}$$

Since $G(\mu_t + m, p)$ is super-critical, it follows from Lemma 2.8 that $\mathrm{Var}(\ell_t(m)) = O(n)$ and $|\,\mathrm{E}[\ell_t(m)] - \phi(\theta_t)n| \le \phi(\theta_t)|m| + O(\sqrt{n})$. Hence,

$$\mathrm{E}[|\ell_t(m) - \phi(\theta_t)n|] \le |m| + O(\sqrt{n}).$$

Plugging this inequality into (3.19), we get

$$\mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t] \le \mathrm{E}[|M_t| \mid X_t, \mathcal{E}_t] + O(\sqrt{n}). \tag{3.20}$$

The following fact, which we prove later, follows straightforwardly from Hoeffding's inequality since by assumption $\sum_{j\ge2} L_j(X_t)^2 = O(n)$.

**Fact 3.29.** $\mathrm{E}[\,|M_t| \mid X_t, \mathcal{E}_t] = O(\sqrt{n})$.

Fact 3.29 and (3.20) imply

$$\mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t] = O(\sqrt{n}),$$

and by the triangle inequality

$$\mathrm{E}[\Delta_{t+1} \mid X_t, \mathcal{E}_t] \;\le\; \mathrm{E}[\Delta'_{t+1} \mid X_t, \mathcal{E}_t] + |\theta_r - \phi(\theta_t)|n \;\le\; |\theta_r - \phi(\theta_t)|n + O(\sqrt{n}). \tag{3.21}$$

Putting (3.17) and (3.21) together, we have

$$\mathrm{E}[\Delta_{t+1} \mid X_t] \le \left(1 - \frac{1}{q}\right)\Delta_t + \frac{|\theta_r - \phi(\theta_t)|n}{q} + O(\sqrt{n}). \tag{3.22}$$

Since by assumption $\theta_t > \Theta_{\mathrm{S}} + \varepsilon$, Lemma 3.9(ii) implies that there exists a constant $\delta \in (0,1)$ such that $\delta|\theta_t - \theta_r| \le |\theta_t - \phi(\theta_t)|$. Together with Lemma 3.9(i), this implies $|\theta_r - \phi(\theta_t)| \le (1-\delta)|\theta_t - \theta_r|$. Plugging this into (3.22), we obtain

$$\mathrm{E}[\Delta_{t+1} \mid X_t] \le (1 - \delta/q)\Delta_t + O(\sqrt{n}),$$

and inducting

$$\mathrm{E}[\Delta_t] \le (1 - \delta/q)^t \Delta_0 + O(\sqrt{n}). \tag{3.23}$$

Therefore, for a sufficiently large $T = O(\log n)$ we have $\mathrm{E}[\Delta_T] = O(\sqrt{n})$, and by Markov's inequality $\Delta_T = O(\sqrt{n})$ with probability $\Omega(1)$.

Finally, by Facts 3.28(i) and 3.28(ii) the number of active vertices at time $T-1$ is $\Omega(n)$ w.h.p., and thus, by Lemma 2.4, $I(X_T) = \Omega(n)$ w.h.p. Since also $L_2(X_T) = O(\log n)$ and $\sum_{j\ge2} L_j(X_T)^2$, the result follows. ◻

We now provide the missing proof of Facts 3.29 and 3.28.

*Proof of Fact 3.29.* Let $W_t$ be a random variable distributed according to the conditional distribution of $|M_t|$ given $X_t$ and $\mathcal{E}_t$. Since $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$, Hoeffding's inequality implies that there exists a constant $c$ such that $\Pr[W_t > a\sqrt{n}] \leq 2\exp(-ca^2)$ for every $a > 0$. Observe also that

$$W_t = \sum_{k=0}^{n} \mathbb{1}(W_t \geq k+1) + \mathbb{1}(k+1 > W_t > k)(W_t - k).$$

Therefore,

$$\mathrm{E}[W_t] \leq \sum_{k=0}^{n} \Pr[W_t > k] \leq 1 + 2\sum_{k=1}^{n} e^{-\frac{ck^2}{n}} \leq 1 + 2\int_0^\infty e^{-\frac{cx^2}{n}} dx = O(\sqrt{n}),$$

as desired. □

*Proof of Fact 3.28.* This proof is similar to that of Fact 3.17. If $\mathcal{L}(X_t)$ is inactive, then Hoeffding's inequality implies that $A_t \in [\hat{\mu}_t - \xi(r), \hat{\mu}_t + \xi(r)]$ with probability $1 - O(n^{-2})$. It is straightforward to check that if $A_t \in [\hat{\mu}_t - \xi(r), \hat{\mu}_t + \xi(r)]$, then $G(A_t, p)$ is a sub-critical random graph, provided $L_1(X_t) > (\Theta_S + \varepsilon)n$. Part (i) then follows from Lemma 2.6.

Part (ii) follows in similar fashion. If $\mathcal{L}(X_t)$ is active, then $A_t \in J_{t,r}$ with probability $1 - O(n^{-2})$ by Hoeffding's inequality. Moreover,

$$\frac{\lambda}{n}(\mu_t - \xi(r)) > \frac{\lambda}{q} + \lambda\left(1 - \frac{1}{q}\right)\frac{L_1(X_t)}{n} - o(1) > 1$$

provided that $\lambda \geq q$, $L_1(X_t) = \Omega(n)$ and that $n$ is large enough. This implies that $G(A_t, p)$ is super-critical when $A_t \in J_{t,r}$. □

### 3.4.2 Proof of preliminary facts

In this section we give the proofs of Lemmas 3.24 and 3.25.

*Proof of Lemma 3.24.* First note that if $\mathcal{L}(X_0)$ is activated, then $A_0 \in J_{0,r}$ w.h.p. by Fact 3.28(ii). The fact that $\lambda > q$ implies that $G(A_0, p)$ is a super-critical random graph for $A_0 \in J_{0,r}$. (This is not necessarily the case when $\lambda = q$ and this why the lemma does not hold for $\lambda = q$.) Hence, $L_1(X_1) = \Omega(n)$ with probability $\Omega(1)$ by Lemma 2.7 and $L_2(X_1) < 2n^{11/12}$ w.h.p. by Lemma 2.5. □

*Proof of Lemma 3.25.* It will be convenient to split the proof in three parts, each corresponding to one of the following claims.

**Claim 3.30.** *If $L_1(X_0) = \Omega(n)$ and $L_2(X_0) < 2n^{11/12}$, then there exists $T = O(\log n)$ such that $L_1(X_T) > (\Theta_S + \varepsilon)n$ and $L_2(X_T) < 2n^{11/12}$ with probability $\Omega(1)$.*

**Claim 3.31.** *If $L_1(X_0) > (\Theta_S + \varepsilon)n$ and $L_2(X_0) < 2n^{11/12}$, then these properties are preserved for a further $T = O(\log n)$ CM steps w.h.p.*

**Claim 3.32.** *If $L_1(X_0) > (\Theta_S + \varepsilon)n$ and $L_2(X_0) < 2n^{11/12}$, then there exists $T = O(\log n)$ such that $L_2(X_T) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_T)^2 = O(n)$ with probability $\Omega(1)$. Moreover, once these properties are obtained they are preserved for a further $T' = O(\log n)$ CM steps w.h.p.*

Lemma 3.25 follows directly from these three claims. □

*Proof of Claim 3.30.* By Lemma 2.5 and a union bound, $L_2(X_t) < 2n^{11/12}$ for all $t \in [0, T]$ w.h.p., provided $T = O(\log n)$. Also, since $L_1(X_0) = \Omega(n)$ by assumption, if $L_1(X_t) = \Omega(n)$, then when $\mathcal{L}(X_t)$ is active $L_1(X_{t+1}) = \Omega(n)$ with probability $1 - O(n^{-1})$ by Fact 3.28(ii) and Lemma 2.7; when $\mathcal{L}(X_t)$ is inactive $L_1(X_{t+1}) \geq L_1(X_t)$. A union bound over the steps implies that $L_1(X_t) = \Omega(n)$ for all $t \in [0, T]$ w.h.p. Hence, it is safe to assume that both of these properties are preserved throughout the proof.

Let $d_t := (\Theta_S + \varepsilon)n - L_1(X_t)$. Then,

$$\mathrm{E}[d_{t+1} \mid X_t, \mathcal{E}_t] = d_t + \theta_t n - \mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t]. \tag{3.24}$$

Let $h^-(\theta_t) := \theta_t n + \frac{(1-\theta_t)n}{q} - \xi(r)$ and let $\ell^-(\theta_t)$ be the size of the largest component of a $G(h^-(\theta_t), p)$ random graph. If $\mathcal{L}(X_t)$ is activated, Fact 3.28(ii) implies that $A_t \in J_{t,r}$ with probability $1 - O\left(n^{-1}\right)$, where we take $r = 2n^{11/12}$. Therefore,

$$\begin{aligned} \mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t] &\geq \sum_{a \in J_{t,r}} \mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t, A_t = a] \Pr[A_t = a \mid X_t, \mathcal{E}_t] \\ &\geq \mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t, A_t = h^-(\theta_t)] - \Omega(1), \end{aligned} \tag{3.25}$$

where the second inequality follows by monotonicity. Since $L_1(X_t) = \Omega(n)$, Fact 3.28(ii) implies that $G(h^-(\theta_t), p)$ is a super-critical random graph. Hence, $\ell^-(\theta_t) = \Omega(n)$ with probability $1 - O(n^{-1})$ by Lemma 2.7. Now, $L_2(X_t) < 2n^{11/12}$, so if $A_t = h^-(\theta_t)$, then $L_1(X_{t+1}) = \ell^-(\theta_t)$ with probability $1 - O(n^{-1})$. Thus, from (3.25), we get

$$\mathrm{E}[L_1(X_{t+1}) \mid X_t, \mathcal{E}_t] \geq \mathrm{E}[\ell^-(\theta_t)] - \Omega(1) \geq \phi(\theta_t)n - \Omega(\xi(r)),$$

where the last inequality follows from Lemma 2.8. Plugging this bound into (3.24), we have

$$\mathrm{E}[d_{t+1} \mid X_t, \mathcal{E}_t] \leq d_t + (\theta_t - \phi(\theta_t))n + O(\xi(r)).$$

Now, $\theta_r > \Theta_S$ by Fact 3.10 and thus $\theta_r > \Theta_S + \varepsilon$ for small enough $\varepsilon$. Also, $L_1(X_t) = \Omega(n)$ by assumption. Therefore, if $d_t > 0$ (i.e., $\Omega(n) = L_1(X_t) < (\Theta_S + \varepsilon)n$), then Lemma 3.9(ii) implies that there exists a constant $\delta \in (0, 1)$ such that $\phi(\theta_t) - \theta_t > \delta(\theta_r - \theta_t) > \delta(\theta_r - \Theta_S - \varepsilon) =: \delta'$. Note that $\delta'$ is a constant in $(0, 1)$ for sufficiently small $\varepsilon$. Moreover, $\xi(r) = o(n)$ and thus there is a constant $\gamma > 0$ such that

$$\mathrm{E}[d_{t+1} - d_t \mid X_t, \mathcal{E}_t] \leq -\delta' n + O(\xi(r)) \leq -\gamma n, \tag{3.26}$$

provided $d_t > 0$.

Assuming $d_0 > 0$ (there is nothing to prove otherwise), let $\tau = \min\{t > 0 : d_t \leq 0\}$ and let $\mathcal{H}_K$ be the event that $\mathcal{L}(X_t)$ is activated for all $t \in [0, K]$, where $K$ is a fixed constant independent of $n$ we choose later. Let $\hat{T} := \min\{\tau, K\}$ and observe that conditioned on $\mathcal{H}_K$, (3.26) holds for all $t < \hat{T}$. Hence, Lemma 2.20 implies $\mathrm{E}[\hat{T}|\mathcal{H}_K] \leq 4/\gamma$, and by Markov's inequality we have that for sufficiently large $K$

$$\Pr[\tau < K/2 \mid \mathcal{H}_K] \geq \Pr[\hat{T} < K/2 \mid \mathcal{H}_K] \geq 1 - \frac{8}{\gamma K} > 0.$$

Since the event $\mathcal{H}_K$ occurs with constant probability $q^{-K}$, we have $L_1(X_T) \geq (\Theta_{\mathrm{S}} + 2\varepsilon)n$ with probability $\Omega(1)$ for some $T = O(1)$. □

*Proof of Claim 3.31.* We show that if $L_1(X_0) > (\Theta_{\mathrm{S}} + \varepsilon)n$ and $L_2(X_0) < 2n^{11/12}$, then $L_1(X_1) > (\Theta_{\mathrm{S}} + \varepsilon)n$ and $L_2(X_1) < 2n^{11/12}$ with probability $1 - O(n^{-1/12})$. A union bound over the steps then implies $L_1(X_t) > (\Theta_{\mathrm{S}} + \varepsilon)n$ and $L_2(X_t) < 2n^{11/12}$ for all $t \in [0, T]$ with probability $1 - O\left(T/n^{1/12}\right)$.

If $\mathcal{L}(X_0)$ is not activated, by Fact 3.28(i), $L_1(X_1) = L_1(X_0) > (\Theta_{\mathrm{S}} + \varepsilon)n$ with probability $1 - O\left(n^{-1}\right)$. Otherwise, let $h^-(\theta_0) := \theta_0 n + \frac{(1-\theta_0)n}{q} - \xi(r)$ and let $\ell^-(\theta_0)$ be the size of the largest component of a $G(h^-(\theta_0), p)$ random graph. By monotonicity

$$\Pr[L_1(X_1) < \phi(\theta_0)n - 2\xi(r) \mid A_0 \in J_{0,r}] \leq \Pr[L_1(X_1) < \phi(\theta_0)n - 2\xi(r) \mid A_0 = h^-(\theta_0)].$$

By Fact 3.28(ii) with $r = 2n^{11/12}$, $G(h^-(\theta_0), p)$ is super-critical with probability $1 - O(n^{-1})$. Hence, if $\mathcal{L}(X_0)$ is activated and $A_0 = h^-(\theta_0)$, $L_1(X_1) = \ell^-(\theta_0)$ with probability $1 - O(n^{-1})$ since $L_2(X_0) < 2n^{11/12}$. Therefore,

$$\Pr[L_1(X_1) < \phi(\theta_0)n - 2\xi(r) \mid \mathcal{E}_0, A_0 \in J_{0,r}] \leq \Pr[\ell^-(\theta_0) < \phi(\theta_0)n - 2\xi(r)],$$

and by Lemma 2.7

$$\Pr[L_1(X_1) < \phi(\theta_0)n - 2\xi(r) \mid \mathcal{E}_0, A_0 \in J_{0,r}] = O(n^{-1}).$$

By Lemma 3.9(i), either $\phi(\theta_0)n - 2\xi(r) > \theta_0 n$ or $\phi(\theta_0)n - 2\xi(r) > \theta_r n$ for sufficiently large $n$ since $\xi(r) = o(n)$. In either case, $\phi(\theta_0)n - 2\xi(r) > (\Theta_{\mathrm{S}} + \varepsilon)n$, since $\theta_0 > \Theta_{\mathrm{S}} + \varepsilon$ by assumption and $\theta_r > \Theta_{\mathrm{S}}$ by Fact 3.10. Thus,

$$\Pr[L_1(X_1) \leq (\Theta_{\mathrm{S}} - \varepsilon)n \mid \mathcal{E}_0, A_0 \in J_{0,r}] = O(n^{-1}).$$

When $\mathcal{L}(X_0)$ is active, $A_0 \in J_{0,r}$ with probability $1 - O(n^{-1})$ by Fact 3.28(i). Hence, $L_1(X_1) > (\Theta_{\mathrm{S}} + \varepsilon)n$ with probability $1 - O\left(n^{-1}\right)$. Finally, note that $L_2(X_1) < 2n^{11/12}$ with probability $1 - O(n^{-1/12})$ and thus the result follows by a union bound. □

*Proof of Claim 3.32.* By Claim 3.31 we may condition on the event $L_1(X_t) > (\Theta_{\mathrm{S}}+\varepsilon)n$ and $L_2(X_t) < 2n^{11/12}$ for all $t \in [0, T]$ with $T = O(\log n)$. By fact 3.28(i) if $\mathcal{L}(X_t)$ is inactive or by Fact 3.28(ii) and Lemma 2.7 if $\mathcal{L}(X_t)$ is active, every new *small* component has size $O(\log n)$ with probability

$1 - O(n^-1)$. A union bound over the steps then implies that every new small component up to time $T$ has size $O(\log n)$ with probability $1 - O(T/n)$. Moreover, the probability that any initial component remains after $T = B \log n$ steps is $O(n^{-1})$ for a sufficiently large constant $B > 0$; therefore, $L_2(X_T) = O(\log n)$ with probability $1 - O(\log n/n)$. Facts 3.28(i), 3.28(ii), Lemma 2.7 and another union bound implies that this property is maintained for an additional $O(\log n)$ steps w.h.p.

To establish that $\sum_{j\geq 2} L_j(X_T)^2 = O(n)$, we consider the one-dimensional random process $\{Z_t\}$ where $Z_t = \sum_{j\geq 2} L_j(X_t)^2$. At any time $t$, the decrease in $Z_t$ as a result of the dissolution of active components is $Z_t/q$ in expectation, and is at least $Z_t/q - o(n)$ with probability $1 - O(n^{-1})$ by Hoeffding's inequality since $L_2(X_t) < 2n^{11/12}$. If $L_1(X_t) > (\Theta_s + \varepsilon)n$, then Facts 3.28(i) and 3.28(ii) imply that when $\mathcal{L}(X_t)$ is active (resp., inactive), then the percolation step is super-critical (resp., sub-critical) with probability $1 - O(n^{-1})$. Hence, Lemma 2.11 implies that the increase in $Z_t$ as a result of the creation of new components in the percolation step is at most $Cn$ with probability $1 - O(n^{-1})$, for some constant $C > 0$. Therefore,

$$\mathrm{E}[Z_{t+1} \mid X_t] \leq Z_t - \frac{Z_t}{q} + Cn + o(n)$$

and thus $Z_T < 8Cqn$ with probability $\Omega(1)$ for some $T = O(\log n)$ by Lemma 2.24. Finally, when $Z_t > 8Cqn$, $Z_t$ decreases by at least $8Cn - o(n)$ and increases by at most $Cn$ with probability $1 - O(n^{-1})$; therefore, $Z_{t+1} \leq Z_t$ with probability $1 - O(n^{-1})$. Moreover, if $Z_t \leq 8Cqn$, then $Z_{t+1} \leq (8q + 1)Cn$ with probability $1 - O(n^{-1})$. Hence, if $Z_0 \leq 8Cqn$, by a union bound $Z_t = O(n)$ for all $t \in [0, T]$ w.h.p., provided $T = O(\log n)$. □

### 3.4.3 Coupling to the same component structure: Proof of Lemma 3.27

In this section we design a coupling of the CM steps which converges quickly to a pair of configurations with the same component structure, assuming the starting configurations have the super-critical properties guaranteed by Lemma 3.26. In particular, we prove Lemma 3.27, which is an extension of Lemma 3.15. For convenience, we reuse the notation introduced in the proof of that lemma.

The following corollary of Lemma 3.26 will be used in the proof.

**Corollary 3.33.** *Let $\lambda \geq q$ and $T = O(\log n)$. Suppose $X_0$ is a random-cluster configuration with a component structure that satisfies properties (i)-(iv) from Lemma 3.27. Then, $|L_1(X_t) - \theta_r n| = O(\sqrt{n}\log^2 n)$, $L_2(X_t) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$ for all $t \in [0, T]$ w.h.p. Moreover, for any $t \in [0, T]$, $|L_1(X_t) - \theta_r n| = O(\sqrt{n})$ with probability $\Omega(1)$.*

*Proof.* By Claim 3.32, $L_2(X_t) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$ for all $t \in [0, T]$ w.h.p. Moreover, if a configuration has these two properties, then the number of active vertices is $\Omega(n)$ with probability $1 - O(n^{-1})$ by Hoeffding's inequality. Lemma 2.4 and a union bound then imply that $I(X_t) = \Omega(n)$ for all $t \in [0, T]$ w.h.p.

Finally note that by (3.23), $\mathrm{E}[|L_1(X_t) - \theta_r n|] = O(\sqrt{n})$ for any $t \in [0, T]$. Hence, by Markov's inequality there exists a constant $c > 0$ such that $\Pr[|L_1(X_t) - \theta_r n| \geq A\sqrt{cn}] \leq 1/A$ for any

$t \geq 0$ and $A > 0$; thus, $|L_1(X_t) - \theta_r n| = O(\sqrt{n})$ with probability $\Omega(1)$. Taking $A = O(\log^2 n)$, a union bound implies that $|L_1(X_t) - \theta_r n| = O(\sqrt{n}\log^2 n)$ for all $t \in [0, T]$ w.h.p., and the result follows. ◻

We are know ready to prove Lemma 3.27.

*Proof of Lemma 3.27.* By Corollary 3.33 we can assume that properties (ii) to (iv) of $X_0$ and $Y_0$ are preserved throughout the $O(\log n)$ steps of this coupling. Corollary 3.33 also implies that $|L_1(X_t) - \theta_r n| = O(\sqrt{n}\log^2 n)$ for all $t \in [0, T]$ w.h.p. provided $T = O(\log n)$; the same holds for $Y_t$. Thus, we also assume that this property is maintained throughout the $O(\log n)$ steps of the coupling.

As mentioned earlier, the coupling we design in this setting is very similar to the one used in Lemma 3.15. Thus, it is convenient to reuse the notation introduced in the proof of that lemma. The coupling will be a composition of three couplings. First we use Coupling I from the proof of Lemma 3.15 to contract $d_t$ as before. Then, we use a one-step coupling which guarantees that the largest components of $\{X_t\}$ and $\{Y_t\}$ have the same size with probability $\Omega(1)$. Once the two configurations agree on the sizes of their largest component and $d_t \leq \omega(n)$, we use Coupling II.

The activations of $\mathcal{L}(X_t)$ and $\mathcal{L}(Y_t)$ are coupled; i.e., both are active with probability $1/q$ and both are inactive otherwise. The remaining components and the edge re-sampling are coupled using Coupling I. Observe that inequalities (3.14) and (3.15) are valid in this setting. Moreover, the expected number of active vertices from $D(X_t)$ and $D(Y_t)$ differ by at most $O(\sqrt{n}\log^2 n)$, since by assumption $|L_1(X_t) - L_1(Y_t)| = O(\sqrt{n}\log^2 n)$. This implies that $|Q_t| = O(\sqrt{n}\log^2 n)$ w.h.p. by Hoefdding's inequality. (In Lemma 3.15 we had $|Q_t| = O(\sqrt{n}\log n)$ w.h.p.) Therefore, inequality (3.16) is also valid here provided the error term $O(\sqrt{n}\log^2 n)$ is replaced by $O(\sqrt{n}\log^3 n)$. Hence, Claim 3.22 holds and thus $d_T \leq \omega(n)$ after $T = O(\log\log n)$ steps with probability $\Omega(1)$.

We take care next of fixing the difference in size between the two largest components. The activation in $X_t \setminus I_{\rm M}(X_t)$ and $Y_t \setminus I_{\rm M}(Y_t)$ is coupled as in Coupling I and we condition on the event that $\mathcal{L}(X_t)$ and $\mathcal{L}(Y_t)$ are activate; this event occurs with probability $1/q$.

First we show that $\rho_t = O(\sqrt{n})$ with probability $\Omega(1)$. By Corollary 3.21, we have that $|L_1(X_t) - L_1(Y_t)| = O(\sqrt{n})$ with probability $\Omega(1)$. If this is the case, then $||D(X_t)| - |D(Y_t)|| = O(\sqrt{n})$. Also, $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$ and $\sum_{j\geq 2} L_j(Y_t)^2 = O(n)$, and thus Hoeffding's inequality implies that the number of active vertices from $D(X_t)$ and $D(Y_t)$ differ by at most $O(\sqrt{n})$ with probability $\Omega(1)$. Thus, $\rho_t = O(\sqrt{n})$ with probability $\Omega(1)$. The activation of the isolated vertices is then coupled using the binomial coupling from Section 2.4. Since $|I_{\rm M}(X_t)| = |I_{\rm M}(Y_t)| = \Omega(n)$ and $\rho_t = O(\sqrt{n})$, Lemma 2.18 implies that this coupling corrects the difference $\rho_t$ with probability $\Omega(1)$. If this is the case, then by coupling the edge re-sampling bijectively as in Coupling I, we ensure that $L_1(X_{t+1}) = L_1(Y_{t+1})$ and $d_{t+1} \leq \omega(n)$ with probability $\Omega(1)$.

