



# Mixing time of vertex-weighted exponential random graphs

Ryan DeMuse, Terry Easlick, Mei Yin \*

Department of Mathematics, University of Denver, Denver, CO 80208, USA



## ARTICLE INFO

### Article history:

Received 1 February 2018

Received in revised form 12 June 2018

### Keywords:

Exponential random graphs

Mixing time

Glauber dynamics

## ABSTRACT

Exponential random graph models have become increasingly important in the study of modern networks ranging from social networks, economic networks, to biological networks. They seek to capture a wide variety of common network tendencies such as connectivity and reciprocity through local graph properties. Sampling from these exponential distributions is crucial for parameter estimation, hypothesis testing, as well as understanding the features of the network in question. We inspect the efficiency of a popular sampling technique, the Glauber dynamics, for vertex-weighted exponential random graphs. Letting  $n$  be the number of vertices in the graph, we identify a region in the parameter space where the mixing time for the Glauber dynamics is  $\Theta(n \log n)$  (the high-temperature phase) and a complementary region where the mixing time is exponentially slow of order  $e^{\Omega(n)}$  (the low-temperature phase). Lastly, we give evidence that along a critical curve in the parameter space the mixing time is  $O(n^{2/3})$ .

© 2018 Elsevier B.V. All rights reserved.

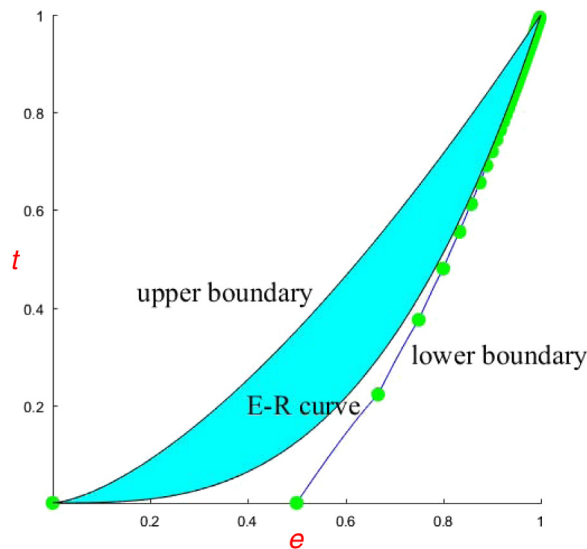
## 1. Introduction

Exponential random graph models are powerful tools in the study of modern networks ranging from social networks, economic networks, to biological networks. By representing the complex global structure of a large network through tractable local properties, these models seek to capture a wide variety of common network tendencies. See for example [1–4], and the references therein. Despite their flexibility, conventionally used exponential random graphs suffer from some deficiencies that may hamper their utility to researchers [5,6]. Consider the dynamics of spreading events in a complex network. There are sensitive control points collectively known as “influential spreaders”, whose infection maximizes the overall fraction of infected vertices. For information diffusion over Twitter for example, the influential spreaders may be thought of as a celebrity or a news source. Vertices in the network thus carry with themselves some distinguishing features, a phenomenon that could not be directly modeled by standard exponential random graphs since their underlying probability space consists of simple graphs only. By placing weights on the vertices, the current work addresses this limitation of the exponential model.

Before proceeding further, we provide another reason why the vertex-weighted model may be of interest [7]. Let  $\mathcal{G}_n$  be the set of all vertex-weighted labeled graphs  $G_n$  on  $n$  vertices. Assume that the vertex weights are iid real random variables subject to a common distribution  $\nu$  supported on  $[0, 1]$ , the edge weight between two vertices is a product of the vertex weights, and the triangle weight among three vertices is a product of the edge weights. Let  $U$  be a random variable distributed according to  $\nu$  and denote its expectation with respect to  $\nu$  by  $\mathbb{E}$ . Further denote the expected edge weight of  $G_n \in \mathcal{G}_n$  by  $e$  and the expected triangle weight by  $t$ . Then we have  $e = \mathbb{E}(U)^2$  and  $t = (\mathbb{E}(U^2))^3$ . Note that by suitably choosing  $\nu$ , the entire

\* Corresponding author.

E-mail addresses: [ryan.demuse@du.edu](mailto:ryan.demuse@du.edu) (R. DeMuse), [terry.easlick@du.edu](mailto:terry.easlick@du.edu) (T. Easlick), [mei.yin@du.edu](mailto:mei.yin@du.edu) (M. Yin).



**Fig. 1.** The cyan region shows where the expected edge and triangle densities lie for a vertex-weighted random graph model.

region between the upper boundary of the realizable edge-triangle densities and the Erdős–Rényi curve ( $e^3 \leq t \leq e^{3/2}$ ) may be attained for vertex-weighted random graphs. See Fig. 1. If we take  $U$  to be Bernoulli, the upper boundary is reproduced:

$$\mathbb{E}(U^2) = \mathbb{E}(U) \implies t = e^{3/2}. \quad (1)$$

If we take  $U$  to be a constant a.s., the Erdős–Rényi curve is recovered:

$$\mathbb{E}(U^2) = \mathbb{E}(U)^2 \implies t = e^3. \quad (2)$$

By contrast, since simple graphs may be interpreted as having iid Bernoulli(.5) weights on the edges, the underlying graph space of standard exponential random graphs lies at a single point  $(1/2, 1/8)$  on the Erdős–Rényi curve. Even without incorporating the exponential construction, assigning vertex weights alone adds intriguing characteristics to the model.

### 1.1. The model

In this paper we will restrict our attention to vertex-weighted exponential random graph models where the vertex weights take values in 0 and 1 only. We include in the exponent a combination of edge and triangle densities, both with non-negative parameters. Even under this simplification, the vertex-weighted model depicts captivating behaviors in large-scale networks. Instead of Erdős–Rényi, it emphasizes the formation of cliques, and is particularly suited for the modeling of a broad range of social networks. Consider the Facebook friend graph for example, where we make the simplifying assumption that a person is either interested (vertex value 1) or not (vertex value 0) in building a friendship. Then having value 1 at three distinct vertices will force the formation of a triangle rather than a two-star, which is in analogy with the common conception that a friend of a friend is more likely to be a friend. An added benefit of this setting is that the model may be considered as an extension of the lattice gas (Ising) model on a graph, and the techniques of spin models may be employed in our investigation [8].

A graph  $G_n \in \mathcal{G}_n$  may be viewed as an element of  $X \in \mathcal{X} := \{0, 1\}^n$ , referred to as “configurations”, that attributes weights 0 or 1 to the ordered vertices of  $G_n$ . Denote by  $X(i)$  the weight of vertex  $i$ . Borrowing terminology from spin models, the vertex weight  $X(i)$  will be called the “spin” at  $i$ . By iid-ness, the spins at different vertices are independent, and subject to a common distribution  $\nu$  that assumes value 0 with probability  $1 - p$  and 1 with probability  $p$  for some  $p \in (0, 1)$ . Let  $H_1$  be the number of edges for the configuration  $X$  and  $H_2$  be the number of triangles. They may be formulated explicitly in terms of lattice gas (Ising) spins:

$$H_1(X) = \sum_{i \neq j} X(i)X(j) \text{ and } H_2(X) = \sum_{i \neq j \neq k} X(i)X(j)X(k), \quad (3)$$

where the inequality  $n_1 \neq n_2 \neq \dots \neq n_k$  means that  $n_i \neq n_j$  for any  $i \neq j$ . We rescale the edge and triangle parameters in the exponent,  $H = (\alpha_1/n)H_1 + (\alpha_2/n^2)H_2$ , so that the total contribution of a single vertex to the weights is  $O(1)$ . We are now ready to introduce a Gibbs distribution on the set of spin configurations  $\mathcal{X}$ . To avoid cumbersome notation, we suppress the  $n$ -dependence in many of the quantities under discussion.