Finally, we use Coupling II until the two configurations have the same component structure. Since each of these couplings succeeds with at least constant probability, the result follows. ◻

## 3.5 Mixing time upper bounds: the $\lambda = \lambda_s$ case

In this section we prove our mixing time upper bound for the CM dynamics for $\lambda = \lambda_s$ and $q > 2$. In particular, the following theorem establishes the upper bound in part (ii) of Theorem 3.11 from Section 3.2.

**Theorem 3.34.** *Let $q > 2$ and $\lambda = \lambda_s$. Then, the mixing time of the CM dynamics is $O(n^{1/3})$.*

The structure of the proof of Theorem 3.34 mimics that of Theorems 3.12 and 3.23 for the sub-critical and super-critical regimes, respectively. We again construct a four-phase coupling by composing a two-phase burn-in period, a coupling to reach two configurations with the same component structure and a final coupling to couple them exactly. For the first, third and fourth phases we are able to reuse Lemmas 3.13, 3.15 and 3.16 from Section 3.3, all of which hold for $\lambda = \lambda_s$ and $q > 2$. However, the second phase of the burn-in period presents significant new challenges.

Recall that after the first phase there is a unique large component, and the goal of the second phase in the sub-critical regime is to reach a configuration where the largest component has size $O(\log n)$. As before, the expected change in size of this component is specified by the drift function $f$. For $\lambda = \lambda_s$ and $q > 2$, the function $f$ has a unique zero $\theta^*$ (see Figure 3.1), and thus the drift can be arbitrarily small near $\theta^*$. In contrast to the super-critical case, where it was only required to get close enough to the zero of $f$, here the evolution of the largest component needs to pass through this regime of "zero drift" and then continue to decrease in size until it reaches a size of $O(\log n)$.

Our approach is to split the second phase of the burn-in period into three parts. Namely, we choose a window $W := [\theta^* n - O(n^{2/3}), \theta^* n + O(n^{2/3})]$, and in the first part we bound the time it takes for the largest component to reach a size in $W$, assuming its initial size was $n$. The magnitude of the drift is roughly given by $f(\theta)n$, and in Section 3.1.2 we showed that $f(\theta) \approx (\theta - \theta^*)^2$ (see Lemma 3.8(i)); hence, $f(\theta)n \approx (\theta - \theta^*)^2 n$. Since $f$ is increasing in $(\theta^*, 1)$, the drift is at least $O(n^{1/3})$ throughout the first part. However, using this very pessimistic lower bound yields only that $O(n^{2/3})$ steps are sufficient to reach $W$. Instead, we split the interval $(\theta^*, 1)$ into $O(\log n)$ subintervals and use a much more precise lower bound for the drift in each subinterval. This gives the desired $O(n^{1/3})$ bound for the number of steps required to reach $W$.

In the second part we bound the time it takes to "escape" $W$ through its left boundary. Inside $W$ the drift is too small to be useful, so it is ignored. Instead, we use the fact that the variance $\sigma^2$ of the process is large; i.e., $\sigma^2 = \Omega(n)$. Since the process is "close" to a martingale (which corresponds to having drift 0), we are able to show that after $O(|W|^2/\sigma^2) = O(n^{1/3})$ steps the process escapes $W$ through its left boundary with probability $\Omega(1)$. For this, we use the tailored supermartingale hitting time bound we proved in Lemma 2.23.

In the third part we show that if the process starts off with a unique large component of size at most $\theta^* n - O(n^{2/3})$, i.e., to the left of $W$, then after $O(n^{1/3})$ steps the largest component will have size $O(\log n)$ with probability $\Omega(1)$. The argument for this part does not differ significantly from the one in the first part. Finally, we point out that the size of the window $W$ is optimized so that each of the three parts takes roughly the same number of steps.

The following lemma is a key part of the proof of Theorem 3.34.

**Lemma 3.35.** *Let $q > 2$ and $\lambda = \lambda_s$. If $X_0$ has at most one large component, then there exists $T = O(n^{1/3})$ such that $L_1(X_T) = O(\log n)$ and $I(X_T) = \Omega(n)$ with probability $\Omega(1)$.*

*Proof of Theorem 3.34.* Let $X_0$ and $Y_0$ be two arbitrary random-cluster configurations. By Lemmas 3.13 and 3.35, after $T = O(n^{1/3})$ steps $L_1(X_T) = O(\log n)$ and $I(X_T) = \Omega(n)$ with probability $\Omega(1)$. The same holds for $Y_T$. Hence, Lemmas 3.15 and 3.16 imply that after an additional $O(\log n)$ steps the two configurations agree with probability $\Omega(1)$. The result then follows from the standard mixing time estimate (2.1). □

Let $\theta^*$ be the unique zero of the function $f$ in the interval $(\theta_{\min}, 1]$ (see Fact 3.6). Lemma 3.35 is direct a consequence of the following three lemmas; these lemmas hold when $q > 2$ and $\lambda = \lambda_s$.

**Lemma 3.36.** *Assume $L_1(X_0) > \theta^* n + Bn^{2/3}$ and $L_2(X_0) < 2n^{11/12}$, where $B > 0$ is a constant independent of $n$. Then, there exists $T = O(n^{1/3})$ such that, with probability $\Omega(1)$, $L_1(X_T) \le \theta^* n + Bn^{2/3}$ and $L_2(X_T) < 2n^{11/12}$.*

**Lemma 3.37.** *Suppose $L_1(X_0) \in [\theta^* n - Bn^{2/3}, \theta^* n + Bn^{2/3}]$ for a sufficiently small constant $B > 0$. If $L_2(X_0) < 2n^{11/12}$, then there exists $T = O(n^{1/3})$ such that, with probability $\Omega(1)$, $L_1(X_T) < \theta^* n - Bn^{2/3}$ and $L_2(X_T) < 2n^{11/12}$.*

**Lemma 3.38.** *Assume $L_1(X_0) < \theta^* n - Bn^{2/3}$ and $L_2(X_0) < 2n^{11/12}$ where $B > 0$ is constant independent of $n$. Then, there exists $T = O(n^{1/3})$ such that, with probability $\Omega(1)$, $L_1(X_T) = O(\log n)$ and $I(X_T) = \Omega(n)$ .*

Let $\varepsilon > 0$ be a small constant (independent of $n$) that we choose later. The following auxiliary facts, which are proved later in Section 3.5.1, will be used in the proofs of these lemmas. Fact 3.39 guarantees additional initial properties of the configuration that will simplify our proofs, while Fact 3.40 establishes useful properties of the one-dimensional process $\{L_1(X_t)\}$ that we will refer to repeatedly.

**Fact 3.39.** *Let $\lambda < q$ and let $X_0$ be a random-cluster configuration.*

*(i) If $X_0$ is such that $L_1(X_0) \ge (\Theta_s + \varepsilon)n$ and $L_2(X_0) < 2n^{11/12}$, then there exists $T = O(\log n)$ such that, with probability $\Omega(1)$, either $L_2(X_T) = O(\log n)$ and $\sum_{j\ge 2} L_j(X_T)^2 = O(n)$ or $L_1(X_T) < (\Theta_s + \varepsilon)n$ and $L_2(X_T) < 2n^{11/12}$.*

*(ii) Suppose $X_0$ is such that $L_1(X_0) \ge (\Theta_s + \varepsilon)n$, $L_2(X_0) = O(\log n)$ and $\sum_{j\ge 2} L_j(X_0)^2 = O(n)$. If $L_1(X_t) \ge (\Theta_s + \varepsilon)n$ for all $t \in [0, T]$, then $L_2(X_t) = O(\log n)$ and $\sum_{j\ge 2} L_j(X_t)^2 = O(n)$ for all $t \in [0, T]$ w.h.p., provided $T = O(n^{1/3})$.*

**Fact 3.40.** *Let $\lambda < q$ and suppose $X_t$ is such that $L_1(X_t) \ge (\Theta_s + \varepsilon)n$, $L_2(X_t) = O(\log n)$ and $\sum_{j\ge 2} L_j(X_t)^2 = O(n)$. Then for sufficiently large $n$:*

*(i)* $\mathrm{Var}(L_1(X_{t+1}) \mid X_t, \mathcal{E}_t) = \Theta(n)$;

*(ii)* $\mathrm{Var}(L_1(X_{t+1}) \mid X_t) = \Omega(n)$*;*

*(iii) There exists a constant* $A > 0$ *such that:*

- $\Pr[L_1(X_{t+1}) \notin [\phi(\theta_t)n - A\sqrt{n}\log n, \phi(\theta_t)n + A\sqrt{n}\log n] \mid X_t, \mathcal{E}_t] = O(n^{-2})$.
- $\Pr[L_1(X_{t+1}) \notin [\phi(\theta_t)n - A\sqrt{n}\log n, \phi(\theta_t)n + A\sqrt{n}\log n] \cup \{L_1(X_t)\} \mid X_t] = O(n^{-2})$.

*(see (3.3) for the definition of* $\phi$*).*

*Proof of Lemma 3.36.* By assumption, $L_1(X_0) > \theta^* n + Bn^{2/3}$ and $L_2(X_0) < 2n^{11/12}$. Since $\theta^* > \Theta_s$ (see Fact 3.6), $L_1(X_0) \ge (\Theta_s + \varepsilon)n$ for small enough $\varepsilon$. Therefore, Fact 3.39 will allow us to assume that $L_2(X_t) = O(\log n)$ and $\sum_{j\ge 2} L_j(X_t)^2 = O(n)$ for all $t \le \tau$ w.h.p., provided $\tau = O(n^{1/3})$. Namely, by Fact 3.39(i) after $O(\log n)$ initial steps these properties are achieved with probability $\Omega(1)$ and by Fact 3.39(ii) they are preserved for $\tau$ steps w.h.p., provided $\tau = O(n^{1/3})$ and that $L_1(X_t) \ge (\Theta_s + \varepsilon)n$ for all $t \le \tau$. (Note that if $L_1(X_t) < (\Theta_s + \varepsilon)n$ for any $t \le \tau$, then there is nothing to prove.)

Let $Y_{t+1} := L_1(X_{t+1}) - \phi(\theta_t)n$ and let

$$\tilde{Y}_t := \begin{cases} Y_t & \text{if } |Y_t| \le A\sqrt{n}\log n; \\ 0 & \text{otherwise,} \end{cases}$$

where $A > 0$ is the constant from Fact 3.40(iii). We shall see that when $\mathcal{L}(X_t)$ is active, $\phi(\theta_t)n$ specifies the expected value of $L_1(X_{t+1})$; hence, $Y_t$ corresponds roughly to the fluctuation of $L_1(X_t)$ around its mean. Moreover, the sequence of random variables $\tilde{Y}_1, \tilde{Y}_2, \ldots$ ignores the very unlikely large fluctuations, which will simplify our arguments.

Now, let $\{Z_t\}$ be the stochastic process given by $Z_{t+1} = \phi(Z_t/n)n + \tilde{Y}_t$ with $Z_0 = L_1(X_0)$. Observe that if the largest component in the configuration is forced to be always active and $Y_t \le A\sqrt{n}\log n$ for all $t \ge 0$, then the one-dimensional process $\{Z_t\}$ keeps track of the size of the largest component.

We show first that $Z_T \le \theta^* n + Bn^{2/3}$ with probability $\Omega(1)$ for some $T = O(n^{1/3})$. Suppose $Z_0 \in [\theta^* n + 2^k Bn^{2/3}, \theta^* n + 2^{k+1}Bn^{2/3}]$ and let $I_k := [\theta^* n + 2^k Bn^{2/3}, \theta^* n + 2^{k+3}Bn^{2/3}]$, where $k$ is a positive integer. (If either $\theta^* n + 2^{k+1}Bn^{2/3} > n$ or $\theta^* n + 2^{k+3}Bn^{2/3} > n$, then the right boundary of the corresponding interval is replaced by $n$.)

We consider first the case when $k$ is small; i.e., $k$ is such that $2^k/n^{1/3} \to 0$. For ease of notation let $\hat{\theta}_t = Z_t/n$. Lemma 3.8(i) implies that for $Z_t \in I_k$ there exist constants $\delta, \delta_1, \delta_2 > 0$ such that

$$\begin{aligned} \phi(\hat{\theta}_t)n &\le Z_t - \delta_1|\hat{\theta}_t - \theta^*|^2 n + \delta_2|\hat{\theta}_t - \theta^*|^3 n \\ &\le Z_t - \delta|\hat{\theta}_t - \theta^*|^2 n \\ &\le Z_t - \delta 2^{2k} B^2 n^{1/3}, \end{aligned} \tag{3.27}$$

and thus

$$Z_{t+1} \le Z_t - \delta 2^{2k} B^2 n^{1/3} + \tilde{Y}_{t+1}.$$

From Fact 3.19(i) we know that $\mathrm{E}[Y_{t+1}|X_t, \mathcal{E}_t] \leq 3n^{1/4}$ and from Fact 3.40(i) that $\mathrm{Var}(Y_{t+1}|X_t, \mathcal{E}_t) = O(n)$. Since conditioned on $\mathcal{E}_t$, $Y_{t+1} = \tilde{Y}_{t+1}$ with probability $1 - O(n^{-2})$ by Fact 3.40(iii), we have that $\mathrm{E}[\tilde{Y}_{t+1}|X_t, \mathcal{E}_t] \leq 4n^{1/4}$ and $\mathrm{Var}(\tilde{Y}_{t+1}|X_t, \mathcal{E}_t) = O(n)$.

We now claim that the process $\{Z_t\}$ in the interval $I_k$ satisfies all the assumptions of Lemma 2.22. To see this, observe that w.l.o.g. we can shift the process by subtracting $\theta^* n + 2^k Bn^{2/3}$ and then take $M = 7 \cdot 2^k Bn^{2/3}$, $\alpha = 1/7$, $\kappa = 2$, $D = \delta 2^{2k} B^2 n^{1/3}$ and $C = \delta' 2^k$, where $\delta'$ is a sufficiently small positive constant. Hence, after $T_k = O(2^{-k} n^{1/3})$ steps, $Z_{T_k} < \theta^* n + 2^k Bn^{2/3}$ with probability $1 - 2\exp(-c2^k)$, for some constant $c > 0$ independent of $n$ and $k$.

We proceed in a similar manner when $2^k/n^{1/3} \nrightarrow 0$; i.e., $2^k = \Omega(n^{1/3})$. In this case, Lemma 3.8(iii) implies that $\phi(\hat{\theta}_t)n \leq Z_t - \gamma n$ for some constant $\gamma > 0$, and so $Z_{t+1} \leq Z_t - \gamma n + \tilde{Y}_{t+1}$. Lemma 2.22 then implies that for some $T_k = O(1)$, $Z_{T_k} \leq \theta^* n + 2^k Bn^{2/3}$ with probability $\Omega(1)$.

Putting all this together, if $Z_0 \in [\theta^* n + 2^k Bn^{2/3}, \theta^* n + 2^{k+1} Bn^{2/3}]$ for some $k \geq 0$, then after $T = \sum_{j=0}^{k} T_j = O(n^{1/3})$ steps $Z_T \leq \theta^* n + Bn^{2/3}$ with probability $\Omega(1)$.

We show next that if $Z_T \leq \theta^* n + Bn^{2/3}$ for some $T = O(n^{1/3})$, then there exists $T' = O(n^{1/3})$ such that $L_1(X_{T'}) \leq \theta^* n + Bn^{2/3}$ w.h.p. For this, let

$$\tilde{Z}_{t+1} = \begin{cases} \phi(\tilde{Z}_t/n)n + \tilde{Y}_{t+1} & \text{if } \mathcal{L}(X_t) \text{ is active;} \\ \tilde{Z}_t & \text{otherwise,} \end{cases}$$

and let $\tilde{Z}_0 = Z_0 = L_1(X_0)$. The process $\{\tilde{Z}_t\}$ is simply "lazy" version of $\{Z_t\}$; namely, $\tilde{Z}_{t+1} = \tilde{Z}_t$ with probability $1 - q^{-1}$, and otherwise $\{\tilde{Z}_t\}$ evolves like $\{Z_t\}$. Therefore, if $Z_T \leq \theta^* n + Bn^{2/3}$ for some $T = O(n^{1/3})$, then $\tilde{Z}_{T'} \leq \theta^* n + Bn^{2/3}$ for some $T' = O(n^{1/3})$ w.h.p.

Now, if $\mathcal{L}(X_t)$ is inactive, then $\mathcal{L}(X_{t+1}) = \mathcal{L}(X_t)$ with probability $1 - O(n^{-1})$ by Fact 3.17(i). Hence,

$$\Pr[\tilde{Z}_{t+1} \neq L_1(X_{t+1}) \mid Z_t = L_1(X_t), \neg\mathcal{E}_t] = O(n^{-1}).$$

Also, if $\mathcal{L}(X_t)$ is active, then Fact 3.40(iii) implies that $Y_{t+1} = \tilde{Y}_{t+1}$ with probability $1 - O(n^{-2})$, so

$$\Pr[\tilde{Z}_{t+1} \neq L_1(X_{t+1}) \mid Z_t = L_1(X_t), \mathcal{E}_t] = O(n^{-2}),$$

and thus

$$\Pr[\tilde{Z}_{t+1} = L_1(X_{t+1}) \mid Z_t = L_1(X_t)] = 1 - O(n^{-1}).$$

Inductively, we then get that $\tilde{Z}_{T'} = L_1(X_{T'})$ w.h.p., and hence $L_1(X_{T'}) \leq \theta^* n + Bn^{2/3}$ with probability $\Omega(1)$. Since also $L_2(X_{T'}) = O(\log n)$ w.h.p. (see discussion at the beginning of the proof), then the result follows. □

*Proof of Lemma 3.37.* By Fact 3.39, we shall assume that $L_2(X_t) = O(\log n)$ and $\sum_{j \geq 2} L_j(X_t)^2 = O(n)$ for all $t \leq T$ w.h.p., where $T = O(n^{1/3})$ will be determined later. Let $Z_t := L_1(X_t) - \theta^* n$; Fact 3.19(i) implies

$$-\frac{f(\theta_t)n}{q} - 2n^{1/4} \leq \mathrm{E}[Z_{t+1} - Z_t \mid X_t] \leq -\frac{f(\theta_t)n}{q} + 2n^{1/4} \leq 2n^{1/4}.$$

Moreover, if $L_1(X_t) \in [\theta^* n - 2Bn^{2/3}, \theta^* n + 2Bn^{2/3}]$, then by Fact 3.8(i) there exists a positive constant $\delta$ such that $|f(\theta_t)|n \le \delta B^2 n^{1/3}$. Hence,

$$-2\delta B^2 n^{1/3} \le \mathrm{E}[Z_{t+1} - Z_t \mid X_t] \le 2\delta B^2 n^{1/3}$$

for sufficiently large $n$. By Fact 3.40(ii), we also know that $\mathrm{Var}(Z_{t+1}|X_t) = \Omega(n)$ and by part (iii) of the same fact we get that, with probability $1 - O(n^{-2})$, either $L_1(X_{t+1}) = L_1(X_t)$ or $L_1(X_{t+1}) \in [\phi(\theta_t)n - a\sqrt{n}\log n, \phi(\theta_t)n + a\sqrt{n}\log n]$ for some constant $a > 0$. Since $|\theta_t - \phi(\theta_t)|n = |f(\theta_t)|n \le \delta B^2 n^{1/3}$, we have that sufficiently large $n$

$$\Pr[|Z_{t+1} - Z_t| > 2a\sqrt{n}\log n \mid X_t] = O(n^{-2}).$$

Also, $|Z_{t+1} - Z_t| < n$ and by assumption $Z_0 \in [-Bn^{2/3}, Bn^{2/3}]$. Hence, Lemma 2.23 with $M = n$, $A = 2Bn^{2/3}$, $C = 2\delta B^2 n^{1/3}$, $\sigma^2 = \Omega(n)$, $L = 2a\sqrt{n}\log n$, $\varepsilon = O(n^{-2})$ and $\kappa = 4$ implies that, for $B$ small enough, $Z_T < -Bn^{2/3}$ with probability $\Omega(1)$ for some $T = O(n^{1/3})$ as desired. □

*Proof of Lemma 3.38.* We show first that for some $T = O(n^{1/3})$, $L_1(X_T) \le (\Theta_s + \varepsilon)n$ and $L_2(X_T) < 2n^{11/12}$ with probability $\Omega(1)$. The proof of this fact is very similar to that of Lemma 3.36. Indeed, by Fact 3.39 we shall again assume that $L_2(X_t) = O(\log n)$ and $\sum_{j \ge 2} L_j(X_t)^2 = O(n)$ for all $t \le T = O(n^{1/3})$, and we consider the same stochastic process $\{Z_t\}$ and sequences of random variables $\{Y_t\}$ and $\{\tilde{Y}_t\}$.

Suppose $Z_0 \in [\theta^* n - 2^{k+3}Bn^{2/3}, \theta^* n - 2^{k+2}Bn^{2/3}]$ and let $I'_k := [\theta^* n - 2^{k+3}Bn^{2/3}, \theta^* n - 2^k Bn^{2/3}]$, where $k \ge 0$ is an integer and if $\theta^* n - 2^{k+3}Bn^{2/3} < (\Theta_s + \varepsilon)n$, the left boundaries of both these intervals is set to $(\Theta_s + \varepsilon)n$ instead. If $Z_t \in I'_k$ and $k$ is small (i.e., $2^k/n^{1/3} \to 0$), then Lemma 3.8(i) guarantees that (3.27) also holds in this setting. Moreover, from Facts 3.19(i), 3.40(i) and 3.40(iii) we deduce as in the proof of Lemma 3.36 that $\mathrm{E}[\tilde{Y}_{t+1}|X_t, \mathcal{E}_t] \le 4n^{1/4}$ and $\mathrm{Var}(\tilde{Y}_{t+1}|X_t, \mathcal{E}_t) = O(n)$. Thus, Lemma 2.22 applied to the process $\{Z_t\}$ in $I'_k$ (with $M = 7 \cdot 2^k Bn^{2/3}$, $\alpha = 4/7$, $\kappa = 2$, $D = \delta 2^{2k} B^2 n^{1/3}$ and $C = \delta' 2^k$, where $\delta' > 0$ is a sufficiently small constant) implies that after $T_k = O(2^{-k} n^{1/3})$ steps, $Z_{T_k} \le \theta^* n - 2^{k+3}Bn^{2/3}$ with probability $1 - 2\exp(-c2^k)$, for some constant $c > 0$ independent of $n$ and $k$. The same holds if $2^k = \Omega(n^{1/3})$ for some $T_k = O(1)$ with probability $\Omega(1)$. Since $\sum_{j \ge 0} T_j = O(n^{1/3})$, $L_1(X_T) \le (\Theta_s + \varepsilon)n$ with probability $\Omega(1)$ for some $T = O(n^{1/3})$. Finally, note that $L_2(X_T) < 2n^{11/12}$ w.h.p. and thus the proof of the first part is complete.

Consider next the interval $J := [(\Theta_s - 2\varepsilon)n, (\Theta_s + 2\varepsilon)n]$. Suppose $L_1(X_0) \in J$ and $L_2(X_0) < 2n^{11/12}$. From Fact 3.18 and a union over the steps, we get $L_2(X_t) < 2n^{11/12}$ for all $t \le T$ w.h.p., provided $T = O(\log n)$. Assuming this, Fact 3.19(ii) and Lemma 3.8(iii) imply

$$\mathrm{E}[L_1(X_{t+1}) - L_1(X_t) \mid X_t] \le -\gamma n + O(n^{1/4})$$

for some constant $\gamma > 0$, provided $L_1(X_t) \in J$. Therefore, by Corollary 2.21 after $O(1)$ steps the largest component has size at most $(\Theta_s - \varepsilon)n$ with probability $\Omega(1)$. Fact 3.20 then implies that after $O(\log n)$ additional steps the largest component has at most $O(\log n)$ vertices with probability $\Omega(1)$.

Finally, if $L_1(X_t) = O(\log n)$ then Facts 3.17(i), 3.17(ii) and Lemma 2.4 imply that $I(X_{t+1}) = \Omega(n)$ w.h.p. Since also $L_1(X_{t+1}) = O(\log n)$ w.h.p. by Fact 3.17(iv), the result follows. □

### 3.5.1 Proofs of auxiliary facts

It remains to go back and prove Facts 3.39 and 3.40.