**Definition 1.** Take  $\alpha_1 \geq 0, \alpha_2 \geq 0$ , and  $p \in (0, 1)$ . Let  $X \in \mathcal{X}$  be a spin configuration. Denote by  $\omega(X)$  the number of vertices with spin 1 in  $X$ . Assign a Gibbs probability measure on  $\mathcal{X}$  as

$$\pi(X) = Z^{-1} \exp(H(X)) p^{\omega(X)} (1-p)^{n-\omega(X)}, \quad (4)$$

where  $H = (\alpha_1/n)H_1 + (\alpha_2/n^2)H_2$  is the combination of edge and triangle weights and  $Z = Z(n, p, \alpha_1, \alpha_2)$  is the normalizing constant (partition function).

The configuration space  $\mathcal{X}$  can be partially ordered in the sense that for  $X, Y \in \mathcal{X}$ , we say that  $X \leq Y$  if and only if  $X(i) \leq Y(i)$  for every  $i \in \{1, \dots, n\}$ . To model the evolution of the network towards equilibrium, we will adopt (single-site) Glauber dynamics, which is a discrete-time irreducible and aperiodic Markov chain  $(X_t)_{t=0}^\infty$  on  $\mathcal{X}$ . Under the Glauber dynamics, the random graph evolves by selecting a vertex  $i$  and updating the spin  $X(i)$  according to  $\pi$  conditioned to agree with the spins at all vertices not equal to  $i$ . By sampling from the exponential distribution using Glauber dynamics, we learn the global structure of the network as well as parameters describing the interactions. Explicitly, let  $X \in \mathcal{X}$  be a configuration and set the initial state  $X_0 = X$ . The next step of the Markov chain,  $X_1$ , is obtained as follows. Choose a vertex  $i$  uniformly from  $\{1, \dots, n\}$ . Let  $X_1(j) = X(j)$  for all  $j \neq i$ ,  $X_1(i) = 1$  with probability  $P_+$  and  $X_1(i) = 0$  with probability  $P_-$ , where the update probabilities  $P_+$  and  $P_-$  are given by

$$P_+(X, i) = \frac{p \exp(H'(X, i))}{p \exp(H'(X, i)) + (1-p)} \quad (5)$$

and

$$P_-(X, i) = \frac{1-p}{p \exp(H'(X, i)) + (1-p)}. \quad (6)$$

Here  $H' = (\alpha_1/n)S + (\alpha_2/n^2)T$  depends only on the spins at vertices other than  $i$ , with

$$S(X, i) = \sum_{j \neq i} X(j) \text{ and } T(X, i) = \sum_{j \neq k \neq i} X(j)X(k) = \frac{S(X, i)(S(X, i) - 1)}{2}. \quad (7)$$

For  $X, Y \in \mathcal{X}$ , the transition matrix for the Glauber dynamics is then

$$P(X, Y) = \frac{1}{n} \sum_i \frac{f(Y(i)) e^{Y(i)H'(X, i)}}{f(Y(i)) e^{Y(i)H'(X, i)} + f(1-Y(i)) e^{(1-Y(i))H'(X, i)}} \mathbf{1}_{\left\{ \begin{smallmatrix} Y(j)=X(j) \\ j \neq i \end{smallmatrix} \right\}}, \quad (8)$$

where  $Y(i) \in \{0, 1\}$  and we define  $f$  such that  $f(0) = 1-p$  and  $f(1) = p$  to lighten the notation.

### 1.2. Mixing time

The Gibbs distribution  $\pi$  is stationary and reversible for the Glauber dynamics chain. By the convergence theorem for ergodic Markov chains, the Glauber dynamics will converge to the stationary distribution and our goal is to obtain some estimates on the mixing time, since it greatly affects the efficiency of simulation studies and sampling algorithms [9,10]. Given  $\varepsilon > 0$ , the mixing time for this Markov chain is defined as

$$t_{\text{mix}}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}, \quad (9)$$

where

$$d(t) = \max_{X \in \mathcal{X}} \|P^t(X, \cdot) - \pi\|_{\text{TV}} \quad (10)$$

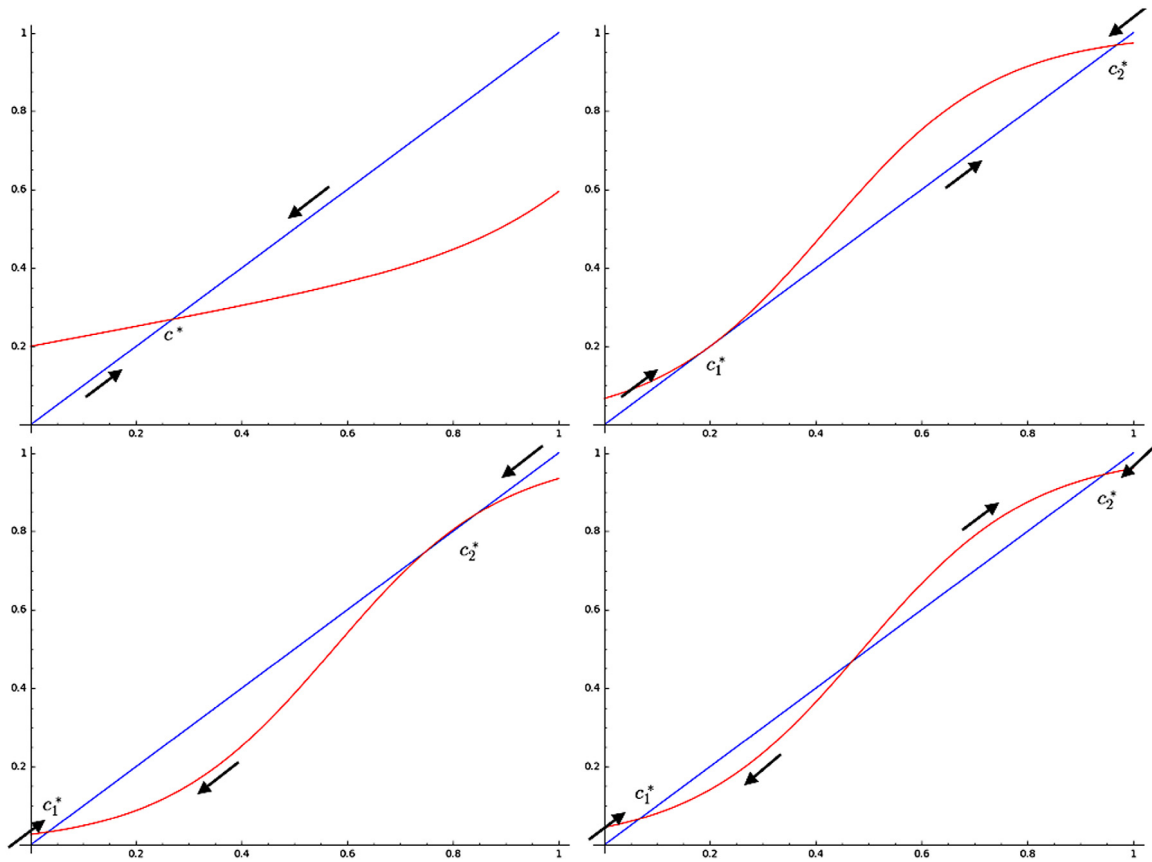
measures the total variation distance to stationarity of the Glauber dynamics chain after  $t$  steps. As is standard, we take  $t_{\text{mix}} := t_{\text{mix}}(1/4)$ . The mixing time is thus defined to be the minimum number of discrete time steps such that, starting from an arbitrary configuration  $X$ , the chain is within total variation distance  $1/4$  from the stationary distribution  $\pi$ . For background on mixing times, see [11,12]. Our results will indicate that the mixing time can vary enormously depending on the choice of parameters.