*Proof of Fact 3.39.* By Fact 3.18 $L_2(X_t) < 2n^{11/12}$ for all $t \le T$ w.h.p., provided $T = O(\log n)$. Assuming this, if $L_1(X_t) < (\Theta_s + \varepsilon)n$ for any $t \le T$, then the proof of part (i) is complete. Thus, we shall also assume that $L_1(X_t) \ge (\Theta_s + \varepsilon)n$ for all $t \le T$. Now, for a suitable $T = O(\log n)$, after the first $T$ steps every initial component in $X_0$ has been activated w.h.p. Moreover, since $L_1(X_t) \ge (\Theta_s + \varepsilon)n$, by Facts 3.17(i) and 3.17(ii) every *new* small component in $X_{t+1}$ has size $O(\log n)$ with probability $1 - O(n^{-1})$. A union bound then implies that $L_2(X_T) = O(\log n)$ w.h.p. To establish that $\sum_{j\ge2} L_j(X_T)^2 = O(n)$ we can use an argument analogous to the one used in the proof of Claim 3.32; the details are thus omitted. This completes the proof of part (i).

For part (ii), note that if $L_1(X_t) \ge (\Theta_S+\varepsilon)n$ and $L_2(X_t) = O(\log n)$, then Facts 3.17(i) and 3.17(ii) imply that $L_2(X_{t+1}) = O(\log n)$ with probability $1-O(n^{-1})$. Again, an analogous argument to the one in the proof of Claim 3.32 yields that $\sum_{j\ge2} L_j(X_{t+1})^2 = O(n)$ with probability $1-O(n^{-1})$. Part (ii) then follows from a union bound over the $T = O(n^{1/3})$ steps. □

*Proof of Fact 3.40.* Let $N_t$ be the size of the largest (new) component created at time $t$, $\mu_t = L_1(X_t)+\frac{n-L_1(X_t)}{q}$, $M_t = A_t - \mu_t$ and let $\ell_m(\theta_t)$ denote the size of the largest component of a $G(\mu_t + m, p)$ random graph. By the law of total variance, we have

$$\mathrm{Var}(N_t \mid X_t, \mathcal{E}_t) = \mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] + \mathrm{Var}_{M_t}(\mathrm{E}[N_t \mid M_t] \mid X_t, \mathcal{E}_t). \tag{3.28}$$

We bound each term in the right hand side of (3.28) separately. If $A_t \in J_t$, then $M_t \in J_t' = [-\xi(r), \xi(r)]$ ($J_t$ was defined in (3.6)). Hence, Fact 3.17(ii) (with $r = O(\log n)$) implies that

$$\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] \le \sum_{m\in J_t'} \mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t, M_t = m] + O(1).$$

Now, if $M_t = m$, then $N_t$ and $\ell_m(\theta_t)$ have the same distribution. Hence,

$$\begin{aligned}\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] &\le \sum_{m\in J_t'} \mathrm{Var}(\ell_m(\theta_t)) \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1) \\ &\le \max_{m\in J_t'} \mathrm{Var}(\ell_m(\theta_t)) + O(1).\end{aligned}$$

Since by assumption $L_1(X_t) \ge (\Theta_s + \varepsilon)n$, $G(\mu_t + m, p)$ is super-critical for $m \in J_t'$ by Fact 3.17(ii). Therefore, Lemma 2.8 implies that $\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] = O(n)$.

Similarly, we obtain that $\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] = \Omega(n)$, and thus

$$\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] = \Theta(n). \tag{3.29}$$

We show next that $\mathrm{Var}_{M_t}(\mathrm{E}[N_t \mid M_t] \mid X_t, \mathcal{E}_t) = O(n)$. For this, let

$$g(m) := \mathrm{E}[N_t \mid M_t = m] = \mathrm{E}[\ell_m(\theta_t)],$$

and let $\Gamma := \mathrm{E}_{M_t}[g(M_t) \mid X_t, \mathcal{E}_t]$. Then,

$$\begin{aligned} \mathrm{Var}_{M_t}(\mathrm{E}[N_t \mid M_t] \mid X_t, \mathcal{E}_t) &= \mathrm{Var}_{M_t}(g(M_t) \mid X_t, \mathcal{E}_t) \\ &= \mathrm{E}_{M_t}[(g(M_t) - \Gamma)^2 \mid X_t, \mathcal{E}_t] \\ &\le \sum_{m \in J_t'} \mathrm{E}_{M_t}[(g(m) - \Gamma)^2] \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1), \end{aligned} \tag{3.30}$$

where the last inequality follows from Fact 3.17(ii). Since $G(\mu_t + m, p)$ is super-critical for $m \in J_t'$, Lemma 2.8 implies that for $m \in J_t'$

$$\phi(\theta_t)(n+m) - 2n^{1/4} \le g(m) = \mathrm{E}[\ell_m(\theta_t)] \le \phi(\theta_t)(n+m) + 2n^{1/4}. \tag{3.31}$$

Also, by Fact 3.17(ii)

$$\begin{aligned} \Gamma &\le \sum_{m \in J_t'} g(m) \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1) \\ &= \sum_{m \in J_t'} \mathrm{E}[\ell_m(\theta_t)] \Pr[M_t = m \mid X_t, \mathcal{E}_t] + O(1) \\ &\le \phi(\theta_t)n + \phi(\theta_t)\,\mathrm{E}[M_t \mid X_t, \mathcal{E}_t] + 3n^{1/4} \\ &= \phi(\theta_t)n + 3n^{1/4}, \end{aligned}$$

where the last inequality follows from Lemma 2.7. Similarly, we can show that $\Gamma \ge \phi(\theta_t)n - 3n^{1/4}$. Together with (3.31) this implies that for any $m \in J_t'$

$$\phi(\theta_t)m - 5n^{1/4} \le g(m) - \Gamma \le \phi(\theta_t)m + 5n^{1/4},$$

and thus

$$(g(m) - \Gamma)^2 \le (\phi(\theta_t)m)^2 + O(n^{1/4}) \cdot |m| + O(\sqrt{n}).$$

Plugging this bound into (3.30), we get

$$\mathrm{Var}_{M_t}(\mathrm{E}[N_t \mid M_t] \mid X_t, \mathcal{E}_t) \le \phi(\theta_t)^2\,\mathrm{E}[M_t^2 \mid X_t, \mathcal{E}_t] + O(n^{1/4}) \cdot \mathrm{E}[|M_t| \mid X_t, \mathcal{E}_t] + O(\sqrt{n}). \tag{3.32}$$

By Fact 3.17(ii), if $\mathcal{L}(X_t)$ is active, then $M_t \in J_t'$ with probability $1 - O(n^{-1})$; therefore, $\mathrm{E}[|M_t| \mid X_t, \mathcal{E}_t] = O(\sqrt{n}\log n)$. To bound $\mathrm{E}[M_t^2 \mid X_t, \mathcal{E}_t]$, let $\xi_2, \xi_3, \ldots$ be independent Bernoulli random variables with parameter $1/q$. Then,

$$\begin{aligned} \mathrm{E}[M_t^2 \mid X_t, \mathcal{E}_t] = \mathrm{E}[(A_t - \mu_t)^2 \mid X_t, \mathcal{E}_t] &= \mathrm{Var}(A_t \mid X_t, \mathcal{E}_t) \\ &= \mathrm{Var}\left(\sum\nolimits_{j \ge 2} \xi_j \cdot L_j(X_t) \mid X_t, \mathcal{E}_t\right) \\ &= \sum\nolimits_{j \ge 2} L_j(X_t)^2 \cdot \mathrm{Var}\left(\xi_j \mid X_t, \mathcal{E}_t\right) = O(n), \end{aligned}$$

where in the last equality we used the fact that $\mathrm{Var}\left(\xi_j \mid X_t, \mathcal{E}_t\right) = \Theta(1)$ for all $j \geq 2$ and that $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$ by assumption. Plugging these bounds into (3.32), we get that $\mathrm{Var}_{M_t}(\mathrm{E}[N_t \mid M_t] \mid X_t, \mathcal{E}_t) = O(n)$. Together with (3.29) this implies that $\mathrm{Var}(N_t \mid X_t, \mathcal{E}_t) = O(n)$.

Now, when $\mathcal{L}(X_t)$ is activated, $L_1(X_{t+1}) \neq N_t$ only when the size of the largest inactive component is at least $N_t$. Since $G(\mu_t + m, p)$ is super-critical with probability $1 - O(n^{-1})$ and $L_2(X_t) = O(\log n)$, Lemma 2.7 implies that $\Pr[L_1(X_{t+1}) \neq N_t] = \exp(-\Omega(n))$. Hence, $\mathrm{Var}(L_1(X_{t+1}) \mid X_t, \mathcal{E}_t) = O(n)$, which establishes part (i).

For part (ii), note that by the law of total variance

$$\begin{aligned}\mathrm{Var}(N_t \mid X_t) &\geq \mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t] \\ &\geq \frac{1}{q}\,\mathrm{E}_{M_t}[\mathrm{Var}(N_t \mid M_t) \mid X_t, \mathcal{E}_t] = \Omega(n).\end{aligned} \tag{3.33}$$

Using Fact 3.17 it is straightforward to check that $\Pr[L_1(X_{t+1}) \neq N_t] = \exp(-\Omega(n))$ and thus $\mathrm{Var}(L_1(X_{t+1}) \mid X_t) = \Omega(n)$, which establishes part (ii).

Finally, for part (iii) observe that when $\mathcal{L}(X_t)$ is inactive, $N_t = O(\sqrt{n})$ with probability $1 - O(n^{-2})$ by Fact 3.17(i); since by assumption $L_1(X_t) \geq (\Theta_s + \varepsilon)n$, $L_1(X_{t+1}) = L_1(X_t)$ with probability $1 - O(n^{-2})$. When $\mathcal{L}(X_t)$ is active, Fact 3.17(ii) implies that $G(A_t, p)$ is super-critical with probability $1 - O(n^{-2})$. Also, $\xi(r) = O(\sqrt{n}\log n)$ since $L_2(X_t) = O(\log n)$. Therefore, by Lemma 2.7

$$\phi(\theta_t)n - O(\sqrt{n}\log n) \leq N_t \leq \phi(\theta_t)n + O(\sqrt{n}\log n)$$

with probability $1-O(n^{-2})$. Part (iii) then follows from the fact that $L_1(X_{t+1}) = N_t$ with probability $1 - \exp(-\Omega(n))$ when $\mathcal{L}(X_t)$ is active. □

## 3.6 Mixing time upper bounds: the $\lambda = q$ case

In this section we prove the upper bound portion of part (iv) of Theorem 3.11 from Section 3.2. In particular we establish the following theorem.

**Theorem 3.41.** *Let $\lambda = q$ and $q > 2$. Then, the mixing time of the CM dynamics is $O(\log n)$.*

This proof follows closely that of Theorem 3.23 in Section 3.4. In fact, we would have been able to use the same proof if Lemma 3.24 held when $\lambda = q$, but this is not the case. After the first phase of the burn-in period, where a configuration with a unique large component is reached, Lemma 3.24 guarantees that after one more step the large component will be of linear size with probability $\Omega(1)$. This is because, when the large component is active, the percolation step is super-critical w.h.p. However, as was pointed out in the proof of Lemma 3.24, this is not the case for $\lambda = q$ where the percolation is critical or close to critical w.h.p., and thus the largest new component will be of sublinear size. Our approach is to re-analyze the first phase of the burn-in period for $\lambda = q$. We show first that after $O(\log n)$ steps there is a unique large component of size $\Omega(n^{2/3})$ with probability $\Omega(1)$. Then, we use the random graph facts for the critical and near-critical regimes in Section 2.3.1 to argue that, after at most $O(\log n)$ steps, the size of this largest component becomes linear. It will also be crucial in our proofs to control the magnitude of the sum of the squares of the sizes of the components, since this quantity determines the variance of the number of active vertices. This again requires us to use the critical random graph facts from Section 2.3.1.

Theorem 3.41 follows immediately from the following two lemmas.

**Lemma 3.42.** *Let $\lambda = q$ and let $X_0$ be an arbitrary random-cluster configuration. Then, there exists $T = O(\log n)$ such that, with probability $\Omega(1)$, $\sum_{j\geq 2} L_j(X_T)^2 = O(n^{4/3})$ and $L_1(X_T) \geq Bn^{2/3}$, for any desired constant $B \geq 0$.*

**Lemma 3.43.** *Let $\lambda = q$ and $q > 2$. Suppose $\sum_{j\geq 2} L_j(X_0)^2 = O(n^{4/3})$ and $L_1(X_0) \geq Bn^{2/3}$ for a sufficiently large constant $B > 0$. Then, there exists $T = O(\log n)$ such that, with probability $\Omega(1)$, $L_1(X_T) = \Omega(n)$ and $L_2(X_T) = O(n^{2/3})$.*

*Proof of Theorem 3.41.* Let $X_0$ and $Y_0$ be two arbitrary random-cluster configurations. By Lemmas 3.42 and 3.43, after $T = O(\log n)$ steps $L_1(X_T) = \Omega(n)$ and $L_2(X_T) = O(n^{2/3})$ with probability $\Omega(1)$; the same holds for $Y_T$. Since $\Theta_S = 1 - q/\lambda = 0$ when $\lambda = q$, the result follows from Lemmas 3.25, 3.26, 3.27 and 3.16. □

We now provide the proofs of Lemmas 3.42 and 3.43.

*Proof of Lemma 3.42.* Let $C(X_t)$ denote the set of connected components of $X_t$. Let $S_0 = \{\mathcal{L}(X_0)\}$, and given $S_t$, $S_{t+1}$ is obtained as follows:

(i) $S_{t+1} = S_t$;

(ii) every component in $S_t$ activated by the CM dynamics at time $t$ is removed from $S_{t+1}$; and

(iii) the largest new component (breaking ties arbitrarily) is added to $S_{t+1}$.

(Note that $S_t \subseteq C(X_t)$ for all $t \geq 0$.)

We analyze first how the size of $S_t$, i.e., the number of components in $S_t$ for which we use $|S_t|$, fluctuates. Observe that for any $t \geq 1$

$$\mathrm{E}[|S_t| \mid S_{t-1}] = \left(1 - \frac{1}{q}\right)|S_{t-1}| + 1,$$

and so

$$\mathrm{E}[|S_t|] = \left(1 - \frac{1}{q}\right)^t |S_0| + \sum_{k=0}^{t-1} \left(1 - \frac{1}{q}\right)^k \leq q + 1.$$

Thus, for any given $t \geq 0$, $|S_t| < 2(q+1)$ with probability at least $1/2$ by Markov's inequality.

Now, let $W_t := \sum_{j\geq 2} |C_j(t)|^2$ where $C_j(t)$ is the $j$-th largest *new* component created at time $t$. By Lemma 2.16, $\mathrm{E}[W_t \mid X_t] = O(n^{4/3})$. Hence, if $Z_t := \sum_{C \in C(X_t)\setminus S_t} |C|^2$, we have

$$\mathrm{E}[Z_{t+1} \mid X_t] = Z_t - \frac{Z_t}{q} + \mathrm{E}[W_t \mid X_t] \leq \left(1 - \frac{1}{q}\right) Z_t + O(n^{4/3}).$$

Lemma 2.24 then implies that there exists $T = O(\log n)$ such that $Z_T = O(n^{4/3})$ with probability $\Omega(1)$.

We show next that at time $T+1$ the configuration has all the desired properties with probability $\Omega(1)$. Suppose all the components in $S_T$ are activated at time $T+1$. Since $|S_T| < 2(q+1)$ with probability at least $1/2$, this occurs with probability at least $1/(2q^{2(q+1)})$. If this is the case, Hoeffding's inequality implies that

$$A_t \geq m + \frac{n-m}{q} - \gamma n^{2/3} =: M$$

with probability $\Omega(1)$, where $m$ is the total number of vertices in $S_T$ and $\gamma > 0$ is a suitable constant. If $A_t = M$, then the largest new component is the largest component of a $G(M, \frac{1+\varepsilon}{M})$ random graph with $\varepsilon = qM/n - 1 \geq -\gamma q n^{-1/3}$. Thus, using monotonicity and Lemma 2.17 we deduce that $\Pr[|C_1(T+1)| \geq Bn^{2/3}] = \Omega(1)$ for any desired constant $B > 0$, and thus $\Pr[L_1(X_{T+1}) \geq Bn^{2/3}] = \Omega(1)$. Finally, since $Z_T = O(n^{4/3})$ and $\mathrm{E}[W_T] = O(n^{4/3})$ by Lemma 2.16, $\mathrm{E}[Z_{T+1}] \leq Z_T + \mathrm{E}[W_T] = O(n^{4/3})$, which concludes the proof. $\square$

*Proof of Lemma 3.43.* Let $Z_t := \sum_{j\geq 2} L_j(X_t)^2$ and $W_t := \sum_{j \geq 1} |C_j(t)|^2$, where $C_j(t)$ is the $j$-th largest *new* component created at time $t$. Let $w_t := L_1(X_t)/n^{2/3}$ and let $\delta$ be a small positive constant we choose later. The result follows by an inductive argument from the following two claims.

**Claim 3.44.**

*(i) Suppose $Z_t = O(n^{4/3})$ and $B \leq w_t \leq \delta n^{1/3}$. Then, there exist constants $\alpha > 1$ and $A, c > 0$ such that*

$$\Pr\left[L_1(X_{t+1}) > \alpha L_1(X_t) \mid X_t, \mathcal{E}_t\right] \geq 1 - A\exp(-cw_t^2);$$

*(ii)* $\Pr[L_1(X_{t+1}) \geq L_1(X_t) \mid X_t, \neg\mathcal{E}_t] = 1.$

**Claim 3.45.** *Suppose $Z_t = O(n^{4/3})$, $w_t \geq B$ and let $C > 0$ be a fixed large constant. Then:*

*(i) There exist constants $a, b > 0$ such that*

$$\Pr\left[Z_{t+1} < Z_t + \frac{Cn^{4/3}}{\sqrt{w_t}} \,\middle|\, X_t, \mathcal{E}_t\right] \geq 1 - \frac{a}{\sqrt{w_t}} - 2\exp(-bw_t^2);$$

*(ii) There exists a constant $d > 0$ such that*

$$\Pr\left[Z_{t+1} < Z_t + \frac{Cn^{4/3}}{\sqrt{w_t}} \,\middle|\, X_t, \neg\mathcal{E}_t\right] \geq 1 - \frac{d}{\sqrt{w_t}}.$$

Let $\hat{t}_0$ be the first time the largest component is activated and let $\tilde{t}_0 = \min\{\hat{t}_0, qw_0^{1/4}\}$. By Claim 3.44(ii), $w_t \geq w_0$ for all $t \leq \hat{t}_0$. Hence, if $Z_t = O(n^{4/3})$ and $t \leq \hat{t}_0$, then Claim 3.45(ii) implies that $Z_{t+1} < Z_t + \frac{Cn^{4/3}}{\sqrt{w_0}}$ with probability at least $1 - \frac{d}{\sqrt{w_0}}$. From this, it follows inductively that

$$Z_{\tilde{t}_0} < Z_0 + \frac{\tilde{t}_0 C n^{4/3}}{\sqrt{w_0}} < Z_0 + \frac{qCn^{4/3}}{w_0^{1/4}}$$

with probability

$$\left(1 - \frac{d}{\sqrt{w_0}}\right)^{\tilde{t}_0} \geq 1 - dqw_0^{-1/4}.$$

Observe that $\hat{t}_0$ is a geometric random variable with parameter $1/q$, and thus by Markov's inequality $\Pr[\hat{t}_0 \geq qw_0^{1/4}] \leq w_0^{-1/4}$. Hence, $\hat{t}_0 \tilde{t}_0$ with probability at least $1 - w_0^{-1/4}$, and thus

$$Z_{\hat{t}_0} < Z_0 + \frac{qCn^{4/3}}{w_0^{1/4}}$$

with probability $1-(dq+1)w_0^{-1/4}$ by a union bound. This, together with Claims 3.44(i) and 3.45(i) imply by a union bound that

$$\begin{aligned}\Pr\left[L_1(X_{\hat{t}_0+1}) > \alpha L_1(X_0)\,,\, Z_{\hat{t}_0+1} < Z_0 + \frac{(q+1)Cn^{4/3}}{w_0^{1/4}}\right] &\geq 1 - A'\exp(-c'w_0^2) - \frac{a}{\sqrt{w_0}} - \frac{dq+1}{w_0^{1/4}} \\ &\geq 1 - A'\exp(-c'w_0^2) - \frac{a+dq+1}{w_0^{1/4}}, \quad (3.34)\end{aligned}$$

for suitable constants $A', c' > 0$.

Now, let $\tilde{T}$ be the first time $L_1(X_t) \geq \delta n$ and $L_2(X_t) = O(n^{2/3})$. Also, let $\hat{T} = 2q\log_\alpha(\delta B^{-1}n^{1/3})$, $T := \min\{\tilde{T}, \hat{T}\}$ and let $0 \leq \hat{t}_0 < \hat{t}_1 < \cdots < \hat{t}_K \leq T$ be the random times at which the largest component in the configuration is activated. From (3.34), it follows inductively that

$$\begin{aligned}\Pr\left[L_1(X_{\hat{t}_K+1}) > \alpha^K L_1(X_0)\,,\, Z_{\hat{t}_K+1} < Z_0 + M\right] &\geq \prod_{k=0}^{K} 1 - A'\exp(-c'w_k^2) - \frac{a+dq+1}{w_k^{1/4}}, \\ &\geq \prod_{k=0}^{K} 1 - A'\exp(-c'w_0^2\alpha^{2k}) - \frac{a+dq+1}{w_0^{1/4}\alpha^{k/4}}, \quad (3.35)\end{aligned}$$

where

$$M := (q+1)Cn^{4/3}\sum_{k=0}^{K}\frac{1}{w_k^{1/4}} = \frac{(q+1)Cn^{4/3}}{w_0^{1/4}}\sum_{k=0}^{K}\frac{1}{\alpha^{k/4}} = O(n^{4/3}).$$

Since by assumption $w_0 \geq B$ and $B$ is large, it follows from (3.35) that for a sufficiently large constant $c'' > 0$

$$\Pr\left[L_1(X_{\hat{t}_K+1}) > \alpha^K L_1(X_0)\,,\, Z_{\hat{t}_K+1} < O(n^{4/3})\right] \geq \prod_{k=0}^{K}\exp\left(-\frac{c''}{\alpha^{k/4}}\right) = \Omega(1).$$

Finally, a Chernoff bound implies that $K \geq \log_{\alpha}(2B^{-1}n^{1/4})$ w.h.p. Hence, $L_1(X_{\hat{t}_K+1}) > \alpha^K Bn^{2/3} = \delta n$ and $L_2(X_{\hat{t}_K+1}) = O(n^{2/3})$ with probability $\Omega(1)$, which completes the proof. □

We complete the proof of Lemma 3.43 by providing the missing proofs of Claims 3.44 and 3.45.