### 1.3. Normalized magnetization

Given a spin configuration  $X \in \mathcal{X}$ , the normalized magnetization  $c$  of  $X$  is defined as

$$c(X) = \frac{1}{n} \sum_{i=1}^n X(i). \quad (11)$$

Adopting (single-site) Glauber dynamics on  $\mathcal{X}$ , the normalized magnetization chain  $(c_t)_{t=0}^\infty$  is a projection of the chain  $(X_t)_{t=0}^\infty$  and so is also aperiodic and irreducible. Set the initial state  $c_0 = c$ . From the mechanism described in Section 1.1, after one



**Fig. 2.** Behavior of the  $\lambda$  function in different regions of the parameter space with arrows indicating whether the fixed point is an attractor or a repeller.

Glauber update,  $c_1$  will take on one of three values:  $c - 1/n$ ,  $c$ , or  $c + 1/n$ . If a spin 0 vertex is chosen and updated to spin 1,  $c$  changes to  $c + 1/n$ . Alternatively, if a spin 1 vertex is chosen and updated to spin 0,  $c$  changes to  $c - 1/n$ . When no spins are updated,  $c$  stays the same. By (5), the probability that we select a spin 0 vertex and update it to spin 1 is

$$P_u = \frac{n - cn}{n} \frac{p \exp(\alpha_1 c + \frac{\alpha_2}{2} c(c - \frac{1}{n}))}{p \exp(\alpha_1 c + \frac{\alpha_2}{2} c(c - \frac{1}{n})) + (1 - p)}. \quad (12)$$

Similarly, by (6), the probability that we select a spin 1 vertex and update it to spin 0 is

$$P_d = \frac{cn}{n} \frac{1 - p}{p \exp(\alpha_1(c - \frac{1}{n}) + \frac{\alpha_2}{2}(c - \frac{2}{n})(c - \frac{1}{n})) + (1 - p)}. \quad (13)$$

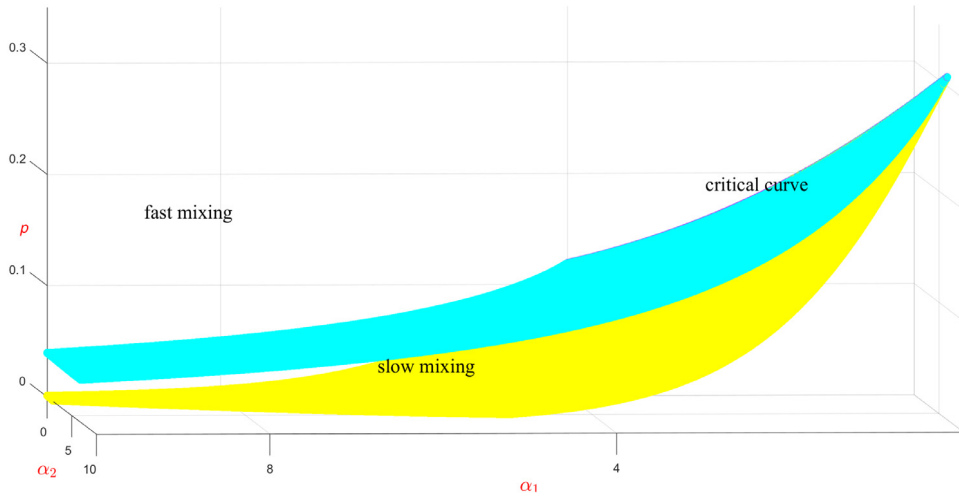
Combining (12) and (13), the magnetization  $c_t$  moves up with probability  $P_u$ , down with probability  $P_d$ , and remains unchanged with probability  $1 - P_u - P_d$ . For  $n$  large enough,  $P_u \asymp (1 - c)\lambda(c)$  and  $P_d \asymp c(1 - \lambda(c))$ , where

$$\lambda(c) = \frac{p \exp(\alpha_1 c + \frac{\alpha_2}{2} c^2)}{p \exp(\alpha_1 c + \frac{\alpha_2}{2} c^2) + (1 - p)} \quad (14)$$

represents the asymptotic probability that a chosen vertex is updated to spin 1. This implies that the expected magnetization drift is asymptotically  $(\lambda(c) - c)/n$ , and a rigorous justification may be found in Lemma 2.

#### 1.4. Phase classification

The magnetization chain  $(c_t)$  is a deciding factor in the convergence of the Glauber dynamics chain  $(X_t)$ . Note that  $0 \leq c_t \leq 1$  and  $\lambda$  is a smooth and increasing function on  $[0, 1]$ . Since  $\lambda(0) > 0$  and  $\lambda(1) < 1$ ,  $\lambda(c) = c$  admits at least one solution in  $(0, 1)$ . If the solution  $c$  is unique and not an inflection point, i.e.  $\lambda'(c) < 1$  (referred to as the “high-temperature phase”), then independent of the initial position all configurations will be driven towards it, and the burn-in stage will cost  $O(n)$  steps. See the upper left plot of Fig. 2. Conversely, if there exist at least two solutions  $c$  such that  $\lambda'(c) < 1$  (referred



**Fig. 3.** Surfaces in the parameter space illustrating the region with fast vs. slow mixing and identifying the critical curve. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to as the “low-temperature phase”), then the burn-in procedure will take the configurations to different attractor states depending on their initial positions. See the lower right plot of Fig. 2. Once the configuration is close to an attractor, the Glauber dynamics allows an exponentially small flow of probability for it to leave. A detailed examination of the burn-in period will be provided in Section 2.

In Section 3, through estimating the average distance after one update between two coupled configurations that agree everywhere except at a single vertex, we show that the Glauber dynamics  $X_t$  mixes in  $O(n \log n)$  steps in the high-temperature phase. Relating to coupon collecting and employing spectral methods, the same asymptotic lower bound  $\Omega(n \log n)$  is validated. While in Section 4, by a conductance argument using the Cheeger inequality, we establish exponentially slow mixing of the Glauber dynamics  $X_t$  in the low-temperature phase. Finally, in Section 5, we give evidence that the burn-in will cost  $O(n^{3/2})$  steps along the “critical curve”, and the Glauber dynamics  $X_t$  is thus expected to mix in  $O(n^{3/2})$  steps. See Fig. 3. The cyan and yellow surfaces separate the high and low-temperature phases, with their intersection marked by the critical curve. Convergence of the Glauber dynamics  $X_t$  elsewhere on the two surfaces corresponds to the situation where  $\lambda(c) = c$  has at least two solutions and one solution  $c$  satisfies  $\lambda'(c) = 1$ . See the upper and right and lower left plots of Fig. 2. The mixing time largely depends on the movement of the chain around the inflection point and is not addressed in this paper.

## 2. Burn-in period

We start by running the Glauber dynamics for an initial burn-in period. This will ensure that the associated magnetization chain is with high probability close to an attractor. Let  $X \in \mathcal{X}$  be any spin configuration. Set the initial state  $X_0 = X$  and let  $c_t$  be the normalized magnetization of  $X_t$  at time  $t$ . We use  $\mathbb{P}_X$  and  $\mathbb{E}_X$  respectively to denote the underlying probability measure and associated expectation. To keep the notation light, we omit the explicit dependence on  $X$  when it is clear from the context.

**Lemma 2.** *The expected drift in  $c_t$  after one step of the Glauber dynamics, starting from the configuration  $X$ , is given by*

$$\mathbb{E}(c_{t+1} - c_t \mid c_t) = \frac{1}{n} (\lambda(c_t) - c_t) + O\left(\frac{1}{n^2}\right), \quad (15)$$

where  $\lambda$  is defined as in (14).

**Proof.** From our discussion in Section 1.3, we compute

$$\begin{aligned} \mathbb{E}(c_{t+1} - c_t \mid c_t) &= \frac{1}{n} (1 - c_t) \frac{p \exp(\alpha_1 c_t + \frac{\alpha_2}{2} c_t (c_t - \frac{1}{n}))}{p \exp(\alpha_1 c_t + \frac{\alpha_2}{2} c_t (c_t - \frac{1}{n})) + (1 - p)} \\ &\quad - \frac{1}{n} c_t \frac{1 - p}{p \exp(\alpha_1 (c_t - \frac{1}{n}) + \frac{\alpha_2}{2} (c_t - \frac{1}{n}) (c_t - \frac{2}{n})) + (1 - p)}. \end{aligned} \quad (16)$$

Note that lower order fluctuations may be extracted from the exponents since

$$\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2} c_t \left(c_t - \frac{1}{n}\right)\right) = \exp\left(-\frac{\alpha_2}{2n} c_t\right) \exp\left(\alpha_1 c_t + \frac{\alpha_2}{2} (c_t)^2\right), \quad (17)$$

and

$$\begin{aligned} & \exp\left(\alpha_1 \left(c_t - \frac{1}{n}\right) + \frac{\alpha_2}{2} \left(c_t - \frac{1}{n}\right) \left(c_t - \frac{2}{n}\right)\right) \\ &= \exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n} c_t + \frac{\alpha_2}{n^2}\right) \exp\left(\alpha_1 c_t + \frac{\alpha_2}{2} (c_t)^2\right), \end{aligned} \quad (18)$$

which gives

$$\begin{aligned} \mathbb{E}(c_{t+1} - c_t \mid c_t) &= \frac{1}{n} (\lambda(c_t) - c_t) \\ &+ \frac{\lambda(c_t)}{n} (1 - c_t) \frac{\exp\left(-\frac{\alpha_2}{2n} c_t\right) - 1}{1 + \left(\frac{p}{1-p}\right) \exp\left(-\frac{\alpha_2}{2n} c_t\right) \exp\left(\alpha_1 c_t + \frac{\alpha_2}{2} (c_t)^2\right)} \\ &+ \frac{\lambda(c_t)}{n} c_t \frac{\exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n} c_t + \frac{\alpha_2}{n^2}\right) - 1}{1 + \left(\frac{p}{1-p}\right) \exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n} c_t + \frac{\alpha_2}{n^2}\right) \exp\left(\alpha_1 c_t + \frac{\alpha_2}{2} (c_t)^2\right)}. \end{aligned} \quad (19)$$

Following standard analytical arguments, the long fractional terms above are  $O(1/n)$ . The conclusion readily follows.  $\square$

Applying [Lemma 2](#), [Theorem 3](#) and [Corollaries 4](#) and [5](#) show that if the associated magnetization  $c_0$  of the initial configuration is significantly different from an attractor  $c^*$  but bounded away from any other solution of the fixed point equation  $\lambda(c) = c$ , then there is a drift of the Glauber dynamics towards a configuration whose normalized magnetization is closer to  $c^*$  than the starting state, i.e.  $c_t \rightarrow c^*$ . [Theorem 3](#) is proved when  $c_0 > c^*$ , and an analogous result holds for  $c_0 < c^*$  using a similar line of reasoning. See [Fig. 2](#) for an illustration of this burn-in procedure.

**Theorem 3.** Suppose  $\lambda(c^*) = c^*$  and  $\lambda'(c^*) < 1$ . Let  $\bar{c} > c^*$  be the smallest value satisfying  $\lambda(\bar{c}) = \bar{c}$ . (If no such  $\bar{c}$  exists, take  $\bar{c} = 1$ .) Let the initial magnetization be  $c_0$  with  $c^* + \mu < c_0 < \bar{c} - \mu$  for some  $\mu > 0$ . Then there exist  $\eta, d > 0$  depending only on  $\mu, p, \alpha_1, \alpha_2$  such that  $T = dn$  and

$$\mathbb{P}(c_T \leq c_0 - \eta) \geq 1 - e^{-\Omega(\sqrt{n})}. \quad (20)$$

**Proof.** Since  $\lambda'(c^*) < 1$ ,  $\lambda(c) - c < 0$  for  $c \in [c^* + \mu, \bar{c} - \mu]$ . By the extreme value theorem, the maximum of the smooth function  $\lambda(c) - c$  is attained for some value  $\underline{c}$  in the compact interval  $[c^* + \mu, \bar{c} - \mu]$ . Define  $\gamma > 0$  as  $\gamma = -(\lambda(\underline{c}) - \underline{c})/2$ . Choose  $\eta > 0$  so that  $[c_0 - 2\eta, c_0 + \eta] \subseteq [c^* + \mu, \bar{c} - \mu]$ . Let

$$D_t(\eta) = \{c_t : c_0 - 2\eta \leq c_t \leq c_0 + \eta\}. \quad (21)$$

By [Lemma 2](#), for  $c_t \in D_t(\eta)$  and  $n$  sufficiently large,  $\mathbb{E}(c_{t+1} - c_t \mid c_t) \leq -\gamma/n$ . Utilizing the negative drift  $-\gamma/n$  of the biased random walk  $c_t$  and employing a moment generating function method, we first show that “bad” magnetization, i.e.  $c_t > c_0 + \eta$  for some  $0 \leq t \leq T$ , occurs with exponentially small probability.