*Proof of Claim 3.44.* Let $\mu_t := L_1(X_t) + \frac{n-L_1(X_t)}{q}$, $J_t := [\mu_t - \gamma w_t n^{2/3}, \mu_t + \gamma w_t n^{2/3}]$ and $m := \mu_t - \gamma w_t n^{2/3}$, where $\gamma > 0$ is a small constant we choose later. Also, let $\varepsilon := mq/n - 1 = \Delta w_t n^{-1/3}$ with $\Delta := q - 1 - \gamma q$. If $A_t = m$, then $C_1(t)$ is the largest component of a $G(m, \frac{1+\varepsilon}{m})$ random graph. Observe that

$$\varepsilon^3 m = \frac{(\Delta w_t)^3}{q}(1 + \Delta w_t n^{-1/3}).$$

Since also $B \leq w_t \leq \delta n^{1/3}$ by assumption and we can choose $B$ large enough, it follows from Lemma 2.12 that for any positive constant $\rho \leq 1/10$ there are constants $c_0, c_1 > 0$ such that

$$\Pr[|C_1(t)| \leq (2-\rho)\varepsilon m \mid X_t, \mathcal{E}_t, A_t = m] \leq c_1 \exp(-c_0 w_t^3).$$

By monotonicity,

$$\Pr[|C_1(t)| \leq (2-\rho)\varepsilon m \mid X_t, \mathcal{E}_t, A_t \in J_t] \leq c_1 \exp(-c_0 w_t^3),$$

and since $L_1(X_{t+1}) \geq |C_1(t)|$,

$$\Pr[L_1(X_{t+1}) \leq (2-\rho)\varepsilon m \mid X_t, \mathcal{E}_t, A_t \in J_t] \leq c_1 \exp(-c_0 w_t^3).$$

Now, note that

$$(2-\rho)\varepsilon m = \frac{(2-\rho)\Delta(1+\Delta w_t n^{-1/3})}{q} L_1(X_t),$$

so we let $\alpha := \frac{(2-\rho)\Delta(1+\Delta w_t n^{-1/3})}{q}$. Observe that if $\delta = \rho = \gamma = 0$, then $\alpha = \frac{2(q-1)}{q}$. Since $\frac{2(q-1)}{q} > 1$ for $q > 2$, we can choose $\gamma$, $\delta$ and $\rho$ small enough such that $\alpha > 1$. Thus,

$$\Pr[L_1(X_{t+1}) > \alpha L_1(X_t) \mid X_t, \mathcal{E}_t, A_t \in J_t] \geq 1 - c_1 \exp(-c_0 w_t^3). \tag{3.36}$$

By assumption $Z_t = O(n^{4/3})$, so Hoeffding's inequality implies that for a suitable constant $c_2 > 0$

$$\Pr[A_t \in J_t \mid X_t, \mathcal{E}_t] \geq 1 - 2\exp(-c_2 w_t^2).$$

Together with (3.36) this implies that

$$\Pr[L_1(X_{t+1}) > \alpha L_1(X_t) \mid X_t, \mathcal{E}_t] \geq (1 - 2\exp(-c_2 w_t^2))(1 - c_1 \exp(-c_0 w_t^3)),$$

from which part (i) follows. Part (ii) holds because if the largest component is inactive at time $t$, then at time $t+1$ there is a component of size $L_1(X_t)$. □

*Proof of Claim 3.45.* Let $W_t' := \sum_{j\geq 2} |C_j(t)|^2$, $\mu_t := L_1(X_t) + \frac{n-L_1(X_t)}{q}$ and $J_t := [\mu_t - \gamma w_t n^{2/3}, \mu_t + \gamma w_t n^{2/3}]$ where $\gamma > 0$ is a small constant we choose later. Then,

$$\mathrm{E}[W_t' \mid X_t, \mathcal{E}_t, A_t \in J_t] = \sum_{m \in J_t} \mathrm{E}[W_t' \mid X_t, \mathcal{E}_t, A_t = m] \Pr[A_t = m \mid X_t, \mathcal{E}_t, A_t \in J_t].$$

Let $\varepsilon(m) = mq/n - 1$, $\Delta^+ = q - 1 + \gamma q$ and $\Delta^- = q - 1 - \gamma q$. If $A_t = m$, then the new components are those of a $G(m, \frac{1+\varepsilon(m)}{m})$ random graph. Moreover, for $m \in J_t$ we have

$$\Delta^- w_t n^{-1/3} \leq \varepsilon(m) \leq \Delta^+ w_t n^{-1/3}.$$

Since $\Delta^- > 0$ for $\gamma$ small, $w_t \geq B$ and $B$ is large by assumption, by Lemma 2.14

$$\mathrm{E}[W_t' \mid X_t, \mathcal{E}_t, A_t \in J_t] = \sum_{m \in J_t} O\left(\frac{m}{\varepsilon(m)}\right) \Pr[A_t = m \mid X_t, \mathcal{E}_t, A_t \in J_t] = O\left(\frac{n^{4/3}}{w_t}\right),$$

and by Markov's inequality

$$\Pr\left[W_t' < \frac{Cn^{4/3}}{\sqrt{w_t}} \;\middle|\; X_t, \mathcal{E}_t, A_t \in J_t\right] \geq 1 - \frac{c_0}{\sqrt{w_t}},$$

for some constant $c_0 > 0$. Since $Z_t = O(n^{4/3})$, Hoeffding's inequality implies that, for suitable constant $c_1 > 0$, $\Pr[A_t \in J_t \mid X_t, \mathcal{E}_t] \geq 1 - 2\exp(-c_1 w_t^2)$. Hence,

$$\Pr\left[W_t' < \frac{Cn^{4/3}}{\sqrt{w_t}} \;\middle|\; X_t, \mathcal{E}_t\right] \geq 1 - \frac{c_0}{\sqrt{w_t}} - 2\exp(-c_1 w_t^2).$$

Now, if $C_I$ is the largest inactive component of $X_t$, then

$$Z_{t+1} \leq Z_t - |C_I|^2 + \min\{|C_1(t)|, |C_I|\}^2 + W_t' \leq Z_t + W_t'.$$

Thus,

$$\Pr\left[Z_{t+1} < Z_t + \frac{Cn^{4/3}}{\sqrt{w_t}} \;\middle|\; X_t, \mathcal{E}_t\right] \geq \Pr\left[W_t' < \frac{Cn^{4/3}}{\sqrt{w_t}} \;\middle|\; X_t, \mathcal{E}_t\right] \geq 1 - \frac{c_0}{\sqrt{w_t}} - 2\exp(-c_1 w_t^2),$$

from which part (i) follows.

We derive part (ii) in a similar fashion. Let $\hat{\mu}_t := \frac{n-L_1(X_t)}{q}$, $\hat{J}_t := [\hat{\mu}_t - \gamma w_t n^{2/3}, \hat{\mu}_t + \gamma w_t n^{2/3}]$ and $\hat{m} = \hat{\mu}_t + \gamma w_t n^{2/3}$. By monotonicity,

$$\mathrm{E}[W_t \mid X_t, \neg\mathcal{E}_t, A_t \in \hat{J}_t] \leq \mathrm{E}[W_t \mid X_t, \neg\mathcal{E}_t, A_t = \hat{m}].$$

Let $\hat{\varepsilon} := \hat{m}q/n - 1 = -w_t n^{-1/3}(1 - \gamma q)$. If $A_t = \hat{m}$, then the new components are those of a $G(\hat{m}, \frac{1+\hat{\varepsilon}}{\hat{m}})$ random graph. So for sufficiently small $\gamma$ Lemma 2.13 implies

$$\mathrm{E}[W_t \mid X_t, \neg\mathcal{E}_t, A_t \in \hat{J}_t] = O\left(\frac{\hat{m}}{|\hat{\varepsilon}|}\right) = O\left(\frac{n^{4/3}}{w_t}\right),$$

and by Markov's inequality

$$\Pr\left[W_t < \frac{Cn^{4/3}}{\sqrt{w_t}} \,\middle|\, X_t, \neg\mathcal{E}_t, A_t \in \hat{J}_t\right] \geq 1 - \frac{c_0'}{\sqrt{w_t}},$$

for a suitable positive constant $c_0'$. Since $Z_t = O(n^{4/3})$ by assumption, by Hoeffding's inequality $\Pr[A_t \in \hat{J}_t \mid X_t, \neg\mathcal{E}_t] \geq 1 - 2\exp(-c_1' w_t^2)$ for a suitable constant $c_1' > 0$. Therefore,

$$\Pr\left[W_t < \frac{Cn^{4/3}}{\sqrt{w_t}} \,\middle|\, X_t, \neg\mathcal{E}_t\right] \geq 1 - \frac{c_0'}{\sqrt{w_t}} - 2\exp(-c_1' w_t^2).$$

Part (ii) then follows from the facts that $Z_{t+1} \leq Z_t + W_t$, $w_t \geq B$ and $B$ is large. □

## 3.7 Mixing time lower bounds

In this section we prove the lower bounds in Theorem 3.11 for the mixing time of the CM dynamics. We point out that for $\lambda < \lambda_s$ the SW dynamics is known to mix in $\Theta(1)$ steps [38, 21], while the CM dynamics requires $\Theta(\log n)$ steps to mix. This is due to the fact that the CM dynamics may require as many steps to activate all the components.

**Theorem 3.46.** *If $q > 2$ and $\lambda \in (\lambda_s, \lambda_S)$, then the mixing time of the CM dynamics is* $\exp(\Omega(\sqrt{n}))$.

*Proof.* The random-cluster model undergoes a phase transition at $\lambda_c$, so it is natural to divide the proof into two cases: $\lambda \in [\lambda_c, \lambda_S)$ and $\lambda \in (\lambda_s, \lambda_c)$.

**Case (i):** $\lambda_c \leq \lambda < \lambda_S = q$. The idea for this bound comes from [25]. Let $S$ be the set of graphs $G$ such that $L_1(G) = \Theta(\sqrt{n})$ and let $X_0 \in S$. Let $\mu := \mathrm{E}[A_0] = n/q$; then by Hoeffding's inequality $\Pr[|A_0 - \mu| > \varepsilon n] \leq 2\exp\left(-2\varepsilon^2\sqrt{n}\right)$. If $A_0 < \mu + \varepsilon n$, the active subgraph is sub-critical for sufficiently small $\varepsilon$. Therefore, Lemma 2.10 implies that $\Pr[X_1 \notin S | X_0 \in S] \leq e^{-c\sqrt{n}}$ for some constant $c > 0$. Hence, $\Pr[X_1, ..., X_t \in S | X_0 \in S] \geq 1 - te^{-c\sqrt{n}} \geq 3/4$ for $t = \lfloor e^{c\sqrt{n}}/4 \rfloor$. The result again follows from Lemma 3.1.

**Case (ii):** $\lambda_s < \lambda < \lambda_c$. The intuition for this case comes directly from Figure 3.1. In this regime, Fact 3.6 implies that the function $f(\theta) = \theta - \phi(\theta)$ has two positive zeros $\theta^*$ and $\theta_r$ in $(\theta_{\min}, 1]$ with $\theta^* < \theta_r$. Moreover, $f$ is negative in the interval $(\theta^*, \theta_r)$. Therefore, any configuration with a unique large component of size $\theta n$ with $\theta \in (\theta^*, \theta_r)$ will "drift" towards a configuration with a bigger large component. However, a typical random-cluster configuration in this regime does not have a large component. This drift in the incorrect direction is sufficient to prove the exponential lower bound in this regime. We now proceed to formalize this intuition.

Let $S$ be the set of graphs $G$ such that $L_1(G) > (\theta^* + \varepsilon)n$ and $L_2(G) = O(\sqrt{n})$ where $\varepsilon$ is a small positive constant to be chosen later. Assume $X_0 \in S$. If $\mathcal{L}(X_0)$ is inactive, by Hoeffding's inequality $A_0 \in I_0 := [(1-\theta_0)n/q - \gamma_0 n, (1-\theta_0)n/q + \gamma_0 n]$ with probability $1 - e^{-\Omega(\sqrt{n})}$ for any desired constant $\gamma_0 > 0$. If $A_0 \in I_0$, then, for a sufficiently small $\gamma_0$, the percolation step is sub-critical, and by Lemma 2.10, $\Pr[X_1 \notin S | X_0 \in S] = e^{-\Omega(\sqrt{n})}$.

When $\mathcal{L}(X_0)$ is active, we show that for any desired constant $\rho > 0$, $L_1(X_1) \in [\phi(\theta_0)n - \rho n, \phi(\theta_0)n + \rho n]$ with probability $1 - e^{-\Omega(\sqrt{n})}$. Let $\mu_0 := \theta_0 n + (1-\theta_0)n/q$; then by Hoeffding's inequality $A_0 \in I_1 := [\mu_0 - \gamma_1 n, \mu_0 + \gamma_1 n]$ with probability $1 - e^{-\Omega(\sqrt{n})}$ for any desired constant $\gamma_1 > 0$. Let $h(\theta_0) = \mu_0 n + \gamma_1 n$ and let $\ell(\theta_0)$ be a random variable distributed as the size of the largest component of a $G(h(\theta_0), p)$ random graph. Then, for any $\rho > 0$,

$$\begin{aligned}\Pr[L_1(X_1) > \phi(\theta_0)n + \rho n] &\leq \sum_{a \in I_1} \Pr[L_1(X_1) > \phi(\theta_0)n + \rho n | A_0 = a] \Pr[A_0 = a] + e^{-\Omega(\sqrt{n})} \\ &\leq \Pr[\,\ell(\theta_0) > \phi(\theta_0)n + \rho n\,] + e^{-\Omega(\sqrt{n})}.\end{aligned}$$

Recall from Section 3.1.2 that when $\lambda < q$, $\lambda(\theta_{\min} + (1-\theta_{\min})q^{-1}) = 1$. Therefore, the $G(h(\theta_0), p)$ random graph is super-critical since $\theta_0 > \theta^* > \theta_{\min}$. Let $\beta = \beta(\lambda')$ with $\lambda' = \lambda h(\theta_0)/n$ where $\beta(\lambda')$ is defined in (2.2). By Lemma 2.9, $\ell(\theta_0) \in [\beta n - \gamma_2 n, \beta n + \gamma_2 n]$ with probability at least $1 - e^{-\Omega(n)}$ for any desired constant $\gamma_2 > 0$. Observe that if $\gamma_1 = 0$, then $\beta = \phi(\theta_0)$ by the definition of $\phi$. Then by continuity, for any constant $\delta > 0$ there exists $\gamma_1$ small enough such that $|\phi(\theta_0) - \beta| < \delta$. Thus, $\ell(\theta_0) \in [\phi(\theta_0)n - \rho n, \phi(\theta_0)n + \rho n]$ with probability $1 - e^{-\Omega(n)}$. Consequently, $\Pr[L_1(X_1) > \phi(\theta_0)n + \rho n] = e^{-\Omega(\sqrt{n})}$. By a similar argument $\Pr[L_1(X_1) < \phi(\theta_0)n - \rho n] = e^{-\Omega(\sqrt{n})}$, and then $L_1(X_1) \in [\phi(\theta_0)n - \rho n, \phi(\theta_0)n + \rho n]$ with probability $1 - e^{-\Omega(\sqrt{n})}$.

Now we show that for suitable positive constants $\varepsilon$ and $\rho$, $\phi(\theta_0) - \rho > \theta^* + \varepsilon$; this implies $L_1(X_1) > (\theta^* + \varepsilon)n$ with probability $1 - e^{-\Omega(\sqrt{n})}$. Note that Lemma 3.9(i) still holds when $\lambda_c > \lambda > \lambda_s$ and $\theta \in (\theta^*, 1)$. Hence, if $\theta_0 > \theta_r - \varepsilon$, then $\phi(\theta_0) > \theta_r - \varepsilon$. Therefore, we can choose $\varepsilon$ and $\rho$ such that $\phi(\theta_0) - \rho > \theta^* + \varepsilon$. If $\theta_0 < \theta_r - \varepsilon$, then $\phi(\theta_0) > \theta_0 > \theta^* + \varepsilon$ since $f$ is negative in this interval. Note that $\phi(\theta_0) - \theta_0 = -f(\theta_0)$, so in this case we can pick $\rho$ to be $-1/2$ of the maximum of $f$ in $[\theta^* + \varepsilon, \theta_r - \varepsilon]$ for a sufficiently small $\varepsilon$. Thus, $L_1(X_1) > (\theta^* + \varepsilon)n$ with probability $1 - e^{-\Omega(\sqrt{n})}$.

By Lemma 2.10, $L_2(X_1) = O(\sqrt{n})$ with probability $1 - e^{-\Omega(\sqrt{n})}$. Hence, $\Pr[X_1 \notin S | X_0 \in S] \leq e^{-c\sqrt{n}}$ for some constant $c > 0$, and then $\Pr[X_1, \ldots, X_t \in S | X_0 \in S] \geq 1 - te^{-c\sqrt{n}} \geq 3/4$ for $t = \lfloor e^{c\sqrt{n}}/4 \rfloor$. The result then follows from Lemma 3.1. $\square$

We prove next a uniform lower bound for the fast mixing regime.

**Theorem 3.47.** *For $q > 1$ and $\lambda \neq \lambda_c$ the mixing time of the CM dynamics is $\Omega(\log n)$.*

*Proof.* Let $X_0$ be a configuration where all the components have size $\Theta(\log^2 n)$ and let $b = q/(q-1)$. The probability that a particular component is not activated in any of the first $T = \lceil \frac{1}{2} \log_b n \rceil$ steps is $(1 - 1/q)^T \leq n^{-1/2}$. Therefore, the probability that all initial components are activated in the first $T$ steps is at least $(1 - n^{-1/2})^K$, where $K = \Theta(n/\log^2 n)$. Moreover, the probability that exactly one initial component is not activated in the first $T$ steps is $\Omega(Kn^{-1/2}(1 - n^{-1/2})^{K-1})$. Hence, after $T$ steps $L_2(X_T) = \Theta(\log^2 n)$ w.h.p., and the result follows from Lemma 3.1. $\square$

We conclude this section by proving a $\Omega(n^{1/3})$ lower bound for the mixing time for $q > 2$ and $\lambda = \lambda_s$.

**Theorem 3.48.** *For $q > 2$ and $\lambda = \lambda_s$ the mixing time of the CM dynamics is $\Omega(n^{1/3})$.*

*Proof.* Let $X_0$ be a random-cluster configuration such that $L_1(X_0) = \theta^* n + n^{3/4}$, $L_2(X_0) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_0)^2 = O(n)$. By Fact 3.39(ii) these properties are maintained for $T$ steps w.h.p., provided $T = O(n^{1/3})$ and $L_1(X_t) \geq (\Theta_s + \varepsilon)n$ for all $t \leq T$, where $\varepsilon > 0$ is the constant from Fact 3.39.

Suppose $X_t$ is such that $\theta^* n - n^{3/4} \leq L_1(X_t) \leq \theta^* n + n^{3/4}$, $L_2(X_t) = O(\log n)$ and $\sum_{j\geq 2} L_j(X_t)^2 = O(n)$. Then, when $\mathcal{L}(X_t)$ is inactive $L_1(X_{t+1}) = L_1(X_t)$ with probability $1 - O(n^{-1})$ by Fact 3.17(i). If $\mathcal{L}(X_t)$ is active, then $L_1(X_{t+1}) \in [\phi(\theta_t)n - A\sqrt{n}\log n, \phi(\theta_t)n + A\sqrt{n}\log n]$ with probability $1 - O(n^{-2})$ by Fact 3.40(iii). Since $\theta^* n - n^{3/4} \leq L_1(X_t) \leq \theta^* n + n^{3/4}$, Lemma 3.8(i) implies that there exist constants $\delta, \delta' > 0$ such that

$$\theta_t - \delta(\theta_t - \theta^*)^2 \leq \phi(\theta_t) \leq \theta_t - \delta'(\theta_t - \theta^*)^2,$$

and thus

$$\theta_t n - \delta\sqrt{n} \leq \phi(\theta_t)n \leq \theta_t n.$$

Therefore, with probability $1 - O(n^{-2})$,

$$L_1(X_{t+1}) \in [\theta_t n - (A+1)\sqrt{n}\log n, \theta_t n + (A+1)\sqrt{n}\log n]$$

Inductively, we get that if $T = O(n^{1/3})$, then $L_1(X_T) \geq L_1(X_0) - T(A+1)\sqrt{n}\log n = \Omega(n)$ w.h.p., and thus the result follows from Lemma 3.1. ◻

## 3.8 Local dynamics

In this section we prove Theorem 1.3 from the Introduction.

### 3.8.1 Standard background

Let $P$ be the transition matrix of a finite, ergodic and reversible Markov chain over state space $\Omega$ with stationary distribution $\pi$, and let $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ denote the eigenvalues of $P$. The *spectral gap* of $P$ is defined by $\lambda(P) := 1 - \lambda^*$, where $\lambda^* = \max\{|\lambda_2|, |\lambda_n|\}$. The following bounds on the mixing time are standard (see, e.g., [37]):

$$\lambda^{-1}(P) - 1 \leq \tau_{\text{mix}}(P) \leq \log\left(2e\pi_{\min}^{-1}\right)\lambda^{-1}(P), \tag{3.37}$$

where $\pi_{\min} = \min_{x\in\Omega} \pi(x)$.

In this section we will need some elementary notions from functional analysis; for extensive background on the application of such ideas to the analysis of finite Markov chains, see [46]. If we endow $\mathbb{R}^{|\Omega|}$ with the inner product $\langle f, g\rangle_\pi = \sum_{x\in\Omega} f(x)g(x)\pi(x)$, we obtain a Hilbert space denoted $L_2(\pi) = (\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle_\pi)$. Note that $P$ defines an operator from $L_2(\pi)$ to $L_2(\pi)$ via matrix-vector multiplication.

Consider two Hilbert spaces $S_1$ and $S_2$ with inner products $\langle\cdot,\cdot\rangle_{S_1}$ and $\langle\cdot,\cdot\rangle_{S_2}$ respectively, and let $R : S_2 \to S_1$ be a bounded linear operator. The *adjoint* of $R$ is the unique operator $R^* : S_1 \to S_2$ satisfying $\langle f, Rg\rangle_{S_1} = \langle R^* f, g\rangle_{S_2}$ for all $f \in S_1$ and $g \in S_2$. If $S_1 = S_2$, $R$ is *self-adjoint* when $R = R^*$. If $R$ is self-adjoint, it is also *positive* if $\forall g \in S_2$, $\langle Rg, g\rangle_{S_2} \geq 0$.

### 3.8.2 A comparison technique for Markov chains

Let $H = (V, E)$ be an arbitrary finite graph and let $\Omega_{\rm E} = \{(V, A) : A \subseteq E\}$ be the set of random-cluster configurations on $H$. Let $P$ be the transition matrix of a finite, ergodic and reversible Markov chain over $\Omega_{\rm E}$ with stationary distribution $\mu = \mu_{p,q}$. For $r \in \mathbb{N}$, let $\Omega_{\rm V} = \{0, 1, \dots, r-1\}^V$ be the set of "$r$-labelings" of $V$, and let $\Omega_{\rm J} = \Omega_{\rm V} \times \Omega_{\rm E}$. Assume $P$ can be decomposed as a product of *stochastic* matrices of the form

$$P = M\left(\prod_{i=1}^{m} T_i\right)M^*, \tag{3.38}$$

where:

(i) $M$ is a $|\Omega_{\rm E}| \times |\Omega_{\rm J}|$ matrix indexed by the elements of $\Omega_{\rm E}$ and $\Omega_{\rm J}$ where $M(A, (\sigma, B)) \neq 0$ only if $A = B$ for all $A \in \Omega_{\rm E}$, $(\sigma, B) \in \Omega_{\rm J}$.

(ii) Each $T_i$ is a $|\Omega_{\rm J}| \times |\Omega_{\rm J}|$ matrix indexed by the elements of $\Omega_{\rm J}$ and reversible w.r.t. the distribution $\nu = \mu M$, and such that $T_i((\sigma, A), (\tau, B)) \neq 0$ only if $\sigma = \tau$ for all $(\sigma, A), (\tau, B) \in \Omega_{\rm J}$.

(iii) $M^*$ is a $|\Omega_{\rm J}| \times |\Omega_{\rm E}|$ matrix such that $M^* : L_2(\mu) \to L_2(\nu)$ is the adjoint of $M : L_2(\nu) \to L_2(\mu)$.

In words, $M$ assigns a (random) $r$-labeling to the vertices of $H$; $(\prod_{i=1}^{m} T_i)$ performs a sequence of $m$ operations $T_i$, each of which updates some edges of $H$; and $M^*$ drops the labels from the vertices. These properties imply that $M^*((\sigma, A), B) = \mathbb{1}(A = B)$ and $MM^* = I$.

Consider now the matrix

$$P_{\rm L} = M\left(\frac{1}{m}\sum_{i=1}^{m} T_i\right)M^*. \tag{3.39}$$

It is straightforward to verify that $P_{\rm L}$ is also reversible w.r.t. $\mu$. The following theorem, which generalizes a recent result of Ullrich [50, 51], relates the spectral gaps of $P$ and $P_{\rm L}$ up to a factor of $O(m \log m)$.

**Theorem 3.49.** *If $M$, $M^*$ and $T_i$ are stochastic matrices satisfying (i)–(iii) above, and the $T_i$'s are idempotent commuting operators, then*

$$\lambda(P_L) \leq \lambda(P) \leq 8m \log m \cdot \lambda(P_L).$$

We pause to note that this fact has a very attractive intuitive basis. As noted above, $P_{\rm L}$ performs a single update $T_i$ chosen u.a.r., while $P$ performs all $m$ updates $T_i$, so by coupon collecting one might expect that $O(m \log m)$ $P_{\rm L}$ steps should suffice to simulate a single $P$ step. However, the proof has to take account of the fact that the $T_i$ updates are interleaved with the vertex re-labeling operations $M$ and $M^*$ in $P_{\rm L}$. The proofs in [50] and [51] are specific to the case where $P$ corresponds to the SW dynamics. Our contribution is the realization that these proofs still go through (without essential modification) under the more general assumptions of Theorem 3.49, as well as the framework described above that provides a systematic way of deducing $P_{\rm L}$ from any $P$ of the form (3.38).