Define  $S_{t_1, t_2}$  as

$$S_{t_1, t_2} = \sum_{t=t_1+1}^{t_2} \left(c_t - c_{t-1} + \frac{\gamma}{2n}\right) \mathbf{1}_{D_{t-1}}. \quad (22)$$

The random variable  $S_{t_1, t_2}$  records the change in “good” magnetization  $c_t$  from  $t_1$  to  $t_2$ , shifted by  $\gamma/(2n)$  per time step. Let  $\mathcal{F}_t$  be the natural filtration. By the tower property of conditional expectation,

$$\mathbb{E}\left(e^{\theta S_{t_1, t_2}}\right) = \mathbb{E}\left(e^{\theta S_{t_1, t_2-1}} \mathbb{E}\left(e^{\theta(c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n}) \mathbf{1}_{D_{t_2-1}}} \mid \mathcal{F}_{t_2-1}\right)\right). \quad (23)$$

Using linearity of conditional expectation, we write

$$\begin{aligned} \mathbb{E}\left(e^{\theta(c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n}) \mathbf{1}_{D_{t_2-1}}} \mid \mathcal{F}_{t_2-1}\right) &= \sum_{k=0}^{\infty} \mathbb{E}\left(\frac{\theta^k (c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n})^k}{k!} \mathbf{1}_{D_{t_2-1}}^k \mid \mathcal{F}_{t_2-1}\right) \\ &= 1 + \mathbb{E}\left(\theta(c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n}) \mathbf{1}_{D_{t_2-1}} \mid \mathcal{F}_{t_2-1}\right) \\ &+ \sum_{k=2}^{\infty} \mathbb{E}\left(\frac{\theta^k (c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n})^k}{k!} \mathbf{1}_{D_{t_2-1}}^k \mid \mathcal{F}_{t_2-1}\right). \end{aligned} \quad (24)$$

Recall that  $\mathbb{E}(c_{t+1} - c_t \mid c_t) \leq -\gamma/n$  for  $c_t \in D_t$ , and so

$$\mathbb{E}\left(e^{\theta(c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n})} \mathbf{1}_{D_{t_2-1}} \mid \mathcal{F}_{t_2-1}\right) \leq 1 - \frac{\gamma\theta}{2n} \mathbf{1}_{D_{t_2-1}} + O\left(\frac{\theta^2}{n^2}\right). \quad (25)$$

Taking  $\theta = c\sqrt{n}$  for a sufficiently small constant  $c$ , the above conditional expectation is less than 1. Iterating this procedure gives

$$\mathbb{E}(e^{\theta S_{t_1, t_2}}) \leq \mathbb{E}(e^{\theta S_{t_1, t_2-1}}) \leq \dots \leq \mathbb{E}(e^{\theta S_{t_1, t_1}}) = 1. \quad (26)$$

By the Chernoff bound,

$$\mathbb{P}\left(S_{t_1, t_2} \geq \frac{\eta}{2}\right) \leq \frac{\mathbb{E}(e^{\theta S_{t_1, t_2}})}{e^{\theta \frac{\eta}{2}}} = e^{-\Omega(\sqrt{n})}. \quad (27)$$

Consider the set

$$B_{t_1, t_2}(\eta) = \left(\bigcap_{t_1 \leq t < t_2} D_t(\eta)\right) \cap \left\{c_{t_2} - c_{t_1} > \frac{\eta}{2}\right\}, \quad (28)$$

consisting of all “good” magnetizations at time  $t_1$  up to time  $t_2$ , with an increase of at least  $\eta/2$  from  $t_1$  to  $t_2$ . Subject to  $c_t \in D_t$  for  $t_1 \leq t < t_2$  and  $c_{t_2} - c_{t_1} > \eta/2$ ,

$$\begin{aligned} S_{t_1, t_2} &= \sum_{t=t_1+1}^{t_2} \left(c_t - c_{t-1} + \frac{\gamma}{2n}\right) \\ &= c_{t_2} - c_{t_1} + \frac{\gamma}{2n}(t_2 - t_1) > \frac{\eta}{2}, \end{aligned} \quad (29)$$

from which the containment  $B_{t_1, t_2}(\eta) \subseteq \{S_{t_1, t_2} \geq \eta/2\}$  follows. Hence

$$\mathbb{P}\left(\bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1, t_2}(\eta)\right) \leq n^2 e^{-\Omega(\sqrt{n})} = e^{-\Omega(\sqrt{n})}. \quad (30)$$

Take  $n$  large enough. Suppose  $c_t > c_0 + \eta$  for some  $0 \leq t \leq T$ . Then there exists a  $t_1$  such that  $c_0 - 2\eta \leq c_s$  for all  $t_1 \leq s \leq t$ . Define  $t_2$  to be the least time greater than  $t_1$  with  $c_{t_2} > c_0 + \eta$ . Then  $c_t \in D_t$  for all  $t_1 \leq t < t_2$  and  $c_{t_2} - c_{t_1} > \eta/2$ . This implies that

$$\{c_t : c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T\} \subseteq \bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1, t_2}, \quad (31)$$

and further implies that

$$\mathbb{P}(c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T) \leq e^{-\Omega(\sqrt{n})}. \quad (32)$$

We have thus shown that the normalized magnetization  $c_t$  remains below  $c_0 + \eta$  for all  $0 \leq t \leq T$  with exponentially high probability, provided that  $c_0$  is suitably bounded away from any other fixed point of  $\lambda$ . (Note: we say an event has exponentially high probability if its complement has exponentially small probability.) Next we show that  $c_T$  ends below  $c_0 - \eta$  with exponentially high probability. We prove this by showing that  $c_t$  actually reaches  $c_0 - 2\eta$  with exponentially high probability, and then by the preceding argument will have exponentially small probability of increasing to  $c_0 - \eta$ . We have

$$\begin{aligned} \mathbb{P}(c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T) &\leq \mathbb{P}\left(\bigcap_{0 \leq t \leq T} D_t(\eta)\right) \\ &\quad + \mathbb{P}(c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T). \end{aligned} \quad (33)$$

Subject to  $c_t \in D_t(\eta)$  for  $0 \leq t \leq T$  and noticing that at worst  $c_0 = 1$  and  $c_T = 0$ ,

$$\begin{aligned} S_{0, T} &= \sum_{t=1}^T \left(c_t - c_{t-1} + \frac{\gamma}{2n}\right) \mathbf{1}_{D_{t-1}} \\ &= c_T - c_0 + \frac{\gamma}{2n}T \geq -1 + \frac{\gamma}{2n}T. \end{aligned} \quad (34)$$

Using the Chernoff bound on  $S_{0,T}$  and assume that  $d > 2/\gamma$ ,

$$\begin{aligned} \mathbb{P}(c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T) &\leq \mathbb{P}\left(S_{0,T} \geq -1 + \frac{\gamma}{2n}T\right) + e^{-\Omega(\sqrt{n})} \\ &\leq \frac{\mathbb{E}(e^{\theta S_{0,T}})}{e^{\theta(-1 + \frac{\gamma}{2n}T)}} + e^{-\Omega(\sqrt{n})} \\ &= e^{-\Omega(\sqrt{n})}. \end{aligned} \quad (35)$$

Finally,

$$\begin{aligned} \mathbb{P}(c_T \geq c_0 - \eta) &\leq \mathbb{P}(c_T \geq c_0 - \eta \text{ and } c_t < c_0 - 2\eta \text{ for some } 0 \leq t \leq T) \\ &\quad + \mathbb{P}(\{c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T\}) \leq e^{-\Omega(\sqrt{n})}, \end{aligned} \quad (36)$$

where to bound the first probability on the right, we apply the bound on “bad” magnetization (32) with minor adaptation: initial magnetization  $c_t$  (for some  $0 \leq t \leq T$ ) in place of  $c_0$ , time interval under consideration  $[t, T]$  in place of  $[0, T]$ , and the increase in magnetization  $c_0 - 2\eta \rightarrow c_0 - \eta$  in place of  $c_0 \rightarrow c_0 + \eta$ .  $\square$

Repeated application of Theorem 3 shows that after a burn-in period on the order of  $O(n)$ , any suitably chosen configuration ends up close to an attractor  $c^*$  with exponentially high probability. Recall the definition of high-temperature phase and low-temperature phase from Section 1.4. The following corollaries are immediate.

**Corollary 4.** *In the high-temperature phase, suppose that  $c^*$  is the unique solution to  $\lambda(c) = c$  and  $\lambda'(c^*) < 1$ . For any  $\varepsilon > 0$ , there exists  $\alpha > 0$  such that for any initial configuration with associated magnetization  $c_0$ , when  $t \geq \alpha n$  we have*

$$\mathbb{P}(c_t \geq c^* + \varepsilon) \leq e^{-\Omega(\sqrt{n})} \quad (37)$$

and

$$\mathbb{P}(c_t \leq c^* - \varepsilon) \leq e^{-\Omega(\sqrt{n})}. \quad (38)$$

**Corollary 5.** *In the low-temperature phase, suppose that  $c^*$  is a solution to  $\lambda(c) = c$  and  $\lambda'(c^*) < 1$ . Take  $\varepsilon > 0$ . If the associated magnetization  $c_0$  for some initial configuration satisfies  $c^* - \varepsilon \leq c_0 \leq c^* + \varepsilon$ , then there exists  $\beta > 0$ ,*

$$\mathbb{P}\left(\sup_{0 < t < e^{\beta\sqrt{n}}} c_t \geq c^* + 2\varepsilon\right) \leq e^{-\Omega(\sqrt{n})} \quad (39)$$

and

$$\mathbb{P}\left(\inf_{0 < t < e^{\beta\sqrt{n}}} c_t \leq c^* - 2\varepsilon\right) \leq e^{-\Omega(\sqrt{n})}. \quad (40)$$

### 3. Fast mixing at high-temperature

In this section we study the mixing time of the Glauber dynamics in the high-temperature phase. We first establish an upper bound  $O(n \log n)$  using path coupling techniques of Bubley and Dyer [13]. Consider two arbitrary spin configurations  $X, Y \in \mathcal{X}$ . Taking “attractive” parameters  $\alpha_i \geq 0$  ensures that we may apply a monotone coupling on the chain  $(X_t, Y_t)$ :  $X_t$  is a version of the Glauber dynamics with starting state  $X$  and  $Y_t$  is a version of the Glauber dynamics with starting state  $Y$ , if  $X_0 \leq Y_0$  then  $X_t \leq Y_t$  for all  $t$ . We write  $\mathbb{P}_{X,Y}$  and  $\mathbb{E}_{X,Y}$  for the underlying probability measure and associated expectation. To keep the notation light, we omit the explicit dependence on  $X$  and  $Y$  when it is clear from the context. To understand how far apart  $X_t$  and  $Y_t$  are, we introduce the Hamming distance, which records the number of vertices where the two configurations disagree. Define  $\rho : \mathcal{X} \times \mathcal{X} \rightarrow \{0, \dots, n\}$  by

$$\rho(X, Y) = \sum_{i=1}^n |X(i) - Y(i)|. \quad (41)$$

Following standard contraction argument, it suffices to estimate the average distance after one Glauber update between two coupled configurations  $X$  and  $Y$  with Hamming distance  $\rho(X, Y) = 1$ .

**Lemma 6.** *Assume that  $\sup_{0 \leq c \leq 1} \lambda'(c) < 1$ . Let  $X, Y \in \mathcal{X}$  be two spin configurations satisfying  $X \leq Y$  and  $\rho(X, Y) = 1$ . Set the initial state  $X_0 = X$  and  $Y_0 = Y$ . Then there exists  $\delta > 0$  depending only on  $p, \alpha_1, \alpha_2$  such that a single step of the Glauber dynamics can be coupled in such a way that when  $n$  is sufficiently large:*

$$\mathbb{E}(\rho(X_1, Y_1)) \leq e^{-\delta/n}. \quad (42)$$



**Proof.** Let  $X, Y \in \mathcal{X}$  be two configurations such that  $X \leq Y$  and there exists a single vertex  $i$  such that  $X(i) = 0$  and  $Y(i) = 1$ . Let  $U$  be a uniform random variable on  $[0, 1]$ . We apply the standard monotone coupling, where  $U$  is used as the common source of noise to update both chains so that they agree as often as possible. From the mechanism described in Section 1.1, the chain evolves by selecting a vertex  $j$  uniformly at random and updating the spin at  $j$ . Set

$$X_1(j) = \begin{cases} 1 & U \leq p_+(X, j), \\ 0 & U > p_+(X, j), \end{cases} \quad Y_1(j) = \begin{cases} 1 & U \leq p_+(Y, j), \\ 0 & U > p_+(Y, j), \end{cases} \quad (43)$$

and  $X_1(k) = X(k)$  and  $Y_1(k) = Y(k)$  for all  $k \neq j$ . Define the function  $f(S)$  as

$$f(S) = \frac{p \exp\left(\frac{\alpha_1}{n}S + \frac{\alpha_2}{2n^2}S(S-1)\right)}{p \exp\left(\frac{\alpha_1}{n}S + \frac{\alpha_2}{2n^2}S(S-1)\right) + (1-p)}. \quad (44)$$

If  $j = i$ , then  $p_+(X, j) = f(S(X, j)) = p_+(Y, j)$  and so  $\rho(X_1, Y_1) = 0$ . For  $j \neq i$ , we have  $p_+(X, j) = f(S(X, j))$  while  $p_+(Y, j) = f(S(X, j) + 1)$ , where  $0 \leq S(X, j) \leq n-2$ . Since  $f$  is a smooth and increasing function, this shows that  $p_+(X, j) \leq p_+(Y, j)$ . Hence  $\rho(X_1, Y_1) = 2$  if  $p_+(X, j) < U \leq p_+(Y, j)$  and  $\rho(X_1, Y_1) = 1$  otherwise.

We wish to find an upper bound for

$$\mathbb{E}(\rho(X_1, Y_1)) = 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \neq i} (p_+(Y, j) - p_+(X, j)). \quad (45)$$

To that end, we compute, by the mean value theorem

$$p_+(Y, j) - p_+(X, j) = f(S(X, j) + 1) - f(S(X, j)) = \frac{g'(\bar{c})}{n}, \quad (46)$$

where  $0 \leq \bar{c} \leq 1$  and  $g'(c)$  is defined as

$$g'(c) = \frac{p(1-p) \exp\left(\alpha_1 c + \frac{\alpha_2}{2}c\left(c - \frac{1}{n}\right)\right)}{p \exp\left(\alpha_1 c + \frac{\alpha_2}{2}c\left(c - \frac{1}{n}\right)\right) + (1-p)} \left(\alpha_1 + \alpha_2 c - \frac{\alpha_2}{2n}\right). \quad (47)$$

Compare  $g'(c)$  against  $\lambda'(c)$ , where  $\lambda(c)$  is defined as in (14),

$$\lambda'(c) = \frac{p(1-p) \exp\left(\alpha_1 c + \frac{\alpha_2}{2}c^2\right)}{p \exp\left(\alpha_1 c + \frac{\alpha_2}{2}c^2\right) + (1-p)} (\alpha_1 + \alpha_2 c). \quad (48)$$