Observe that Theorem 3.49 relates the spectral gaps of $P$ and $P_{\mathrm{L}}$. We shall see next how to use this technology to obtain mixing time bounds for the heat-bath dynamics using the CM bounds from Sections 3.2 and 3.7.

### 3.8.3 Application to local dynamics

Let $P_{\mathrm{CM}}$ and $P_{\mathrm{HB}}$ be the transition matrices of the Chayes-Machta (CM) and heat-bath (HB) dynamics respectively. In this subsection we show that $P_{\mathrm{CM}}$ can be expressed as a product of stochastic matrices equivalent to (3.38) and that $P_{\mathrm{HB}}$ is closely related to the corresponding matrix $P_{\mathrm{L}}$ in (3.39). Then, we use Theorem 3.49 to relate the spectral gaps $\lambda(P_{\mathrm{HB}})$ and $\lambda(P_{\mathrm{CM}})$ and hence prove Theorem 1.3 via (3.37).

In this case, $\Omega_{\mathrm{V}} = \{0,1\}^V$ is the set of possible "active-inactive" labelings of $V$. Consider the $|\Omega_{\mathrm{E}}| \times |\Omega_{\mathrm{J}}|$ stochastic matrix $M$ defined by

$$M(B,(\sigma,A)) = \mathbb{1}(A = B)\, \mathbb{1}(A \subseteq E(\sigma))(q-1)^{f(\sigma,A)} q^{-c(A)},$$

where $E(\sigma) = \{(u,v) \in E : \sigma(u) = \sigma(v)\}$ and $f(\sigma,A)$ is the number of inactive connected components in $(\sigma,A)$. The adjoint of $M$ is the $|\Omega_{\mathrm{J}}| \times |\Omega_{\mathrm{E}}|$ stochastic matrix $M^*((\sigma,A),B) = \mathbb{1}(A = B)$. Consider also the family of $|\Omega_{\mathrm{J}}| \times |\Omega_{\mathrm{J}}|$ stochastic matrices $T_e$ defined for each $e = (u,v) \in E$ as follows:

$$T_e((\sigma,A),(\tau,B)) = \mathbb{1}(\sigma = \tau)\begin{cases} p & \text{if } B = A \cup e, \quad \sigma(u) = \sigma(v) = 1; \\ 1-p & \text{if } B = A \setminus e, \quad \sigma(u) = \sigma(v) = 1; \\ 1 & \text{if } A(e) = B(e), \quad \sigma(u) = 0 \ \text{ or } \ \sigma(v) = 0; \\ 0 & \text{if } A(e) \neq B(e), \quad \sigma(u) = 0 \ \text{ or } \ \sigma(v) = 0 \end{cases}$$

where $\sigma(v) = 1$ (resp., 0) if vertex $v$ is active (resp., inactive) in $\sigma$ and $A(e) = 1$ (resp., $A(e) = 0$) if the edge $e$ is present (resp., not present) in $A$.

In words, the matrix $M$ assigns a random active-inactive labeling to a random-cluster configuration, while $M^*$ drops the active-inactive labeling from a joint configuration. The matrix $T_e$ samples $e$ with probability $p$ provided both its endpoints are active. The key observation, which we prove later, is that we can naturally express the CM dynamics as the product of these matrices:

**Lemma 3.50.** *$P_{CM} = M\left(\prod_{e\in E} T_e\right) M^*$.*

Now consider the Markov chain given by the matrix

$$P_{\mathrm{SU}} = M\left(\frac{1}{|E|}\sum_{e\in E} T_e\right) M^*,$$

which we call the *Single Update (SU) dynamics* and corresponds to the matrix $P_{\mathrm{L}}$ defined in (3.39). Hence, $P_{\mathrm{SU}}$ is reversible w.r.t. to $\mu = \mu_{p,q}$. Observe that $M$ and $M^*$ clearly satisfy the assumptions of Theorem 3.49. Moreover, we can easily verify that the $T_e$'s also satisfy these assumptions:

**Fact 3.51.** *The $T_e$'s defined above are idempotent commuting operators from $L_2(\nu)$ to $L_2(\nu)$. Moreover, each $T_e$ is reversible w.r.t. $\nu = \mu M$.*

*Proof.* The distribution $\nu$ corresponds to the joint Edwards-Sokal measure over $\Omega_{\mathrm{J}}$:

$$\nu(\sigma, A) \propto \left(\frac{p}{1-p}\right)^{|A|} (q-1)^{f(\sigma,A)}\, \mathbb{1}(A \subseteq E(\sigma))$$

(see, e.g., [10]). From this representation, it is straightforward to check that $T_e$ is reversible w.r.t. to $\nu$. Also, from the definition of $T_e$ it follows that $T_e = T_e^2$ and $T_e T_{e'} = T_{e'} T_e$, which completes the proof. □

In light of Lemma 3.50 and Fact 3.51, we may apply Theorem 3.49 to obtain

$$\lambda(P_{\mathrm{SU}}) \leq \lambda(P_{\mathrm{CM}}) \leq 8|E|\log|E| \cdot \lambda(P_{\mathrm{SU}}). \tag{3.40}$$

The SU dynamics is closely related to the HB dynamics. Specifically, their spectral gaps are very similar, as the following fact which we will prove in a moment shows:

**Claim 3.52.** *Let $\alpha = (q(1-p)+p)/q^2$; then,*

$$\alpha\lambda(P_{HB}) \leq \lambda(P_{SU}) \leq \lambda(P_{HB}).$$

Putting together this claim and (3.40) yields

$$\alpha\lambda(P_{\mathrm{HB}}) \leq \lambda(P_{\mathrm{CM}}) \leq 8|E|\log|E| \cdot \lambda(P_{\mathrm{HB}}),$$

which relates the spectral gaps of $P_{\mathrm{HB}}$ and $P_{\mathrm{CM}}$ up to a factor of $\tilde{O}(n^2)$. (Note that $\alpha \in [1/q^2, 1/q]$, and thus $\alpha = \Theta(1)$.) Using (3.37) this relationship can be translated to the mixing times at the cost of a further factor of $\log(\mu_{\min}^{-1})$, which is $\tilde{O}(n^2)$ in the mean-field case. Theorem 1.3 now follows immediately from the mixing time bounds on the CM dynamics proved in Theorem 3.11.

**Remark.** In the final version of [5], which is currently in preparation, we are able to translate the relation between the spectral gaps of $P_{\mathrm{HB}}$ and $P_{\mathrm{CM}}$ to the mixing times losing only an $\tilde{O}(n)$ factor. This strengthens all the bounds in Theorem 1.2 from the Introduction by an $O(n)$ factor.

It remains only for us to supply the missing proofs of Lemma 3.50 and Claim 3.52.

*Proof of Lemma 3.50.* Let $\mathcal{A}(\sigma) = \{(u,v) \in E : \sigma(u) = \sigma(v) = 1\}$ and let $\mathcal{T} = \prod_{e \in E} T_e$. Observe that

$$\mathcal{T}((\sigma, A), (\tau, B)) = \mathbb{1}(\sigma = \tau)\, \mathbb{1}(A \setminus \mathcal{A}(\sigma) = B \setminus \mathcal{A}(\sigma)) p^{|\mathcal{A}(\sigma)\cap B|}(1-p)^{|\mathcal{A}(\sigma)| - |\mathcal{A}(\sigma)\cap B|}.$$

Then, from the definitions of $M$, $M^*$ and $\mathcal{T}$, we obtain

$$\begin{aligned} M\mathcal{T}M^*(A,B) &= \sum_{(\sigma,C)} \sum_{(\tau,D)} M(A,(\sigma,C))\mathcal{T}((\sigma,C),(\tau,D))M^*((\tau,D),B) \\ &= \sum_{\sigma\in\Omega_{\mathrm{V}}} \mathbb{1}(A \subseteq E(\sigma)) \frac{(q-1)^{f(\sigma,A)}}{q^{c(A)}}\, \mathbb{1}(A \setminus \mathcal{A}(\sigma) = B \setminus \mathcal{A}(\sigma)) p^{|\mathcal{A}(\sigma)\cap B|}(1-p)^{|\mathcal{A}(\sigma)| - |\mathcal{A}(\sigma)\cap B|}. \end{aligned}$$

Now observe that if $A \subseteq E(\sigma)$, then sub-step (i) of the CM dynamics chooses $\sigma \in \Omega_V$ with probability $(q-1)^{f(\sigma,A)} q^{-c(A)}$. Moreover, if after sub-step (i) the joint configuration obtained is $(\sigma, A)$, then the probability of obtaining $B$ in sub-step (iii) is $p^{|\mathcal{A}(\sigma)\cap B|}(1-p)^{|\mathcal{A}(\sigma)|-|\mathcal{A}(\sigma)\cap B|}$ provided $A$ and $B$ differ only in the active part of the configuration $\mathcal{A}(\sigma)$. Thus, $M\mathcal{T}M^*(A,B) = P_{\mathrm{CM}}(A,B)$. ◻

*Proof of Claim 3.52.* First we show that $P_{\mathrm{SU}}$ is positive. This was already shown for $P_{\mathrm{HB}}$ in [50, Lemma 2.7]. Recall each $T_e$ is reversible w.r.t. to $\nu$, and thus the operator $T_e : L_2(\nu) \to L_2(\nu)$ is self-adjoint (see, e.g., [46]). Since $T_e$ is self-adjoint and idempotent, it is also positive for every $e \in E$ (see, e.g. [34, Thm. 9.5-1 & 9.5-2]). Therefore, for $f \in \mathbb{R}^{|\Omega_{\mathrm{E}}|}$ we have

$$\langle P_{\mathrm{SU}} f, f\rangle_\mu = \frac{1}{|E|}\sum_{e\in E}\langle MT_eM^* f, f\rangle_\mu = \frac{1}{|E|}\sum_{e\in E}\langle T_eM^* f, M^* f\rangle_\nu \geq 0,$$

and thus $P_{\mathrm{SU}}$ is positive.

Given a random-cluster configuration $(V, A)$, it is straightforward to check that one step of the SU dynamics is equivalent to the following discrete steps:

(i) *activate* each connected component of $(V, A)$ independently with probability $1/q$;

(ii) pick $e \in E$ u.a.r.;

(iii) if both endpoints of $e$ are active, add $e$ with probability $p$ and remove it otherwise. (If either endpoint of $e$ is inactive, do nothing.)

Similarly, recalling the definition of the HB dynamics from the Introduction, it is easy to check that each step is equivalent to the following:

(i) pick an edge $e \in E$ u.a.r.;

(ii) include the edge $e$ in the new configuration with probability $p_e$, where

$$p_e = \begin{cases} \frac{p}{p+q(1-p)} & \text{if } e \text{ is a cut edge in } (V, A\cup\{e\}); \\ p & \text{otherwise.} \end{cases}$$

(The rest of the configuration is left unchanged.)

Note that $e$ is a cut edge in $(V, A \cup \{e\})$ iff changing the current configuration of $e$ changes the number of connected components.

Using these definitions for the SU and HB dynamics, it is an easy exercise to check that for $A \neq B$ and $\alpha = (q(1-p)+p)/q^2$,

$$\alpha P_{\mathrm{HB}}(A,B) \leq P_{\mathrm{SU}}(A,B) \leq P_{\mathrm{HB}}(A,B).$$

Since $P_{\mathrm{SU}}$ is positive, the result follows from Lemma 2.5 in [50]. ◻

# Chapter 4

# Random-Cluster Dynamics in $\mathbb{Z}^2$

## 4.1 Preliminaries

In this section we gather a number of definitions and background results that we will refer to repeatedly. More details and proofs can be found in the books [26, 37].

**Random-cluster model on $\mathbb{Z}^2$.** Let $\mathbb{L} = (\mathbb{Z}^2, \mathbb{E})$ be the square lattice graph, where for $u, v \in \mathbb{Z}^2$, $(u, v) \in \mathbb{E}$ iff $d(u, v) = 1$ with $d(\cdot, \cdot)$ denoting the Euclidean distance. Let $\Lambda_n \subseteq \mathbb{Z}^2$ be the set of vertices of $\mathbb{L}$ contained in a square box of side length $n$, and let $\Lambda = (\Lambda_n, E_n)$ be the graph whose edge set $E_n$ contains all edges in $\mathbb{E}$ with both endpoints in $\Lambda_n$. We use $\partial\Lambda$ to denote the *boundary* of $\Lambda$; that is, the set of vertices in $\Lambda_n$ connected by an edge in $\mathbb{E}$ to $\Lambda_n^c = \mathbb{Z}^2 \setminus \Lambda_n$.

A *random-cluster configuration* on $\Lambda$ corresponds to a subset $A$ of $E_n$. Alternatively, it is sometimes convenient to think of $A$ as a vector in $\{0, 1\}^{|E_n|}$ indexed by the edges, where $A(e) = 1$ iff $e \in A$. Edges belonging to $A$ are called *open*, and edges in $E_n \setminus A$ *closed*.

For any random-cluster configuration $A^c$ on $\Lambda_n^c$, we may consider the conditional random-cluster measure induced in $\Lambda_n$ by $A^c$. To make this precise, we introduce the standard concept of *boundary conditions*. A boundary condition for $\Lambda$ is a partition $\eta = (P_1, P_2, \dots, P_k)$ of $\partial\Lambda$ which encodes how the vertices of $\partial\Lambda$ are connected in a fixed configuration $A^c$ on $\Lambda_n^c$; i.e., for all $u, v \in \partial\Lambda$, $u, v \in P_i$ iff $u$ and $v$ are connected by a path in $A^c$ (see Figure 4.1(a)). In this case we also say that $u$ and $v$ are *wired* in $\eta$.

For $A \subseteq E_n$ and a boundary condition $\eta$, let $c(A, \eta)$ be the number of connected components of $(\Lambda_n, A)$ when the connectivities from the boundary condition $\eta$ are also considered. More precisely, if $C_1, C_2$ are connected components of $A$, and there exist $u \in C_1 \cap \partial\Lambda$ and $v \in C_2 \cap \partial\Lambda$ such that $u$ and $v$ are wired in $\eta$, then $C_1$ and $C_2$ are identified as the same connected component in $A$. The random-cluster measure on $\Lambda$ with boundary condition $\eta$ and parameters $p \in (0, 1)$ and $q > 0$ is then given by

$$\mu^{\eta}_{\Lambda,p,q}(A) = \frac{p^{|A|}(1-p)^{|E_n \setminus A|}q^{c(A,\eta)}}{Z^{\eta}_{\Lambda,p,q}}, \tag{4.1}$$

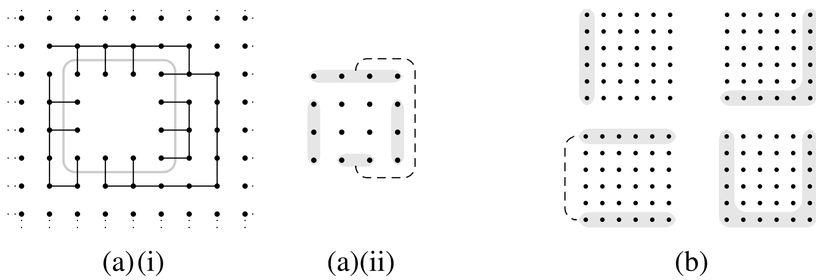


Figure 4.1: *(a) (i) $\Lambda_4 \subset \mathbb{Z}^2$ with a random-cluster configuration $A^c$ in $\Lambda_4^c$, (ii) the boundary condition induced in $\Lambda_4$ by $A^c$; (b) examples of side-homogeneous boundary conditions.*

where $Z^\eta_{\Lambda,p,q}$ is the normalizing constant, or *partition function*. (Cf. equation (1.1) in the Introduction, which corresponds to the special case when the boundary condition $\eta$ is "free"; see below.) When $\Lambda$, $p$ and $q$ are clear from the context we will just write $\mu^\eta$.

**Free, wired and side-homogeneous boundary conditions.** Some boundary conditions will be of particular interest to us. In the *free* boundary condition no two vertices of $\partial\Lambda$ are wired. At the other extreme, in the *wired* boundary condition all vertices of $\partial\Lambda$ are pairwise wired. We will use $\mu^0_{\Lambda,p,q}$ and $\mu^1_{\Lambda,p,q}$ to denote the random-cluster measures on $\Lambda$ with free and wired boundary conditions, respectively.

We consider another class of boundary conditions which we call *side-homogeneous*. Let $L_1, L_2, L_3, L_4 \subset \partial\Lambda$ be the sets of vertices on each *side* of the square box $\Lambda_n$. (A corner vertex of $\Lambda_n$ belongs to two sides.) The class of side-homogeneous boundary conditions contains all $\eta = (P_1, \ldots, P_k)$ satisfying:

(P1) $|P_i| > 1$ for at most one $i$; and

(P2) If $|P_i| > 1$, then $P_i$ is the union of some of the sets $L_j$; i.e., $P_i = \bigcup_{j\in\kappa} L_j$, for some $\kappa \subseteq \{1, 2, 3, 4\}$.

(See Figure 4.1(b).) Note that both the free and wired boundary conditions are side-homogeneous and there are in total 16 distinct side-homogeneous boundary conditions.

**Monotonicity.** For any pair of boundary conditions $\eta$ and $\psi$, we say $\eta \le \psi$ if the partition $\eta$ is a refinement of $\psi$; i.e., if the connectivities induced by $\eta$ in $\partial\Lambda$ are also induced by $\psi$. When $q \ge 1$, $\eta \le \psi$ implies $\mu^\eta_{\Lambda,p,q} \preceq \mu^\psi_{\Lambda,p,q}$, where $\preceq$ denotes *stochastic domination*; i.e., $\mu^\eta_{\Lambda,p,q}(\mathcal{E}) \le \mu^\psi_{\Lambda,p,q}(\mathcal{E})$ for all increasing events $\mathcal{E}$. (An event $\mathcal{E}$ is *increasing* if it is preserved by the addition of edges.)

**Planar duality.** Let $\Lambda^* = (\Lambda^*_n, E^*_n)$ denote the planar dual of $\Lambda$ in the usual sense. That is, $\Lambda^*_n$ corresponds to the set of faces of $\Lambda$, and for each $e \in E_n$, there is a dual edge $e^* \in E^*_n$ connecting the two faces bordering $e$. The random-cluster measure satisfies $\mu_{\Lambda,p,q}(A) = \mu_{\Lambda^*,p^*,q}(A^*)$, where $A^*$ is the dual configuration to $A$ (i.e., $e^* \in E_n^*$ iff $e \in E_n$), and

$$p^* = \frac{q(1-p)}{p + q(1-p)}.$$

(This duality relation is a consequence of Euler's formula.) The unique value of $p$ satisfying $p = p^*$, denoted $p_{sd}(q)$, is called the *self-dual point*.

**Infinite measure and phase transition.** The random-cluster measure may also be defined on the infinite lattice $\mathbb{Z}^2$ by considering the sequence of random-cluster measures on $\Lambda_n$ with free boundary conditions as $n \to \infty$. This sequence converges to a limiting measure $\mu_{\mathbb{L},p,q}$, which is known as the *random-cluster measure on* $\mathbb{L}$. The measure $\mu_{\mathbb{L},p,q}$ exhibits a phase transition corresponding to the appearance of an infinite connected component. That is, there exists a critical value $p = p_c(q)$ such that if $p < p_c(q)$ (resp., $p > p_c(q)$), then all components are finite (resp., there is at least one infinite component) with high probability.

For $q \geq 1$, the exact value of $p_c(q)$ for $\mathbb{L}$ was only recently settled in breakthrough work by Beffara and Duminil-Copin [3], who proved the long standing conjecture

$$p_c(q) = p_{sd}(q) = \frac{\sqrt{q}}{\sqrt{q}+1}.$$

**Exponential decay of connectivies and spatial mixing.** In [3], it was also established that the phase transition is very sharp, meaning that as soon as $p < p_c(q)$ there is exponential decay of connectivities. More formally, for $q \geq 1$ and any fixed $p < p_c(q)$, there exist positive constants $C, \lambda$ such that for all $u, v \in \mathbb{Z}^2$,

$$\mu_{\mathbb{L},p,q}(u \leftrightarrow v) \leq C\mathrm{e}^{-\lambda\, d(u,v)},$$

where $u \leftrightarrow v$ denotes the event that $u$ and $v$ are connected by a path of open edges. In work predating [3], Alexander [1] showed that exponential decay of connectivities implies exponential decay of *finite volume* connectivities uniformly over all boundary conditions. That is, for any boundary condition $\eta$ on $\Lambda$, and all $u, v \in \Lambda_n$,

$$\mu^{\eta}_{\Lambda,p,q}(u \overset{\Lambda_n}{\leftrightarrow} v) \leq C\mathrm{e}^{-\lambda\, d(u,v)}, \tag{4.2}$$

where $u \overset{\Lambda_n}{\leftrightarrow} v$ is the event that $u$ and $v$ are connected by a path of open edges in $\Lambda_n$.

The notion of decay of connectivities for the random-cluster model is analogous to the notion of decay of correlations in spin systems, which is ubiquitous in the spin systems literature. As in spin systems, we require in our analysis of the dynamics a stronger form of decay of connectivities known as *spatial mixing*.

For $e \in E_n$, let $B(e,r) \subset \Lambda_n$ be the set of vertices in the minimal square box around $e$ such that $d(\{e\}, v) \geq r$ for all $v \in \Lambda_n \setminus B(e,r)$. Note that if $d(\{e\}, \partial\Lambda) > r$, then $B(e,r)$ is just a square

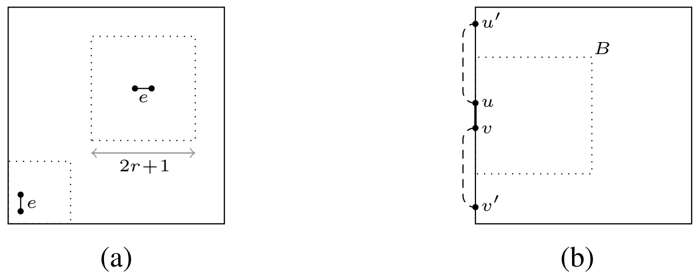


Figure 4.2: *(a) $B(e,r)$ for two edges $e$ of $\Lambda$; (b) a boundary condition $\psi$ where the spatial mixing property does not hold.*

box of side length $2r+1$ centered at $e$; otherwise $B(e,r)$ intersects $\partial\Lambda$ (see Figure 4.2(a)). Let $E(e,r)$ be the set of edges in $E_n$ with both endpoints in $B(e,r)$, and let $E^c(e,r) = E_n \setminus E(e,r)$. The spatial mixing property we use, which is slightly weaker than that defined in [1], states that for all $e \in E_n$ and for every pair of configurations $A_1^c, A_2^c$ on $E^c = E^c(e,r)$, we have

$$\left|\mu^{\eta}_{\Lambda,p,q}(\, e = 1 \mid A_1^c\,) - \mu^{\eta}_{\Lambda,p,q}(\, e = 1 \mid A_2^c\,)\right| \le \mathrm{e}^{-\lambda r} \tag{4.3}$$

for some constant $\lambda > 0$. Alexander [1] showed that (4.2) implies (4.3) for a certain class of boundary conditions $\eta$ when $q$ is an integer. In Section 4.3 we will show, using the machinery developed in [1], that (4.3) holds for all side-homogeneous boundary conditions $\eta$ for any (not necessarily integer) $q \ge 1$. We shall see that (4.3) does not hold for arbitrary boundary conditions (see, e.g., Figure 4.2(b), together with the detailed explanation in Section 4.3).

**Glauber dynamics.** Let $\mathcal{M}$ denote the heat-bath Glauber dynamics on $\Lambda = (\Lambda_n, E_n)$ reversible with respect to $\mu^{\eta}_{\Lambda,p,q}$. Recall from Chapter 1 that given a random-cluster configuration $A_t \subseteq E_n$ at time $t$, a step of $\mathcal{M}$ results in a new configuration $A_{t+1}$ as follows:

(i) pick $e \in E_n$ u.a.r;

(ii) let $A_{t+1} = A_t \cup \{e\}$ with probability

$$\begin{cases} \frac{p}{p+q(1-p)} & \text{if } e \text{ is a "cut edge" in } (\Lambda_n, A_t); \\ p & \text{otherwise;} \end{cases}$$

(iii) else let $A_{t+1} = A_t \setminus \{e\}$.