Via standard analytical arguments as in the proof of Lemma 2, the difference  $g'(c) - \lambda'(c) = O(1/n)$ . Therefore

$$\begin{aligned} \mathbb{E}(\rho(X_1, Y_1)) &\leq 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \neq i} \left( \frac{1}{n} \sup_{0 \leq c \leq 1} \lambda'(c) + O\left(\frac{1}{n^2}\right) \right) \\ &\leq 1 - \frac{1}{n} + \frac{n-1}{n^2} \sup_{0 \leq c \leq 1} \lambda'(c) + O\left(\frac{1}{n^2}\right) \\ &\leq 1 - \frac{1 - \sup_{0 \leq c \leq 1} \lambda'(c)}{n} + O\left(\frac{1}{n^2}\right). \end{aligned} \quad (49)$$

Define  $\delta > 0$  as  $\delta = (1 - \sup_{0 \leq c \leq 1} \lambda'(c)) / 2$ . Then for  $n$  sufficiently large, we obtain

$$\mathbb{E}(\rho(X_1, Y_1)) \leq 1 - \frac{\delta}{n} \leq e^{-\delta/n}. \quad \square \quad (50)$$

The requirement  $\sup_{0 \leq c \leq 1} \lambda'(c) < 1$  in Lemma 6 may be weakened. By Corollary 4, in the high-temperature phase, with exponentially high probability, after  $O(n)$  time steps the associated magnetization of all configurations are within an  $\varepsilon$ -neighborhood of the unique solution  $c^*$  of  $\lambda$  with  $\lambda'(c^*) < 1$ . The supremum referenced in (49) thus need not be taken over the entire interval  $[0, 1]$  but just  $[c^* - \varepsilon, c^* + \varepsilon]$ , and is guaranteed to be less than 1 using smoothness of  $\lambda$ . Now take any two configurations  $X, Y \in \mathcal{X}$  with  $\rho(X, Y) = k$ , where  $1 \leq k \leq n$ . (At worst  $X(i) = 0$  and  $Y(i) = 1$  for all  $i \in \{1, \dots, n\}$ .) There is a sequence of states  $X_0, \dots, X_k$  such that  $X_0 = X, X_k = Y$ , and each neighboring pair  $X_i, X_{i+1}$  is unit Hamming distance apart. Applying Lemma 6 for configurations at unit distance, we have  $\mathbb{E}(\rho(X_1, Y_1)) \leq ne^{-\delta/n}$ . Iterating gives

$$\mathbb{E}(\rho(X_t, Y_t)) \leq ne^{-\delta t/n}. \quad (51)$$

**Theorem 7.** In the high-temperature phase, the mixing time for the Glauber dynamics is  $O(n \log n)$ .

**Proof.** By Theorem 14.6 and Corollary 14.7 of [12], (51) implies

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{n(\log n - \log \varepsilon)}{\delta} \right\rceil. \quad (52)$$

Setting  $\varepsilon = 1/4$ ,

$$t_{\text{mix}} \leq \left\lceil \frac{n(\log n + \log 4)}{\delta} \right\rceil. \quad \square \quad (53)$$

Next in [Theorem 8](#), by checking the total variance distance from the stationary distribution at time  $t^* := (n \log n)/4$ , we establish a matching lower bound  $\Omega(n \log n)$  for the Glauber dynamics in the high-temperature phase. Together with [Theorem 7](#), the correct order for the mixing time,  $\Theta(n \log n)$ , is validated.

**Theorem 8.** *In the high-temperature phase, the mixing time for the Glauber dynamics is  $\Omega(n \log n)$ .*

**Proof.** Let  $f(X)$  count the number of vertices with spin 1 in a configuration  $X \in \mathcal{X}$ . Denote by  $\gamma$  the spectral gap associated with the Glauber dynamics. By Lemma 13.12 and Remark 13.13 of [12],

$$\gamma \leq \frac{\varepsilon(f)}{\text{Var}_\pi(f)}, \quad (54)$$

where the Dirichlet form

$$\varepsilon(f) = \frac{1}{2} \sum_{X, Y \in \mathcal{X}} (f(X) - f(Y))^2 \pi(X) P(X, Y), \quad (55)$$

and the variance under stationary distribution

$$\text{Var}_\pi(f) = \sum_{X \in \mathcal{X}} (f(X))^2 \pi(X) - \left( \sum_{X \in \mathcal{X}} f(X) \pi(X) \right)^2. \quad (56)$$

From the mechanism described in Section 1.1, for configurations  $X, Y \in \mathcal{X}$ ,  $P(X, Y)$  defined in (8) is zero unless  $X$  and  $Y$  differ at at most one vertex. This implies that

$$\varepsilon(f) \leq \frac{1}{2} \sum_{X, Y \in \mathcal{X}} \pi(X) P(X, Y) = \frac{1}{2} \sum_{X \in \mathcal{X}} \pi(X) = \frac{1}{2}, \quad (57)$$

and when applied to (54), further implies that  $2\text{Var}_\pi(f) \leq 1/\gamma$ . Hence

$$\log 2 (2\text{Var}_\pi(f) - 1) \leq \log 2 \left( \frac{1}{\gamma} - 1 \right) \leq t_{\text{mix}}, \quad (58)$$

where the second inequality uses spectral representation techniques (for details, see for example Theorem 12.4 of [12]). By [Theorem 7](#),  $t_{\text{mix}} = O(n \log n)$ , which then gives  $\text{Var}_\pi(f) = O(n \log n)$ . Let configuration  $X$  be chosen according to the stationary distribution  $\pi$ . By Chebyshev's inequality,

$$\pi(|f(X) - \mathbb{E}_\pi(f(X))| > n^{2/3}) \leq \frac{\text{Var}_\pi(f(X))}{n^{4/3}} = O(n^{-1/3} \log n). \quad (59)$$

Therefore asymptotically

$$\pi(|f(X) - \mathbb{E}_\pi(f(X))| \leq n^{2/3}) \rightarrow 1. \quad (60)$$

Let  $X^+, X^- \in \mathcal{X}$  be configurations such that  $X^+(i) = 1$  and  $X^-(i) = 0$  for every  $i \in \{1, \dots, n\}$ . Assume that  $X^+$  and  $X^-$  are coupled using the standard monotone coupling as described in the proof of [Lemma 6](#). So  $X_t^+ \geq X_t^-$  for all  $t$ . Let  $R_{t^*}$  denote the number of vertices not yet selected by the Glauber Dynamics by time  $t^*$ . By a coupon collecting argument,

$$\mathbb{E}(R_{t^*}) \asymp n^{3/4} \text{ and } \text{Var}(R_{t^*}) \leq \mathbb{E}(R_{t^*}). \quad (61)$$

(For details, see for example Lemma 7.12 of [12].) Let  $\varepsilon > 0$ . By Chebyshev's inequality, asymptotically

$$\mathbb{P}(|R_{t^*} - \mathbb{E}(R_{t^*})| > (1 - \varepsilon)\mathbb{E}(R_{t^*})) \leq \frac{\text{Var}(R_{t^*})}{(1 - \varepsilon)^2 (\mathbb{E}(R_{t^*}))^2} \leq O(n^{-3/4}), \quad (62)$$

which using set containment implies that

$$\mathbb{P}(R_{t^*} < \varepsilon \mathbb{E}(R_{t^*})) \leq \mathbb{P}(|R_{t^*} - \mathbb{E}(R_{t^*})| > (1 - \varepsilon)\mathbb{E}(R_{t^*})) \leq O(n^{-3/4}). \quad (63)$$

It follows that  $R_{t^*} = \Omega(n^{3/4})$  with probability tending to 1 asymptotically. Since  $f(X_t^+) - f(X_t^-) \geq R_{t^*}$ , we conclude that  $f(X_{t^*}^+) - f(X_{t^*}^-) = \Omega(n^{3/4})$  with probability tending to 1 asymptotically.