In this setting, $e$ is a cut edge in $(\Lambda_n, A_t)$ iff changing the current configuration of $e$ changes the number of connected components of $A_t$, but also taking into account the connectivities introduce by the boundary condition $\eta$.

**Identity coupling.** Couplings of Markov chains and their connection with mixing times were discussed in Section 2.2. One coupling of the heat-bath dynamics will be of particular interest to us. Namely, consider the coupling that couples the evolution of two copies of $\mathcal{M}$, $\{X_t\}$ and $\{Y_t\}$, by using the same random $e \in E_n$ in step (i) and the same uniform random number $r \in [0, 1]$ to decide whether to add or remove $e$ in steps (ii) and (iii). We call this the *identity coupling*. It is straightforward to verify that, when $q \geq 1$, the identity coupling is a *monotone coupling*, in the sense that if $X_t \subseteq Y_t$ then $X_{t+1} \subseteq Y_{t+1}$ with probability 1. In fact, the identity coupling can be extended to a simultaneous coupling of *all* configurations that preserves the partial order $\subseteq$. Therefore, the coupling time starting from any pair of configurations is bounded by the coupling time for initial configurations $Y_0 = \emptyset$ and $X_0 = E_n$, which are the unique minimal and maximal elements in the partial order [45].

## 4.2 The speed of disagreement percolation

In spin systems, a central idea in the analysis of local Markov chains is to bound the speed at which information propagates. *Disagreement percolation* (or *path of disagreements*) arguments provide bounds of this sort (see, e.g., [4, 16]). These arguments are based on the idea that in spin systems interactions only occur between neighboring sites, and thus if two configurations agree everywhere except in some region $A$, then it takes many steps for a local Markov chain under the identity coupling to propagate these disagreements to regions that are far from $A$.

In this section, we provide a bound on the speed of propagation of disagreements for the Glauber dynamics of the random-cluster model on $\Lambda = (\Lambda_n, E_n)$, under side-homogeneous boundary conditions. A random-cluster configuration may exhibit long range interactions in the form of arbitrarily long paths, so disagreements could potentially propagate arbitrarily fast. Our insight is to restrict attention to pairs of configurations where one of them is *stationary*; i.e., it has law $\mu^\eta_{\Lambda,p,q}$. Then by (4.2), the probability of long paths decays exponentially with the length of the path, which makes the long range interactions manageable.

Throughout this section we will use the notation introduced in Section 4.1. In addition, for a random-cluster configuration $A$ on $\Lambda$ and any $D \subseteq \Lambda_n$, we use $A(D)$ to denote the configuration induced by $A$ on the edges with both endpoints in $D$. Also, we use $\partial D$ and $\partial E(D)$ to denote the set of vertices and edges, respectively, on the boundary of $D$; that is, $\partial D$ is the set of vertices in $D$ connected by an edge in $E_n$ to $\Lambda_n \setminus D$ and $\partial E(D) := \{(u, v) \in E_n : u \in D, v \notin D\}$. (Note that if $v \in D \cap \partial\Lambda$, then $v \notin \partial D$.) We are now ready to state and proof the main result of this section.

**Lemma 4.1.** *Let $p < p_c(q)$, $q \geq 1$ and consider two copies $\{X_t\}$, $\{Y_t\}$ of the Glauber dynamics on $\Lambda = (\Lambda_n, E_n)$ with a side-homogeneous boundary condition $\eta$. Assume $Y_0$ has law $\mu^\eta_{\Lambda,p,q}$ and that $X_0(B(e, r)) = Y_0(B(e, r))$ for some $e \in E_n$ and $r \geq 1$. If the evolutions of $\{X_t\}$ and $\{Y_t\}$ are coupled using the identity coupling, then there exist absolute constants $c, C, \lambda > 0$ (independent of $r$ and $n$) such that, for $m = |E_n|$, $r \geq c$ and $1 \leq k \leq r^{1/4}/(4e^2)$, we have*

$$\Pr[\, X_{km}(e) \neq Y_{km}(e)\,] \;\leq\; C e^{-\lambda r^{1/4}}.$$


(a) (b)

Figure 4.3: *(a) a random-cluster configuration $A$ in $B$, with $R_0 = \Gamma(A, B)$ (shaded); (b) $R_1$, assuming edge $e$ was updated in the first step.*

*Proof.* Let $B := B(e, r)$. For some fixed $\ell$ (to be chosen later) and each $t \geq 0$ consider the event

$$\mathcal{E}_{\ell,t} := \{u \overset{Y_t}{\nleftrightarrow} v \quad \forall u, v \in B \text{ s.t. } d(u, v) > \ell\},$$

where $u \overset{Y_t}{\nleftrightarrow} v$ denotes the event that $u$ and $v$ are not connected by a path in $Y_t(B)$. Let $\mathcal{E}_\ell := \bigcap_{t=0}^{km} \mathcal{E}_{\ell,t}$; then,

$$\Pr[\, X_{km}(e) \neq Y_{km}(e)\, ] \leq \Pr[\, X_{km}(e) \neq Y_{km}(e) \,|\, \mathcal{E}_\ell\, ] + \Pr[\, \neg\mathcal{E}_\ell\, ]. \tag{4.4}$$

We bound each term on the right hand side of (4.4) separately.

For any random-cluster configuration $A$ on $\Lambda$, let

$$\Gamma(A, B) := B \setminus \bigcup\nolimits_{v \in \partial B} C(v, A), \tag{4.5}$$

where $C(v, A)$ is the set of vertices in the connected component of $v$ in $(\Lambda_n, A)$.

Consider the sequence $R_0 \supseteq R_1 \supseteq \ldots$ of subsets (or *regions*) of $B$, such that $R_0 = \Gamma(Y_0, B)$ and

$$R_{t+1} = \begin{cases} R_t & \text{if no edge from } \partial E(R_t) \text{ is updated at time } t; \\ \\ R_t \setminus C(a_t, Y_t) & \text{if } (a_t, b_t) \in \partial E(R_t) \text{ with } a_t \in R_t, b_t \notin R_t \text{ is the edge updated} \\ & \text{at time } t; \end{cases}$$

(see Figures 4.3(a), 4.3(b)). The second case above applies regardless of the state of $(a_t, b_t)$ in $X_{t+1}$ and $Y_{t+1}$. Observe that $R_t$ need not be a connected region of $\mathbb{Z}^2$ and that every edge in $\partial E(R_t)$ is closed in both $X_t$ and $Y_t$.

The key observation is that, for all $t \geq 0$, $R_t$ is a region of $B$ in which $X_t(R_t) = Y_t(R_t)$. We prove this by induction on $t$. Assume $X_t(R_t) = Y_t(R_t)$ and let $\psi_X$ (resp., $\psi_Y$) be the boundary condition induced in $R_t$ by $X_t(\Lambda_n \setminus R_t)$ (resp., $Y_t(\Lambda_n \setminus R_t)$) and $\eta$. Both $\psi_X$ and $\psi_Y$ are partitions of $\partial R_t \cup (R_t \cap \partial\Lambda)$. (Recall that, by definition, if $v \in R_t \cap \partial\Lambda$ then $v \notin \partial R_t$.)

We consider three cases based on the location of the edge $(a_t, b_t)$ with respect to $R_t$. First, if $\{a_t, b_t\} \cap R_t = \emptyset$, then clearly $X_{t+1}(R_{t+1}) = Y_{t+1}(R_{t+1})$. The second possibility is that $\{a_t, b_t\} \cap R_t = \{a_t\}$; if this is the case, then $R_{t+1} = R_t \setminus C(a_t, Y_t)$ by definition. Moreover, $C(a_t, X_t) = C(a_t, Y_t)$ because $X_t(R_t) = Y_t(R_t)$ and all edges in $\partial E(R_t)$ are closed in both $X_t$ and $Y_t$; as a result $X_{t+1}(R_{t+1}) = Y_{t+1}(R_{t+1})$.

Finally, when $\{a_t, b_t\} \subseteq R_t$ we show that $\psi_X = \psi_Y$; from this follows that $X_{t+1}(R_{t+1}) = Y_{t+1}(R_{t+1})$. First observe that every vertex in $\partial R_t$ is a singleton in both $\psi_X$ and $\psi_Y$, since every edge in $\partial E(R_t)$ is closed in $X_t$ and $Y_t$. Therefore, if $|R_t \cap \partial\Lambda| \le 1$, then $\psi_X, \psi_Y$ are both the free boundary condition on $R_t$ and we are done. Otherwise, assume that $|R_t \cap \partial\Lambda| \ge 2$ and let $u, v$ be any two distinct vertices in $R_t \cap \partial\Lambda$. If $u$ and $v$ are wired in $\eta$, then they are also wired in $\psi_X$ and $\psi_Y$. Moreover, if $u$ and $v$ are not wired in $\eta$, then property (P1) of side-homogeneous boundary conditions implies that one of them (say, $v$) is necessarily a singleton element in $\eta$. Since there is no path of open edges from $v$ to $\Lambda_n \setminus R_t$ in either $X_t$ or $Y_t$, then $v$ is also a singleton element (and thus not wired to $u$) in both $\psi_X$ and $\psi_Y$. Hence, $u, v \in R_t \cap \partial\Lambda$ are wired in $\psi_X$ (resp., $\psi_Y$) iff they are wired in $\eta$. Since all the vertices in $\partial R_t$ are singletons in both $\psi_X$ and $\psi_Y$, we conclude that $\psi_X = \psi_Y$.

We now have that $X_t(R_t) = Y_t(R_t)$ for all $t \ge 0$. Hence, if both endpoints of $e$ lie in $R_{km}$, then $X_{km}(e) = Y_{km}(e)$. Also, since we are conditioning on $\mathcal{E}_\ell$ and we will choose $\ell \ll r$, both endpoints of $e$ lie in $R_0$. So, if $X_{km}(e) \ne Y_{km}(e)$, we may take $v_0$ to be the first endpoint of $e$ to be removed from $R_{t_0}$, at some time $t_0 \le km$. Let $e_1 \in \partial E(R_{t_0})$ be the edge whose update is responsible for removing $v_0$ from $R_{t_0}$. Starting from $e_1$ we can then construct a sequence of edges $e_1, e_2, \ldots$ such that $e_i = (u_i, v_i) \in \partial R_{t_{i-1}}$, with $u_i \in R_{t_{i-1}}$ and $v_i \notin R_{t_{i-1}}$, is the edge that removes $v_{i-1}$ from $R_{t_{i-1}}$ at time $t_{i-1}$. Note that $t_0 > t_1 > \ldots$ and that the sequence $e_1, e_2, \ldots$ stops once it reaches an edge $e_{\hat{t}}$ incident to a vertex outside $R_0$. We call the sequence $e_1, e_2, \ldots, e_{\hat{t}}$ a *witness* for the fact that $X_{km}(e) \ne Y_{km}(e)$.

The vertices $v_{i-1}$ and $u_i$ are in the same connected component in $Y_{t_{i-1}}$, and since we are conditioning on $\mathcal{E}_\ell$, we have $d(v_{i-1}, u_i) \le \ell$. Therefore, the number of witnesses of length $L$ is (crudely) at most $(4(\ell+1)^2)^L$. Note also that every witness contains at least $r/(\ell+1)$ edges, otherwise it cannot reach any of the vertices outside $R_0$. Moreover, the probability that a given witness of length $L$ is updated by the identity coupling in $km$ steps is $\binom{km}{L}\left(\frac{1}{m}\right)^L$. Hence,

$$
\begin{aligned}
\Pr[\, X_{km}(e) \ne Y_{km}(e) \mid \mathcal{E}_\ell \,] \le & \sum_{L \ge \frac{r}{\ell+1}} \binom{km}{L}\left(\frac{1}{m}\right)^L (4(\ell+1)^2)^L \\
\le & \sum_{L \ge \frac{r}{\ell+1}} \left(\frac{4\mathrm{e}k(\ell+1)^2}{L}\right)^L \\
\le & \ \omega^{\frac{r}{\ell+1}} \sum_{L \ge 0} \omega^L,
\end{aligned}
$$

where $\omega = \frac{4\mathrm{e}k(\ell+1)^3}{r}$. By taking $\ell = r^{1/4} - 1$ and using the fact that $k \le r^{1/4}/(4\mathrm{e}^2)$, we have

$$
\Pr[\, X_{km}(e) \ne Y_{km}(e) \mid \mathcal{E}_\ell \,] \le \frac{\mathrm{e}}{\mathrm{e}-1} \cdot \mathrm{e}^{-r^{3/4}}. \tag{4.6}
$$

Now we turn our attention to the second term on the right hand side of (4.4). Let $N$ be the number of updates the identity coupling performs in $B$ up to time $km$, and let $M$ be the number of edges in $B$; i.e., $M := |E(e,r)| = \Theta(r^2)$. A Chernoff bound implies that $N > 2kM$ with probability $\exp(-\Omega(kM))$, and thus

$$\Pr[\,\neg\mathcal{E}_\ell\,] \le \Pr[\,\neg\mathcal{E}_\ell \mid N \le 2kM\,] + \mathrm{e}^{-\Omega(kM)}.$$

Observe that $\neg\mathcal{E}_\ell = \bigcup_{t=0}^{km} \neg\mathcal{E}_{\ell,t}$, and if the edge update at time $t$ occurs outside $B$, we have $\neg\mathcal{E}_{\ell,t} = \neg\mathcal{E}_{\ell,t+1}$. Hence, a union bound implies

$$\Pr[\,\neg\mathcal{E}_\ell\,] \;\le\; 2kM \max_{0\le t\le km} \Pr[\,\neg\mathcal{E}_{\ell,t}\,] + \mathrm{e}^{-\Omega(kM)}. \tag{4.7}$$

To bound $\Pr[\,\neg\mathcal{E}_{\ell,t}\,]$ we use the exponential decay of finite volume connectivities (4.2). To do this, recall that $Y_0$ (and thus $Y_t$ for all $t \ge 0$) has law $\mu^\eta$. Also, there are only $O(r^4)$ pairs of vertices in $B$. Hence, (4.2) and a union bound imply

$$\Pr[\,\neg\mathcal{E}_{\ell,t}\,] \le O(r^4)\cdot \mathrm{e}^{-\Omega(\ell)} \le O(r^4)\cdot \mathrm{e}^{-\Omega(r^{1/4})}.$$

Since $1 \le k \le r^{1/4}/(4\mathrm{e}^2)$ and $M = \Theta(r^2)$, (4.7) gives

$$\Pr[\,\neg\mathcal{E}_\ell\,] \le O(r^{6.25})\cdot \mathrm{e}^{-\Omega(r^{1/4})} + \mathrm{e}^{-\Omega(r^2)}.$$

Together with (4.4) and (4.6), this implies that there exist constants $c, C, \lambda > 0$ such that for all $r \ge c$, we have

$$\Pr[\,X_{km}(e) \ne Y_{km}(e)\,] \le C\mathrm{e}^{-\lambda r^{1/4}},$$

as desired. □

## 4.3 Spatial mixing for side-homogeneous boundary conditions

In this section we show that the spatial mixing (4.3) holds for the class of side-homogeneous boundary conditions on $\Lambda = (\Lambda_n, E_n)$. Let $e \in E_n$ and let $B = B(e,r)$ for some $r \ge 1$. Spatial mixing holds when the influence on $e$ of the configuration in $E^c = E^c(e,r)$ decays exponentially with $r$. This is easy to establish when $B \cap \partial\Lambda = \emptyset$, since such influence is present only if there are paths of open edges from $e$ to $\partial B$, and, by (4.2), the probability of such paths decays exponentially with $r$. However, if $B \cap \partial\Lambda \ne \emptyset$, the influence from $E^c$ could also propagate along $B \cap \partial\Lambda$ via the boundary condition on $\Lambda$. This is why (4.3) does not hold for arbitrary boundary conditions, as the following concrete example illustrates.

With a slight abuse of notation, we use $\{E^c = 1\}$ (resp., $\{E^c = 0\}$) to denote the event that all the edges in $E^c$ are open (resp., closed). Suppose $e = (u,v)$ is an edge in $\partial\Lambda$ that is far from the corners of $\Lambda$, and let $\psi$ be the boundary condition on $\Lambda$ where $u$ is wired to a vertex $u' \in \partial\Lambda \setminus B$

and $v$ is wired to a different vertex $v' \in \partial\Lambda \setminus B$ (see Figure 4.2(b)). When $p = 1/2$ and $q = 3$, we have $\mu^{\psi}(e = 1|E^c = 1) = 1/2$. Also, by considering a small box around $e$, is easy to check that $\mu^{\psi}(e = 1|E^c = 0) \leq 2/5$. Both these bounds are independent of $r$ and $n$; consequently, $\psi$ does not have the spatial mixing property.

It turns out that side-homogeneous boundary conditions (and in particular property (P1)) rule out the possibility of influence propagating along $B \cap \partial\Lambda$. As a result, we are able to establish the spatial mixing property for side-homogeneous boundary conditions, as stated in the following lemma; the proof uses the machinery developed in [1].

**Lemma 4.2.** *Let $p < p_c(q)$, $q \geq 1$ and let $\eta$ be a side-homogeneous boundary condition for $\Lambda = (\Lambda_n, E_n)$. For any $e \in E_n$, there exist constants $c, \lambda > 0$ such that for all $r \geq c$ and every pair of configurations $A_1^c$, $A_2^c$ on $E^c$:*

$$|\mu^{\eta}(\, e = 1 \,|\, A_1^c\,) - \mu^{\eta}(\, e = 1 \,|\, A_2^c\,)| \;\leq\; \mathrm{e}^{-\lambda r}. \tag{4.8}$$

*Proof.* Consider the measure $\mu^w := \mu^{\eta}(\,\cdot\,|E^c = 1)$ on $B = B(e, r)$. Let $\Gamma(\cdot, \cdot)$ be defined as in (4.5). We derive the result from the following key fact, which we prove later.

**Claim 4.3.** *There exists a coupling $\pi$ of the distributions $\mu^{\eta}(\,\cdot\,|\,A_1^c\,)$, $\mu^{\eta}(\,\cdot\,|\,A_2^c\,)$ and $\mu^w$ such that $\pi(A_1, A_2, A_w) > 0$ only if $A_1 \subseteq A_w$, $A_2 \subseteq A_w$ and $A_1$, $A_2$ agree on all edges with both endpoints in $\Gamma(A_w, B)$.*

Let $\Gamma^c(A_w, B) := \Lambda_n \setminus \Gamma(A_w, B)$. Given the coupling $\pi$, we have

$$\begin{aligned}
|\mu^{\eta}(\, e = 1 \,|\, A_1^c\,) - \mu^{\eta}(\, e = 1 \,|\, A_2^c\,)| &\leq \pi(\, A_1(e) \neq A_2(e)\,) \\
&\leq \pi\,(\,\Gamma^c(A_w, B) \cap \{e\} \neq \emptyset\,) \\
&\leq \mu^w\,(\,\{e\} \leftrightarrow \partial B\,)\,,
\end{aligned}$$

where $\{e\} \leftrightarrow \partial B$ denotes the event that there is a path from $e$ to $\partial B$.

Now, exponential decay of finite volume connectivities (4.2) and a union bound over the boundary vertices imply

$$\mu^w(\,\{e\} \leftrightarrow \partial B\,) \leq 2C|\,\partial B|\mathrm{e}^{-\lambda r}.$$

Since $|\,\partial B| = \Theta(r)$, we obtain (4.8) for all $r \geq c$ for some constant $c > 0$, and hence the lemma. □

We conclude this section by providing the missing proof of Claim 4.3.

*Proof.* Let $\theta_1$ (resp., $\theta_2$) be the boundary condition induced on $B = B(e, r)$ by $A_1^c$ (resp., $A_2^c$) and $\eta$. Note that $\mu^{\theta_1}$, $\mu^{\theta_2}$ and $\mu^w$ are random-cluster measures on $B$ with different boundary conditions, and clearly $\mu^w \geq \mu^{\theta_1}$ and $\mu^w \geq \mu^{\theta_2}$. Strassen's theorem (see, e.g., [47]) then implies the existence of monotone couplings $\mu_1$ for $\mu^{\theta_1}$ and $\mu^w$, and $\mu_2$ for $\mu^{\theta_2}$ and $\mu^w$. (Recall that $\mu_1$ is a monotone coupling for $\mu^{\theta_1}$ and $\mu^w$ iff every sample $(A_{\theta_1}, A_w)$ from $\mu_1$ satisfies $A_{\theta_1} \subseteq A_w$.) We show next how to use $\mu_1$ and $\mu_2$ to construct the desired coupling $\pi$.

First, let $\Delta := \Gamma(A_w, B)$ and let $\xi$ be the boundary condition induced in $\Delta$ by $\eta$ and the configuration of $A_w$ in $\Gamma^c(A_w, B)$. We construct $\pi$ as follows:

(i) sample $(A_{\theta_1}, A_w)$ from $\mu_1$;

(ii) sample $A_{\theta_2}$ from $\mu_2(\,\cdot \mid A_w\,)$; and

(iii) sample $A_\gamma$ from $\mu^{\xi}_{\Delta,p,q}$.

Let $\pi$ be the distribution of

$$(\{A_{\theta_1} \setminus E(\Delta)\} \cup A_\gamma, \{A_{\theta_2} \setminus E(\Delta)\} \cup A_\gamma, \{A_w \setminus E(\Delta)\} \cup A_\gamma)$$

after step (iii), where $E(\Delta)$ denotes the set of edges with both endpoints in $\Delta$.

A straightforward calculation reveals that $A_{\theta_2}$ has law $\mu^{\theta_2}$, and thus after step (ii) the distribution of $(A_{\theta_1}, A_{\theta_2}, A_w)$ has all the right marginals. Moreover, since $\mu_1$ and $\mu_2$ are monotone couplings, $A_{\theta_1} \subseteq A_w$ and $A_{\theta_2} \subseteq A_w$.

We argue next that replacing the configuration in $\Delta$ with $A_\gamma$ in step (iii) has no effect on the distribution. For this, let $\xi_1$ (resp., $\xi_2$) be the boundary condition induced in $\Delta$ by $\eta$ and the configuration of $A_{\theta_1}$ (resp., $A_{\theta_2}$) in $\Gamma^c(A_w, B)$. (Note that $\xi$, $\xi_1$, $\xi_2$ are partitions of $\partial\Delta \cup (\Delta \cap \partial\Lambda)$.)

We show that $\xi = \xi_1 = \xi_2$. This is easy to see when $\Delta \cap \partial\Lambda = \emptyset$, since in this case all three of them are the free boundary condition on $\Delta$. This is because $A_{\theta_1} \subseteq A_w$, $A_{\theta_2} \subseteq A_w$ and every edge from $\partial\Delta$ to $\Delta^c$ is closed in $A_w$.

When $\Delta \cap \partial\Lambda$ is not trivial, only vertices in $\Delta \cap \partial\Lambda$ may be wired in $\xi$. Property (P1), together with the fact that every edge from $\partial\Delta$ to $\Delta^c$ is closed in $A_w$, implies that two vertices from $\Delta \cap \partial\Lambda$ are wired in $\xi$ iff they are wired in $\eta$. The same holds for $\xi_1$ and $\xi_2$; hence, $\xi = \xi_1 = \xi_2$. (Note that this argument is essentially the same as the one used in the proof of Lemma 4.1 to show that the boundary conditions induced in $R_t$ by $X_t$ and $Y_t$ are the same.)

Finally, the *domain Markov property* of random-cluster measures (see, e.g., [26]) ensures that indeed replacing $\Delta$ with $A_\gamma$ has no effect on the distribution. Hence, $\pi$ is a coupling of the measures $\mu^{\theta_1}$, $\mu^{\theta_2}$, and $\mu^w$ with all the desired properties. □

## 4.4 Mixing time upper bound in the sub-critical regime

In this section we prove our main result: the upper bound for the mixing time in Theorem 1.4 for $p < p_c(q)$. We state two theorems whose combination establishes the desired upper bound for $p < p_c(q)$. In Theorem 4.4 we show that spatial mixing for the class of side-homogeneous boundary conditions, as established in Section 4.3, implies a bound of $O(n^2 \log n (\log\log n)^2)$ for the mixing time of the Glauber dynamics on $\Lambda = (\Lambda_n, E_n)$, for any $n$ and any side-homogeneous boundary condition. The proof is inductive and makes crucial use of property (P2) of side-homogeneous boundary conditions, which ensures that for any $e \in E_n$ and $r \geq 1$, the boundary condition induced in $B(e,r)$ by the events $\{E^c = 1\}$ or $\{E^c = 0\}$ is also side-homogeneous.

In Theorem 4.5 we show that a sufficiently good upper bound on the mixing time of the Glauber dynamics—in fact $O\left(n^{2.25}/\log n\right)$ suffices—can be bootstrapped to the desired upper bound of $O(n^2 \log n)$. The proof of Theorem 4.5 crucially uses the bounds on the speed of propagation of disagreements from Section 4.2. Our proofs are inspired by those in the spin systems literature, in particular those in [16, 41, 44].

**Theorem 4.4.** *Let $p < p_c(q)$, $q \geq 1$ and let $\eta$ be a side-homogeneous boundary condition for $\Lambda = (\Lambda_n, E_n)$. There exists a fixed constant $C > 0$ such that for all $n$, the mixing time of the Glauber dynamics in $\Lambda$ is at most $T(m) = Cm \log m (\log \log m)^2$, where $m := |E_n| = \Theta(n^2)$.*

*Proof.* We bound the coupling time $T_{\text{coup}}$ of the identity coupling; the result then follows from the fact that $\tau_{\text{mix}} \leq T_{\text{coup}}$. Consider two copies $\{X_t\}$, $\{Y_t\}$ of the Glauber dynamics coupled with the identity coupling. We may assume $X_0 = E_n$ and $Y_0 = \emptyset$, since we know from Section 4.1 that this is the worst pair of starting configurations. We prove that

$$\Pr[\, X_T(e) \neq Y_T(e)\, ] = O\left(m^{-2}\right)$$

for $T = T(m)$ and for all $e \in E_n$. The bound for the coupling time then follows from a union bound over the $m$ edges.

To bound $\Pr[\, X_T(e) \neq Y_T(e)\, ]$, we introduce two additional copies $\{Z_t^+\}$, $\{Z_t^-\}$ of the Glauber dynamics. These two copies will only update edges with both endpoints in the box $B = B(e, r)$, for a suitable $r$ we choose later. We set $Z_0^+ = X_0 = E_n$ and $Z_0^- = Y_0 = \emptyset$. The four Markov chains $\{X_t\}$, $\{Y_t\}$, $\{Z_t^+\}$ and $\{Z_t^-\}$ are coupled with the identity coupling, and the updates outside $B$ are ignored by $\{Z_t^+\}$ and $\{Z_t^-\}$. By monotonicity of the identity coupling, we have $Z_t^- \subseteq Y_t \subseteq X_t \subseteq Z_t^+$ for all $t \geq 0$. Hence,

$$\begin{aligned} \Pr[\, X_t(e) \neq Y_t(e)\, ] &\leq \Pr[\, Z_t^+(e) \neq Z_t^-(e)\, ] \\ &= \Pr[\, Z_t^+(e) = 1\, ] - \Pr[\, Z_t^-(e) = 1\, ]. \end{aligned}$$

The stationary distributions of $\{Z_t^+\}$ and $\{Z_t^-\}$ are $\mu_\Lambda^\eta(\,\cdot\,|\, E^c = 1\,)$ and $\mu_\Lambda^\eta(\,\cdot\,|\, E^c = 0)$, respectively, where as usual $\{E^c = 1\}$ (resp., $\{E^c = 0\}$) denotes the event that all the edges in $E_n \backslash E(e, r)$ are open (resp., closed). The triangle inequality then implies

$$\begin{aligned} \Pr[\, X_t(e) \neq Y_t(e)\, ] \;\leq\; & \left|\Pr[\, Z_t^+(e) = 1\, ] - \mu_\Lambda^\eta(\, e = 1 \,|\, E^c = 1\,)\right| && (4.9) \\ & + \left|\mu_\Lambda^\eta(\, e = 1 \,|\, E^c = 1\,) - \mu_\Lambda^\eta(\, e = 1 \,|\, E^c = 0\,)\right| && (4.10) \\ & + \left|\mu_\Lambda^\eta(\, e = 1 \,|\, E^c = 0\,) - \Pr[\, Z_t^-(e) = 1\, ]\right|. && (4.11) \end{aligned}$$

The chains $\{Z_t^+\}$ and $\{Z_t^-\}$ are (lazy) Glauber dynamics on the smaller square box $B$. Moreover, the boundary conditions induced in $B$ by $\eta$ and the events $\{E^c = 1\}$, $\{E^c = 0\}$ are side-homogeneous. Hence, we proceed inductively.

First note that for any fixed $m_0$, the result holds for all square boxes of volume at most $m_0$ by simply adjusting the constant $C = C(m_0)$. Now, let $r = c \log m$ for some constant $c$ we choose later, and assume the mixing time bound holds for square boxes of volume $M := |E(e, r)|$ with side-homogeneous boundary condition. After $T(m)$ steps, the expected number of updates in $B$ is

$$T(m)\frac{M}{m} = CM\log m(\log\log m)^2 \geq \lceil 4\log_2 m\rceil T(M),$$

where we choose $m_0$ such that the last inequality holds for all $m > m_0$. Hence, a Chernoff bound implies that the number of updates in $B$ is at least $\lceil 2\log_2 m\rceil T(M)$ with probability at least $1-m^{-2}$.

The induction hypothesis implies that the mixing time of $\{Z_t^+\}$ is at most $T(M)$. Hence, if indeed $\lceil 2\log_2 m\rceil T(M)$ updates happen in $B$, then

$$||\, Z_t^+(\cdot) - \mu_\Lambda^\eta(\,\cdot\,|\, E^c = 1\,)\,||_{\mathrm{TV}} \leq \frac{1}{m^2}.$$

(Here we used the fact from Section 4.1 that $\tau_{\mathrm{mix}}(\varepsilon) \leq \lceil\log_2 \varepsilon^{-1}\rceil\, \tau_{\mathrm{mix}}$.) Combining this with the above Chernoff bound, we have

$$\Big|\Pr[\, Z_t^+(e) = 1\,] - \mu_\Lambda^\eta(\, e = 1 \,|\, E^c = 1)\Big| \leq \frac{2}{m^2}.$$

The quantity in (4.11) is bounded similarly.

Finally, taking $c = 2/\lambda$, the spatial mixing property (Lemma 4.2) implies that (4.10) is at most $1/m^2$. Putting these bounds together we get

$$\Pr[\, X_t(e) \neq Y_t(e)\,] \leq \frac{5}{m^2},$$

as desired. ◻

**Theorem 4.5.** *Let $p<p_c(q)$, $q\geq 1$ and let $m_0$, $c$ be sufficiently large and sufficiently small constants, respectively. Assume that the mixing time of the Glauber dynamics in any square box of volume $m_0$ with side-homogeneous boundary conditions is at most $\frac{c\, m_0^{9/8}}{\log_2 m_0}$. Then, the mixing time of the Glauber dynamics in $\Lambda$ with side-homogeneous boundary conditions is $O(n^2\log n)$.*

*Proof.* Let $m := |E_n| = \Theta(n^2)$ and the let $\eta$ be a side-homogeneous boundary condition for $\Lambda$. Also, let $\{X_t\}$, $\{Y_t\}$ be two copies of the Glauber dynamics in $\Lambda$ coupled with the identity coupling. We prove that for $1 \leq k = o(m^{1/8})$, we have

$$\Pr[X_{km}(e) \neq Y_{km}(e)] \leq \mathrm{e}^{-\Omega(k)} \tag{4.12}$$

for any $e \in E_n$ and any pair of initial configurations $X_0, Y_0$. Hence, for some $k = O(\log m)$ we have $\Pr[X_{km}(e) \neq Y_{km}(e)] \leq 1/(4m)$, and a union bound over the edges implies $\tau_{\mathrm{mix}} \leq T_{\mathrm{coup}} = O(m\log m)$, as desired.

We bound $\Pr[X_{km}(e) \neq Y_{km}(e)]$ for the case where $X_0 = E_n$ and $Y_0$ is sampled from $\mu^\eta$. The proof for the case where $X_0 = \emptyset$ and $Y_0$ is sampled from $\mu^\eta$ is identical. The monotonicity of the identity coupling discussed in Section 4.1 implies then that this bound holds for arbitrary initial configurations $X_0$, $Y_0$.

Let $\mathcal{E}_k$ be the event $\{X_{km}(B(e,r)) = Y_{km}(B(e,r))\}$ for some $r$ we will choose later, and let

$$\rho(k) := \max_{e \in E_n} \Pr[\, X_{km}(e) \neq Y_{km}(e) \,],$$

where the probability is over both the choice of $Y_0$ and the steps of the Markov chain. We will show that $\rho(k) \le \exp(-\Omega(k))$.

Since $Y_0$ has law $\mu_\Lambda^\eta$, the bound on disagreement percolation in Lemma 4.1 implies that

$$\Pr[\, X_{2km}(e) \neq Y_{2km}(e) \mid \mathcal{E}_k \,] \le \mathrm{e}^{-\Omega(r^{1/4})},$$

provided $1 \le k \le r^{1/4}/(4\mathrm{e}^2)$ and $r$ is large enough. Thus,

$$\Pr[\, X_{2km}(e) \neq Y_{2km}(e) \,] \le \Pr[\, X_{2km}(e) \neq Y_{2km}(e) \mid \neg\mathcal{E}_k \,] \Pr[\, \neg\mathcal{E}_k \,] + \mathrm{e}^{-\Omega(r^{1/4})}.$$

The monotonicity of the identity coupling implies that

$$\Pr[\, X_{2km}(e) \neq Y_{2km}(e) \mid \neg\mathcal{E}_k \,] \le \rho(k),$$

and a union bound over the edges in $B(e,r)$ implies $\Pr[\, \neg\mathcal{E}_k \,] \le \Theta(r^2)\rho(k)$. Putting these bounds together and taking $r = \Theta(k^4)$, we obtain

$$\rho(2k) \le Ck^8\rho^2(k) + \mathrm{e}^{-\lambda k},$$

for some suitable constants $C > 1$ and $\lambda > 0$. (Note that since $r = \Theta(k^4)$ and $r < n$, our proof of inequality (4.12) does not hold for arbitrarily large $k$; hence the restriction $k = o(m^{1/8})$.)

Now, let

$$\phi(k) := 2^8(Ck^8 + 1)\max\{\rho(k), \mathrm{e}^{-\lambda k/2}\}.$$

We show next that $\phi(2k) \le \phi(k)^2$. For this observe that $2^8(C(2k)^8 + 1)\rho(2k) \le \phi(k^2)$. Hence, if $\rho(2k) \ge e^{-\lambda k}$, we get $\phi(2k) \le \phi(k^2)$ directly. Otherwise, we have

$$\phi(2k) \le (2^8)^2(Ck^8 + 1)\mathrm{e}^{-\lambda k} \le \phi(k)^2.$$

Thus, for any integer $\alpha > 0$, we get $\phi(k) \le \phi(k/2^\alpha)^{2^\alpha}$. The result follows from the following fact, which provides a stopping point for this recurrence.

**Claim 4.6.** *Let $l = m_0^{1/8}/A$, with $A = 2(8C\mathrm{e})^{1/8}$. Then $\rho(l) \le \frac{1}{2^8\,\mathrm{e}(Cl^8+1)}$ for a sufficiently large $m_0$.*

As a result, $\phi(l) \le 1/\mathrm{e}$ for a sufficiently large constant $l$, and thus $\rho(k) \le \phi(k) \le \exp(-k/l)$ as desired. □

We conclude this section with the proof of Claim 4.6. The proof is similar to that of Theorem 5.1 and makes crucial use of the hypothesis on the mixing time in square boxes of volume $m_0$.

*Proof.* Let $e \in E_n$ and choose $r'$ such that $|E(e,r')| = m_0$. (Note that as a result $r' = \Theta(m_0^{1/2})$.) The proof proceeds along the same lines as that of Theorem 4.4. In fact we consider the same processes $\{Z_t^+\}$, $\{Z_t^-\}$, where $Z_0^+ = X_0 = E_n$, $Z_t^- = \emptyset$ and $\{Z_t^+\}$, $\{Z_t^-\}$ only update edges with both endpoints in $B(e,r')$. As before, we couple the four chains $\{X_t\}$, $\{Y_t\}$, $\{Z_t^+\}$, $\{Z_t^-\}$ with the identity coupling, ignoring the updates outside $B(e,r')$ in $\{Z_t^+\}$ and $\{Z_t^-\}$. The monotonicity of the identity coupling then implies that $Z_t^- \subseteq Y_t \subseteq X_t \subseteq Z_t^+$ for all $t \geq 0$. Hence, we obtain inequality (4.9)-(4.11). (Note that in this case $Y_t$ has law $\mu_\Lambda^\eta$ for all $t \geq 0$.)

Lemma 4.2 implies that (4.10) is at most $\exp(-\Omega(r'))$. To bound (4.9), note that if we run the identity coupling for $lm$ steps, a Chernoff bound implies that with probability at least $1 - \exp(-lm_0/8)$ the number of updates in $B(e,r')$ is at least $(2A)^{-1}m_0^{9/8}$. If indeed this many steps hit $B(e,r')$, then the hypothesis of the theorem (with $c = (2A)^{-1}$) implies

$$\left|\Pr[\, Z_t^+(e) = 1\,] - \mu_\Lambda^\eta(\, e = 1 \mid E^c = 1)\right| \leq ||\, Z_t^+(\cdot) - \mu_\Lambda^\eta(\,\cdot \mid E^c = 1\,)\,||_{\mathrm{TV}} \leq \frac{1}{m_0}.$$

(Here we also used the fact that $\tau_{\mathrm{mix}}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil \tau_{\mathrm{mix}}$.) We do the same to bound (4.11), and then

$$\begin{aligned}\Pr[\, X_{lm}(e) \neq Y_{lm}(e)\,] &\leq \frac{2}{m_0} + e^{-\Omega(m_0^{9/8})} + e^{-\Omega(m_0^{1/2})} \\ &\leq \frac{4}{m_0} = \frac{4}{(Al)^8} \leq \frac{1}{2^8\,\mathrm{e}(Cl^8+1)},\end{aligned}$$

for a sufficiently large constant $m_0$. Hence, $\phi(l) \leq 1/\,\mathrm{e}$ for a sufficiently large $l$, as desired. ◻

## 4.5 Mixing time lower bound in the sub-critical regime

In this section we prove the lower bound from Theorem 1.4 for $p < p_c(q)$. (The lower bound for $p > p_c(q)$ is derived in Section 4.6.) In the setting of spin systems, [28] provides a general mixing time lower bound for Glauber dynamics. As mentioned earlier, the random-cluster model is not a spin system in the usual sense because of the long range interactions, but we are still able to adapt the techniques in [28] to the random-cluster setting. In fact, our proof follows closely the argument in the proof of Theorem 4.1 in [28], the main difference being that we require a more subtle choice of the starting configuration to limit the effect of the long range interactions. We derive the following theorem.

**Theorem 4.7.** *Let $p < p_c(q)$, $q \geq 1$ and let $\eta$ be a side-homogeneous boundary condition for $\Lambda = (\Lambda_n, E_n)$. The mixing time of the Glauber dynamics in $\Lambda$ is $\Omega(n^2 \log n)$.*

It is convenient to carry out our proof in continuous time. The continuous time Glauber dynamics is obtained by adding a rate 1 Poisson clock to each edge; when the clock at edge $e$ rings, $e$ is updated as in discrete time.

The switch to continuous time requires us to extend the bound in Section 4.2 for the speed of propagation of disagreements to the continuous time dynamics. In addition, we will require slightly different assumptions about the initial configuration $Y_0$. This is established in the following lemma, whose proof is very similar to that of Lemma 4.1 (only requiring minor adjustments) and is deferred to Appendix A.

**Lemma 4.8.** *Let $p < p_c(q)$, $q \geq 1$ and let $\eta$ be a side-homogeneous boundary condition for $\Lambda = (\Lambda_n, E_n)$. Also, let $B = B(e, r)$ for some $e \in E_n$ and $r \geq 1$. Consider two copies $\{X_t\}$, $\{Y_t\}$ of the continuous time Glauber dynamics on $\Lambda$ such that:*

- *$X_0(B) = Y_0(B)$;*
- *$Y_0(B)$ has law $\mu^0_{B,p,q}(\,\cdot \mid e = b)$ for some $b \in \{0, 1\}$;*
- *$Y_0(e') = 0$ for all $e' \in E^c(e, r)$ incident to $\partial B$; and*
- *$\{Y_t\}$ only performs edge updates in $B$.*

*If the evolutions of $\{X_t\}$ and $\{Y_t\}$ are coupled using the identity coupling, then there exist absolute positive constants $c$, $C$, and $\lambda$ (independent of $r$ and $n$) such that, for all $r \geq c$ and $1 \leq T \leq r^{1/4}/(4\mathrm{e}^2)$, we have*

$$\Pr[\, X_T(e) \neq Y_T(e)\,] \;\leq\; C\mathrm{e}^{-\lambda r^{1/4}}.$$

We are now ready to prove Theorem 4.7.

*Proof.* Let $\{X_t\}$ and $\{X_t^{\mathcal{D}}\}$ be copies of the continuous and discrete time Glauber dynamics in $\Lambda$, respectively, such that $X_0^{\mathcal{D}} = X_0$. The following standard inequality holds for all $t \geq 0$:

$$||X^{\mathcal{D}}_{t'} - \mu^\eta||_{\mathrm{TV}} \geq ||X_t - \mu^\eta||_{\mathrm{TV}} - 2e^{-t'},$$

where $t' = |E_n|t/2$ (see, e.g., Proposition 2.1 in [28]).

We will show that $||X_T - \mu^\eta||_{\mathrm{TV}} > 1/3$ for some $T = \Omega(\log n)$; as a result $||X^{\mathcal{D}}_{T'} - \mu^\eta||_{\mathrm{TV}} > 1/4$ for some $T' = \Omega(n^2 \log n)$ and sufficiently large $n$. This implies that the mixing time of the discrete time dynamics is $\Omega(n^2 \log n)$ as desired.

First we introduce some notation. Assume w.l.o.g. that $4r + 1$ divides $n - 1$ for some $r$ to be chosen later, and split $\Lambda_n$ into $(n-1)^2/(4r+1)^2$ square boxes of side length $4r + 1$. Each of these boxes corresponds to $B(e, 2r)$ for some edge $e \in E_n$; let $C \subseteq E_n$ be the set of these edges. Also, let $\hat{E} := E_n \setminus \bigcup_{e \in C} E(e, r)$ and let $\mathcal{E}$ be the event that every edge $e' \in \hat{E}$ incident to $B(e, r)$ for some $e \in C$ is closed.

Let $A$, $A_{\mathcal{E}}$ be random-cluster configurations sampled from $\mu^\eta$ and $\mu^\eta(\,\cdot \mid \mathcal{E})$, respectively, and let $\beta := \mathrm{E}[f(A_{\mathcal{E}})]$, where $f(A_{\mathcal{E}})$ is the fraction of edges $e \in C$ such that $A_{\mathcal{E}}(e) = b$ for some fixed $b \in \{0, 1\}$. Consider the following threshold for a value of $\varepsilon > 0$ that will be chosen later:

$$\hat{\beta} = \begin{cases} \beta + \varepsilon & \text{if } \beta < 1/2; \\ \beta & \text{if } \beta = 1/2; \\ \beta - \varepsilon & \text{if } \beta > 1/2. \end{cases}$$

We pick $b$ such that $\Pr[f(A) > \hat{\beta}] \le 1/2$.

As in [28], our goal is to choose $X_0$ such that $\Pr[f(X_T) \ge \hat{\beta}] \to 1$ as $n \to \infty$ for some $T = \Omega(\log n)$; then $||X_T - \mu^\eta||_{\mathrm{TV}} > 1/3$ for large enough $n$, as desired.

Let $\Phi$ be the set of random-cluster configurations in $\Lambda$ such that $\hat{A} \in \Phi$ iff for all $e \in C$, $\hat{A}(e) = b$ and $\hat{A}(e') = 0$ for all $e' \in \hat{E}$ incident to $B(e, r)$. For each $\hat{A} \in \Phi$, let

$$\pi_0(\hat{A}) := \frac{\mu^\eta(\hat{A})}{\mu^\eta(\Phi)}. \tag{4.13}$$

The starting configuration $X_0$ is sampled from $\pi_0$.

Consider now an auxiliary copy $\{Y_t\}$ of the continuous time Glauber dynamics such that $Y_0 = X_0$. The two chains $\{X_t\}$,$\{Y_t\}$ are coupled using the identity coupling, except that $\{Y_t\}$ does not update edges in $\hat{E}$. First we establish a bound for $\Pr[f(Y_T) \le \hat{\beta} + \varepsilon/2]$. To do this, we use the following monotonicity property which is a straightforward consequence of Lemma 3.5 in [28].

**Fact 4.9.** *For each $e \in C$, let $\alpha_e := \mu^0_{B,p,q}(e = b)$ where $B = B(e, r)$. Then, for all $t \ge 0$,*

$$\Pr[Y_t(e) = b] \ge \alpha_e + (1 - \alpha_e)\mathrm{e}^{-t/(1-\alpha_e)}.$$

From this fact, we follow the steps in the proof of Theorem 4.1 in [28] to derive the following bound:

$$\mathrm{E}[f(Y_T)] \ge \beta + (1 - \beta)\mathrm{e}^{-T/(1-\beta)}, \tag{4.14}$$

for all $T \ge 0$. Taking $\varepsilon = 1/(4\exp(2T))$, the right hand side of (4.14) is at least $\hat{\beta} + \varepsilon$. Also, since $Y_0$ is sampled from $\pi_0$, the configurations of the edges in $C$ are independent in $Y_T$. Hence, $|C|f(Y_T)$ is the sum of $|C|$ independent Bernoulli random variables; a Chernoff bound then implies

$$\Pr[f(Y_T) \le \hat{\beta} + \varepsilon/2] \le \mathrm{e}^{-\Omega(\varepsilon^2|C|)}. \tag{4.15}$$

The second step in the proof is to bound $\Pr[f(X_T) \le f(Y_T) - \varepsilon/2]$. For this, we use Lemma 4.8, which is tailored precisely to our setting. Thus, for all $e \in C$ and $1 \le T \le r^{1/4}/(4\mathrm{e}^2)$, we have

$$\Pr[X_T(e) \ne Y_T(e)] \le C\mathrm{e}^{-\lambda r^{1/4}},$$

provided $r \ge c$, where $c$ is a sufficiently large constant. Therefore, the expected number of disagreements between $X_T$ and $Y_T$ in the set $C$ is at most $|C|C\exp(-\lambda r^{1/4})$, and by Markov's inequality,

$$\Pr[f(X_T) \le f(Y_T) - \varepsilon/2] \le \frac{2C}{\varepsilon}\mathrm{e}^{-\lambda r^{1/4}}. \tag{4.16}$$

Putting together the bounds in (4.15) and (4.16), we get

$$\Pr[f(X_T) \le \hat{\beta}] \le \frac{2C}{\varepsilon}\mathrm{e}^{-\lambda r^{1/4}} + \mathrm{e}^{-\Omega(\varepsilon^2|C|)}.$$

Finally, observe that $|C| = \Theta(\frac{n^2}{r^2})$; thus, when $r = (\frac{1}{4}\log n)^4$ and $T = \min\{\lambda/4, 1\}r^{1/4}$, we get $\Pr[f(X_T) > \hat{\beta}] \to 1$ as $n \to \infty$ as desired. $\square$

## 4.6 Mixing time in the super-critical regime

In this section we prove Theorem 1.4 from the Introduction for $p > p_c(q)$. We will make use of planar duality, discussed in Section 4.1, in order to reduce the proof to the sub-critical case.

**Theorem 4.10.** *For $p > p_c(q)$ and $q \geq 1$, the mixing time of the Glauber dynamics on $\Lambda = (\Lambda_n, E_n)$ with free or wired boundary conditions is $\Theta(n^2 \log n)$.*

*Proof.* We focus on the free boundary condition case; the wired case follows from an analogous argument. The planar dual $\Lambda^* = (\Lambda_n^*, E_n^*)$ of the graph $\Lambda$ consists of an $(n-1)\times(n-1)$ box with an additional outer vertex corresponding to the infinite face of $\Lambda$. The dual measure $\mu_{\Lambda^*,p^*,q}$, with

$$p^* = \frac{q(1-p)}{p + q(1-p)},$$

is equivalent to the measure $\mu^1_{\Lambda',p^*,q}$ where $\Lambda' = (\Lambda_{n+1}, E_{n+1})$ (see, e.g., [3]). Note that $p > p_c(q)$ iff $p^* < p_c(q)$.

We say that two random-cluster configurations $A$ on $\Lambda$ and $A'$ on $\Lambda'$ are *compatible* if the configuration resulting from $A'$ by contracting all the vertices in the boundary of $\Lambda'$ into a single vertex is $A^*$, the dual configuration of $A$. Note that each $A'$ has a unique compatible $A$, while each $A$ has multiple compatible $A'$ that differ only in the disposition of the edges in the boundary $\partial\Lambda_{n+1}$. Observe also that any edge $e'$ of $\Lambda'$ with at most one endpoint incident to $\partial\Lambda_{n+1}$ corresponds to a unique dual edge $e^* \in E_n^*$ and thus to a unique edge $e \in E_n$.

In order to analyze the Glauber dynamics on $\Lambda$ when $p > p_c(q)$, we consider instead the Glauber dynamics on $\Lambda'$ with parameter $p^* < p_c(q)$, which we denote $\{X_t'\}$. We shall show that $\{X_t'\}$ induces a Markov chain $\{X_t\}$ on $\Lambda$ which is essentially the same as the Glauber dynamics on $\Lambda$ with parameter $p$, and that the mixing times of $\{X_t\}$ and $\{X_t'\}$ are equal up to constant factors. Since $p^* < p_c(q)$, the results in Sections 4.4 and 4.5 imply that mixing time $\{X_t'\}$ (and hence of $\{X_t\}$) is $\Theta(n^2 \log n)$.

To define the induced dynamics, let $e_t'$ be the edge chosen u.a.r. from $E_{n+1}$ at time $t$ by $\{X_t'\}$, and let $e_t$ be the corresponding edge in $\Lambda$ if there is one. $X_{t+1}$ is obtained from $X_t$ as follows:

(i) if both endpoints of $e_t'$ are in $\partial\Lambda_{n+1}$, then $X_{t+1} = X_t$;

(ii) else if $X_{t+1}' = X_{t+1}' \cup \{e_t'\}$, then $X_{t+1} = X_t \setminus \{e_t\}$;

(iii) else if $X_{t+1}' = X_{t+1}' \setminus \{e_t'\}$, then $X_{t+1} = X_t \cup \{e_t\}$.

The initial configuration $X_0$ is the unique configuration compatible with $X_0'$.

We show first that $\{X_t\}$ is in fact a lazy version of the Glauber dynamics on $\Lambda$. To see this, note that $X_{t+1} = X_t$ whenever both endpoints of $e_t'$ are in $\partial\Lambda_{n+1}$. Otherwise, it is straightforward to check that $e_t \in X_t$ is a cut edge iff $e_t' \in X_t'$ is not a cut edge. Hence, $X_{t+1} = X_t \cup \{e_t\}$ iff

$X'_{t+1} = X_t \setminus \{e'_t\}$ and thus $X_{t+1} = X_t \cup \{e_t\}$ with probability:

$$\begin{cases} 1 - \frac{p^*}{q(1-p^*)+p^*} = p & \text{if } e'_t \text{ is a cut edge;} \\ \\ 1 - p^* = \frac{p}{q(1-p)+p} & \text{otherwise.} \end{cases}$$

This implies that $\{X_t\}$ does not move with probability $\Theta(n^{-1})$, and otherwise evolves exactly like the Glauber dynamics on $\Lambda$. Hence, it is sufficient for us to establish the mixing time of $\{X_t\}$. To do this, we show that the mixing times of $\{X_t\}$ and $\{X'_t\}$ are essentially the same.

Let $\Omega_{\mathrm{RC}}$ be the set of random-cluster configurations on $\Lambda$, and let $C(A)$ be the set of configurations compatible with a configuration $A$ on $\Lambda$. The first observation is that when $\{X'_t\}$ mixes, so does $\{X_t\}$. This follows from:

$$\begin{aligned} \|\, X_t(\cdot) - \mu_{\Lambda,p,q}(\cdot)\,\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{A\in\Omega_{\mathrm{RC}}} |X_t(A) - \mu_{\Lambda,p,q}(A)| \\ &= \frac{1}{2}\sum_{A\in\Omega_{\mathrm{RC}}} |X_t(A) - \mu_{\Lambda^*,p^*,q}(A^*)| \\ &\le \frac{1}{2}\sum_{A\in\Omega_{\mathrm{RC}}}\sum_{A'\in C(A)} \left|X'_t(A') - \mu^1_{\Lambda',p^*,q}(A')\right| \\ &= \|\, X'_t(\cdot) - \mu^1_{\Lambda',p^*,q}(\cdot)\,\|_{\mathrm{TV}}, \end{aligned}$$

where in the first and last equality we use the definition of total variation distance, in the second equality we use planar duality, and the third inequality follows from the triangle inequality and the correspondence between the configurations of $\Lambda$ and $\Lambda'$. Hence, by the results in Section 4.4, the mixing time of $\{X_t\}$ is $O(n^2 \log n)$.

We show next that the mixing time of $\{X_t\}$ is $\Omega(n^2 \log n)$. For this, note that in Theorem 4.7 we showed that there is an initial distribution $\pi_0$ for $X'_0$, defined in (4.13), such that

$$\|X'_T(\cdot) - \mu^1_{\Lambda',p^*,q}(\cdot)\|_{\mathrm{TV}} > 1/4$$

for some $T = \Omega(n^2 \log n)$. We will prove that when $X'_0$ is sampled from $\pi_0$, then

$$\|X'_t(\cdot) - \mu^1_{\Lambda',p^*,q}(\cdot)\|_{\mathrm{TV}} = \|X_t(\cdot) - \mu_{\Lambda,p,q}(\cdot)\|_{\mathrm{TV}}$$

for all $t \ge 0$. To show this we introduce some additional notation.

Let $\Lambda'' := (\Lambda_{n+1}, E_{n+1} \setminus \partial E_{n+1})$, where $\partial E_{n+1} \subseteq E_{n+1}$ is the set of edges with both endpoints in $\partial\Lambda_{n+1}$. Also, for any random-cluster configuration $A'$ on $\Lambda'$, we use $\partial A'$ to denote the random-cluster configuration induced in $\partial\Lambda'$ by $A'$.

Under the wired boundary condition, we have that for any $e \in \partial\Lambda'$, $\mu^1_{\Lambda',p^*,q}(e = 1) = p^*$. Hence, $\mu^1_{\Lambda',p^*,q}$ is the product measure of the distributions $\mu^1_{\Lambda'',p^*,q}$ and $\mu_{\partial\Lambda',p^*,1}$; the latter is the distribution on $\partial\Lambda'$ where every edge is sampled independently with probability $p^*$. Thus we have

$$\mu^1_{\Lambda',p^*,q}(A') = \mu^1_{\Lambda'',p^*,q}(A' \setminus \partial A') \cdot \mu_{\partial\Lambda',p^*,1}(\partial A').$$

By the correspondence between the configurations of $\Lambda^*$ and $\Lambda'$, we have that $\mu^1_{\Lambda'',p^*,q}(A' \setminus \partial A') = \mu_{\Lambda^*,p^*,q}(A^*)$. (As in Section 4.1, $A^*$ denotes the dual of the unique configuration $A$ compatible with $A'$.) Moreover, by planar duality $\mu_{\Lambda^*,p^*,q}(A^*) = \mu_{\Lambda,p,q}(A)$, and thus

$$\mu^1_{\Lambda',p^*,q}(A') = \mu_{\Lambda,p,q}(A) \cdot \mu_{\partial\Lambda',p^*,1}(\partial A'). \tag{4.17}$$

Also, under the wired boundary condition the configuration on the boundary of $X'_0$ is sampled according to $\mu_{\partial\Lambda',p^*,1}$. Hence, the distribution on the boundary of $X'_t$ has law $\mu_{\partial\Lambda',p^*,1}$ for all $t \geq 0$. Thus,

$$X'_t(A') = \mu_{\partial\Lambda',p^*,1}(\partial A') \cdot X'_t(A' \setminus \partial A') = \mu_{\partial\Lambda',p^*,1}(\partial A') \cdot X_t(A). \tag{4.18}$$

Hence,

$$\begin{aligned}
\| X'_t(\cdot) - \mu^1_{\Lambda',p^*,q}(\cdot) \|_{\mathrm{TV}} &= \frac{1}{2} \sum_{A\in\Omega_{\mathrm{RC}}} \sum_{A'\in C(A)} \left| X'_t(A') - \mu^1_{\Lambda',p^*,q}(A') \right| \\
&= \frac{1}{2} \sum_{A\in\Omega_{\mathrm{RC}}} \sum_{A'\in C(A)} \mu_{\partial\Lambda',p^*,1}(\partial A') \left| X_t(A) - \mu_{\Lambda,p,q}(A) \right| \\
&= \| X_t(\cdot) - \mu_{\Lambda,p,q}(\cdot) \|_{\mathrm{TV}},
\end{aligned}$$

where in the first and last equality we used the definition of total variation distance and the second follows from (4.17) and (4.18).

The results in Section 4.5 then imply that the mixing time of $\{X_t\}$ is $\Omega(n^2 \log n)$. Consequently, the Glauber dynamics on $\Lambda$ with $p > p_c(q)$ mixes in $\Theta(n^2 \log n)$ steps, as desired. □

# A proof of Lemma 4.8

We show first that the measure that results from conditioning on the state of a single edge maintains the exponential decay of finite volume connectivities (4.2).

**Fact 4.11.** *Let $p < p_c(q)$, $q \geq 1$, and let $\eta$ be a boundary condition for $\Lambda = (\Lambda_n, E_n)$. Consider a copy $\{Y_t\}$ of the continuous time Glauber dynamics on $\Lambda$, and assume $Y_0$ is sampled from the distribution $\mu^\eta_{\Lambda,p,q}(\cdot \mid e = b)$, for some $e \in E_n$ and $b \in \{0,1\}$. Then, for all $u, v \in \Lambda_n$, there exists positive constant $C$ and $\lambda$ such that*

$$\Pr[u \overset{Y_t}{\leftrightarrow} v] \leq C e^{-\lambda\, d(u,v)},$$

*where $u \overset{Y_t}{\leftrightarrow} v$ denotes the event that $u$ and $v$ are connected by a path of open edges in $Y_t$.*

*Proof.* Let $\{Z_t\}$ be a second instance of the continuous time Glauber dynamics. The evolution of $\{Z_t\}$ is coupled with that of $\{Y_t\}$ via the identity coupling, except that $\{Z_t\}$ never updates the edge $e$. The initial configuration of $\{Z_t\}$ is sampled according to the distribution $\mu^\eta(\,\cdot\,|\, e = 1)$ such that $Y_0 \subseteq Z_0$. This is always possible because $\mu^\eta(\,\cdot\,|\, e = 1) \geq \mu^\eta$. Then, $Y_t \subseteq Z_t$ and $Z_t$ has law $\mu^\eta(\,\cdot\,|\, e = 1)$ for all $t \geq 0$. We establish that the measure $\mu^\eta(\,\cdot\,|\, e = 1)$ has exponential decay of finite volume connectivities and thus so does the distribution of $Y_t$ for all $t \geq 0$. By (4.2), for all $u, v \in \Lambda_n$, we have

$$\mu^\eta(\, u \leftrightarrow v \,|\, e = 1)\mu^\eta(e = 1) \leq \mu^\eta(u \leftrightarrow v) \leq C\mathrm{e}^{-\lambda\, d(u,v)},$$

where $C, \lambda$ are positive constants. If $p' = \frac{p}{q(1-p)+p}$, then $\mu^\eta \geq \mu^\eta_{\Lambda,p',1}$ (see, e.g., [19]), and thus $\mu^\eta(e = 1) \geq p'$. Since $q \geq 1$,

$$\mu^\eta(\, u \leftrightarrow v \,|\, e = 1) \leq \frac{qC}{p}\mathrm{e}^{-\lambda\, d(u,v)}.$$

The result then follows immediately when $p = \Omega(1)$. Otherwise, the measure $\mu^\eta$ is stochastically dominated by any random-cluster measure $\mu^\eta_{\Lambda,p'',q}$ with $p'' = \Omega(1)$, for which we just established exponential decay of finite volume connectivities; the result follows by monotonicity. □

We are now ready to prove the lemma.

*Proof.* Let $Q_t$ be the random time at which the $t$-th edge is updated by the identity coupling. For some fixed $\ell$ to be chosen later, and each $t \geq 0$, consider the event

$$\mathcal{E}_{\ell,t} := \{u \overset{Y_{Q_t}}{\not\leftrightarrow} v \quad \forall u, v \in B \text{ s.t. } d(u, v) > \ell\},$$

where $u \overset{Y_{Q_t}}{\not\leftrightarrow} v$ denotes the event that $u$ and $v$ are not connected by a path in $Y_{Q_t}(B)$. Also, let $\mathcal{E}_\ell := \bigcap_{t:Q_t \leq T} \mathcal{E}_{\ell,t}$. Then,

$$\Pr[\, X_T(e) \neq Y_T(e)\,] \leq \Pr[\, X_T(e) \neq Y_T(e) \,|\, \mathcal{E}_\ell\,] + \Pr[\,\neg\mathcal{E}_\ell\,] \tag{4.19}$$

(cf. equation (4.4)). We bound each term on the right hand side of (4.19) separately.

Conditioned on the event $\mathcal{E}_\ell$, a witness for the fact that $X_T(e) \neq Y_T(e)$ can be constructed as in discrete time. However, the probability that a given witness of length $L$ is updated by the continuous time dynamics is instead bounded using the following fact from [28].

**Fact 4.12.** *Consider $L$ independent rate 1 Poisson clocks. Then, the probability that there is an increasing sequence of times $0 \leq t_1 < \ldots < t_L \leq T$ such that clock $i$ rings at time $t_i$ is at most $\left(\frac{eT}{L}\right)^L$.*

Recall from Section 4.2 that the number of witnesses of length $L$ is at most $(4(\ell + 1)^2)^L$ (crudely). Hence, following the same steps as in the proof Lemma 4.1, and taking $\ell = r^{1/4} - 1$, we get

$$\Pr[\, X_T(e) \neq Y_T(e) \,|\, \mathcal{E}_\ell\,] \leq \frac{\mathrm{e}}{\mathrm{e} - 1} \cdot \mathrm{e}^{-r^{3/4}}, \tag{4.20}$$

using the fact that $T \le r^{1/4}/(4\mathrm{e}^2)$ (cf. equation (4.6)).

To bound the second term on the right hand side of (4.19), let $N$ be the number of edge updates in $B$ up to time $T$. Observe that $N$ is a Poisson random variable with rate $M := |E(B,r)| = \Theta(r^2)$. Using standard bounds for Poisson tail probabilities we get that $\Pr[N > \mathrm{e}^2 MT] = \exp(-\Omega(MT))$ for all $T \ge 1$. Therefore,

$$\Pr[\,\neg\mathcal{E}_\ell\,] \le \Pr[\,\neg\mathcal{E}_\ell \mid N \le \mathrm{e}^2 MT\,] + \mathrm{e}^{-\Omega(MT)}.$$

Also, $\neg\mathcal{E}_\ell := \bigcup_{t:Q_t \le T} \neg\mathcal{E}_{\ell,t}$, and if the edge update at time $Q_t$ occurs outside $B$, we have $\neg\mathcal{E}_{\ell,t} = \neg\mathcal{E}_{\ell,t+1}$. Hence, a union bound implies

$$\Pr[\,\neg\mathcal{E}_\ell\,] \;\le\; \mathrm{e}^2 MT \max_{t:Q_t \le T} \Pr[\,\neg\mathcal{E}_{\ell,t}\,] + \mathrm{e}^{-\Omega(MT)}.$$

Fact 4.11 establishes exponential decay of finite volume connectivities (4.2) for the distribution of $Y_t$ in $B$ for all $t \ge 0$. Then, as in Lemma 4.1, we obtain

$$\Pr[\,\neg\mathcal{E}_\ell\,] \le O(r^{6.25}) \cdot \mathrm{e}^{-\Omega(r^{1/4})} + \mathrm{e}^{-\Omega(r^2)}.$$

Together with (4.20), this implies there exist constants $c, C, \lambda > 0$ such that for all $r \ge c$ we have $\Pr[\,X_T(e) \ne Y_T(e)\,] \le C\exp(-\lambda r^{1/4})$, as desired. $\square$

# Bibliography


[1] K.S. Alexander. “Mixing properties and exponential decay for lattice systems in finite volumes”. In: *The Annals of Probability* 32.1A (2004), pp. 441–487.

[2] K.S. Alexander. “On weak mixing in lattice models”. In: *Probability Theory and Related Fields* 110.4 (1998), pp. 441–471.

[3] V. Beffara and H. Duminil-Copin. “The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$”. In: *Probability Theory and Related Fields* 153 (3-4 2012), pp. 511–542.

[4] J. van den Berg. “A uniqueness condition for Gibbs measures, with applications to the 2-dimensional Ising antiferromagnet”. In: *Communications in Mathematical Physics* 152.1 (1993), pp. 161–166.

[5] A. Blanca and A. Sinclair. “Dynamics for the mean-field random-cluster model”. In: *Proceedings of the 19th International Workshop on Randomization and Computation* (2015), pp. 528–543.

[6] A. Blanca and A. Sinclair. “Random-Cluster Dynamics in $\mathbb{Z}^2$”. In: *Probability Theory and Related Fields, To appear. Extended abstract appeared in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016)* (2016), pp. 498–513.

[7] B. Bollobás, G.R. Grimmett, and S. Janson. “The random-cluster model on the complete graph”. In: *Probability Theory and Related Fields* 104.3 (1996), pp. 283–317. ISSN: 0178-8051.

[8] C. Borgs, J. Chayes, and P. Tetali. “Swendsen-Wang algorithm at the Potts transition point”. In: *Probability Theory and Related Fields* 152 (2012), pp. 509–557.

[9] C. Borgs et al. “Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics”. In: *Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS)* (1999), pp. 218–229.

[10] L. Chayes and J. Machta. “Graphical representations and cluster algorithms II”. In: *Physica A* 254 (1998), pp. 477–516.

[11] A. Coja-Oghlan, C. Moore, and V. Sanwalani. “Counting connected graphs and hypergraphs via the probabilistic method”. In: *Random Structures & Algorithms* 31.3 (2007), pp. 288 –329.

[12] C. Cooper et al. “Mixing Properties of the Swendsen-Wang Process on the Complete Graph and Narrow Grids”. In: *Journal of Mathematical Physics* 41 (2000), pp. 1499–1527.

[13] P. Cuff et al. “Glauber Dynamics for the Mean-Field Potts Model”. In: *Journal of Statistical Physics* 149.3 (2012), pp. 432–477.

[14] H. Duminil-Copin, V. Sidoravicius, and V. Tassion. “Continuity of the phase transition for planar random-cluster and Potts models with $1 \leq q \leq 4$”. In: *Preprint* (2015).

[15] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.

[16] M. Dyer et al. “Mixing in Time and Space for Lattice Spin Systems: A Combinatorial View”. In: *Random Structures & Algorithms* 24 (2004), pp. 461–479.

[17] R.G. Edwards and A.D. Sokal. “Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm”. In: *Physical Review D* 38.6 (1988), pp. 2009–2012.

[18] W. Feller. *An introduction to probability theory and its applications, Vol. 2*. 3rd. Wiley Mathematical Statistics Series. Wiley, 1966.

[19] C.M. Fortuin. “On the random-cluster model: III. The simple random-cluster model”. In: *Physica* 59.4 (1972), pp. 545–570.

[20] C.M. Fortuin and P.W. Kasteleyn. “On the random-cluster model I. Introduction and relation to other models”. In: *Physica* 57.4 (1972), pp. 536–564.

[21] A. Galanis, D. Štefankovič, and E. Vigoda. “Swendsen-Wang algorithm on the Mean-Field Potts Model”. In: *Proceedings of the 19th International Workshop on Randomization and Computation* (2015), pp. 815–828.

[22] T. Garoni. Personal communication. 2015.

[23] Q. Ge and D. Štefankovič. “A graph polynomial for independent sets of bipartite graphs”. In: *Combinatorics, Probability and Computing* 21.5 (2012), pp. 695–714.

[24] Reza Gheissari and Eyal Lubetzky. “Mixing times of critical 2D Potts models”. In: *arXiv preprint arXiv:1607.02182* (2016).

[25] V.K. Gore and M.R. Jerrum. “The Swendsen-Wang Process Does Not Always Mix Rapidly”. In: *Journal of Statistical Physics* 97.1-2 (1999), pp. 67–86. ISSN: 0022-4715.

[26] G.R. Grimmett. *The Random-Cluster Model*. Vol. 333. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Berlin: Springer-Verlag, 2006.

[27] Heng Guo and Mark Jerrum. “Random cluster dynamics for the Ising model is rapidly mixing”. In: *arXiv preprint arXiv:1605.00139* (2016).

[28] T.P. Hayes and A. Sinclair. “A general lower bound for mixing of single-site dynamics on graphs”. In: *Annals of Applied Probability* 17.3 (2007), pp. 931–952.

[29] Remco Van Der Hofstad. *Random graphs and complex networks.* 2008.

[30] M. Huber. “A bounding chain for Swendsen-Wang.” In: *Random Structures & Algorithms* 22.1 (2003), pp. 43–59.

[31] S. Janson. Personal communication. 2012.

[32] S. Janson and M. Luczak. “Susceptibility in subcritical random graphs”. In: *Journal of Mathematical Physics* 49.12 (2008), p. 125207.

[33] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs.* Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[34] E. Kreyszig. *Introductory Functional Analysis With Applications.* Wiley Classics Library. John Wiley & Sons, 1978.

[35] L. Laanait et al. “Interfaces in the Potts model I: Pirogov-Sinai theory of the Fortuin-Kasteleyn representation”. In: *Communications in Mathematical Physics* 140.1 (1991), pp. 81–91.

[36] D.A. Levin, M.J. Luczak, and Y. Peres. “Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability”. In: *Probability Theory and Related Fields* 146.1 (2010), pp. 223–265.

[37] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Society, 2008.

[38] Y. Long et al. “A power law of order 1/4 for critical mean-field Swendsen-Wang dynamics”. In: *Memoirs of the American Mathematical Society* 232.1092 (2011).

[39] E. Lubetzky and A. Sly. “Critical Ising on the square lattice mixes in polynomial time”. In: *Communications in Mathematical Physics* 313.3 (2012), pp. 815–836.

[40] M.J. Luczak and T. Łuczak. “The phase transition in the cluster-scaled model of a random graph”. In: *Random Structures & Algorithms* 28.2 (2006), pp. 215–246.

[41] F. Martinelli. *Lectures on Glauber dynamics for discrete spin models.* Vol. 1717. Springer Lecture Notes in Mathematics. Springer Verlag, 1999, pp. 93–191.

[42] F. Martinelli, E. Olivieri, and R.H. Schonmann. “For 2-d lattice spin systems weak mixing implies strong mixing”. In: *Communications in Mathematical Physics* 165.1 (1994), pp. 33–47.

[43] C. McDiarmid. “Concentration”. In: *Probabilistic Methods for Algorithmic Discrete Mathematics* 16 (1998), pp. 195–248.

[44] E. Mossel and A. Sly. “Exact thresholds for Ising–Gibbs samplers on general graphs”. In: *The Annals of Probability* 41.1 (2013), pp. 294–328.

[45] J. Propp and D. Wilson. “Exact sampling with coupled Markov chains and applications to statistical mechanics”. In: *Random Structures & Algorithms* 9 (1996), pp. 223–252.

[46] L. Saloff-Coste. *Lectures on finite Markov chains.* Ed. by Pierre Bernard. Vol. 1665. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1997, pp. 301–413.

[47] V. Strassen. "The existence of probability measures with given marginals". In: *The Annals of Mathematical Statistics* 36 (1965), pp. 423–439.

[48] R.H. Swendsen and J.S. Wang. "Nonuniversal critical dynamics in Monte Carlo simulations". In: *Physical Review Letters* 58 (2 1987), pp. 86–88.

[49] M. Ullrich. "Comparison of Swendsen-Wang and heat-bath dynamics". In: *Random Structures and Algorithms* 42.4 (2013), pp. 520–535.

[50] M. Ullrich. "Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions". In: *Dissertationes Mathematicae* 502 (2014), p. 64.

[51] M. Ullrich. "Swendsen-Wang Is Faster than single-bond Dynamics". In: *SIAM Journal on Discrete Mathematics* 28.1 (2014), pp. 37–48.