Define sets  $A$  and  $B$  respectively as

$$A := \{|f(X_{t^*}^+) - \mathbb{E}_\pi(f(X))| \leq n^{2/3}\} \quad (64)$$

and

$$B := \{|f(X_{t^*}^-) - \mathbb{E}_\pi(f(X))| \leq n^{2/3}\}. \quad (65)$$

Then by the triangle inequality, their intersection  $A \cap B$ , if nonempty, satisfies  $f(X_{t^*}^+) - f(X_{t^*}^-) = O(n^{2/3})$ . Since  $n^{3/4} > n^{2/3}$ , this contradicts what was established in the previous paragraph. Therefore the sets  $A$  and  $B$  are asymptotically disjoint and so one of them has probability bounded above by  $1/2 - o(1)$ . Without loss of generality, say

$$\mathbb{P}(|f(X_{t^*}^+) - \mathbb{E}_\pi(f(X))| \leq n^{2/3}) \leq \frac{1}{2} - o(1). \quad (66)$$

By definition of the total variation distance and the bounds (60), (66),

$$\|P^{t^*}(X^+, \cdot) - \pi\|_{\text{TV}} \geq |P^{t^*}(X^+, A) - \pi(A)| \geq 1 - \frac{1}{2} + o(1). \quad (67)$$

Since at  $t_{\text{mix}}$ , the default for  $t_{\text{mix}}(\varepsilon)$ , the distance must be less than or equal to  $1/4$ , this shows that the mixing time is asymptotically bigger than  $t^* = (n \log n)/4$  proving the lower bound.  $\square$

#### 4. Slow mixing at low-temperature

In this section we study the mixing time of the Glauber dynamics in the low-temperature phase. Rather than analyzing the spin update probability  $\lambda$  defined in (14) directly, we find asymptotic expressions for components of the partition function  $Z$  (see Definition 1). For  $k \in \{0, \dots, n\}$ , define  $A_k = \{X : |\{i : X(i) = 1\}| = k\}$ . That is, the set  $A_k$  consists of spin configurations  $X \in \mathcal{X}$  whose number of vertices with spin 1 is  $k$  and number of vertices with spin 0 is  $n - k$ . Then

$$\pi(A_k) = \frac{1}{Z} \binom{n}{k} \exp\left(\frac{\alpha_1}{n} \binom{k}{2} + \frac{\alpha_2}{n^2} \binom{k}{3}\right) p^k (1-p)^{n-k}. \quad (68)$$

Let  $a_k$  be such that  $a_k = Z\pi(A_k)$ . Notice that  $\sum_{k=0}^n a_k = Z$ .

**Lemma 9.** Let  $c \in [0, 1]$  and  $a_k$  be defined as above, we have

$$\log(a_{\lfloor cn \rfloor}) = n(\varphi_{p, \alpha_1, \alpha_2}(c) + o(1)), \quad (69)$$

where

$$\varphi(c) = \frac{\alpha_1}{2} c^2 + \frac{\alpha_2}{6} c^3 - c \log \frac{c}{p} - (1-c) \log \frac{1-c}{1-p}. \quad (70)$$

**Proof.** Stirling's formula states that

$$n! \asymp \sqrt{2\pi} e^{-n} n^n \sqrt{n}. \quad (71)$$

Setting  $k = \lfloor cn \rfloor$ , the binomial coefficient admits an asymptotic formula:

$$\binom{n}{k} \asymp \frac{1}{\sqrt{2\pi c^{cn} \sqrt{c(1-c)}^{(1-c)n} \sqrt{1-c} \sqrt{n}}}. \quad (72)$$

Hence

$$\begin{aligned} \log(a_{\lfloor cn \rfloor}) &\asymp n \left( \frac{\alpha_1}{2} c^2 + \frac{\alpha_2}{6} c^3 - c \log \frac{c}{p} - (1-c) \log \frac{1-c}{1-p} \right) \\ &\quad - \log \left( \sqrt{2\pi c(1-c)n} \right) \\ &= n(\varphi(c) + o(1)). \quad \square \end{aligned} \quad (73)$$

Next in Lemmas 10 and 11, we reveal a deep relationship between  $\varphi$  (commonly referred to as the “free energy density”) and the spin update probability  $\lambda$ . As we will see, local maximizers for  $\varphi$  correspond to fixed points of  $\lambda$ , and concavity of  $\varphi$  at the local maximizer (indicating whether it is a local maximum or minimum) translates to the attractor/repellor characterization on the fixed point of  $\lambda$  previously described in Section 1.4.

**Lemma 10.** Let  $\lambda$  and  $\varphi$  be respectively defined as in (14) and (70). Then  $\lambda(c) = c \iff \varphi'(c) = 0$ .

**Proof.** We have the following string of equivalences

$$\begin{aligned}\lambda(c) = c &\iff \left(\frac{p}{1-p}\right) \exp\left(\alpha_1 c + \frac{\alpha_2}{2} c^2\right) = \frac{c}{1-c} \\ &\iff \alpha_1 c + \frac{\alpha_2}{2} c^2 + \log \frac{p}{1-p} - \log \frac{c}{1-c} = 0 \\ &\iff \varphi'(c) = 0. \quad \square\end{aligned}\tag{74}$$

**Lemma 11.** Let  $\lambda$  and  $\varphi$  be respectively defined as in (14) and (70). Suppose  $\varphi'(c) = 0$ . Then  $\varphi''(c) > 0 \iff \lambda'(c) > 1$ ,  $\varphi''(c) < 0 \iff \lambda'(c) < 1$ , and  $\varphi''(c) = 0 \iff \lambda'(c) = 1$ .

**Proof.** Suppose  $\varphi'(c) = 0$ . From Lemma 10,  $\lambda(c) = c$ . Therefore

$$\lambda'(c) = c(1-c)(\alpha_1 + \alpha_2 c). \tag{75}$$

The claim readily follows since

$$\varphi''(c) = \alpha_1 + \alpha_2 c - \frac{1}{c(1-c)}. \quad \square \tag{76}$$

Using a conductance argument whose idea goes back at least to Griffiths et al. [14], we now show that in the region where  $\varphi(c)$  has at least two local maximizers, the mixing time for the Glauber dynamics is at least exponential. In the language of  $\lambda$ , this establishes exponentially slow mixing of the Glauber dynamics when  $\lambda(c) = c$  has at least two solutions  $c$  satisfying  $\lambda'(c) < 1$ . Define the bottleneck ratio (Cheeger constant) of a Markov chain with stationary distribution  $\pi$  as

$$\Phi_* = \min_{S: \pi(S) \leq 1/2} \frac{Q(S, S^c)}{\pi(S)}, \tag{77}$$

where  $S$  is a set in the configuration space and  $Q$  is the edge measure given by

$$Q(X, Y) = \pi(X)P(X, Y) \text{ and } Q(A, B) = \sum_{X \in A, Y \in B} Q(X, Y). \tag{78}$$

Recall that  $P(X, Y)$  defined in (8) is the transition probability from configuration  $X$  to configuration  $Y$ , and so  $Q(A, B)$  is the probability of moving from set  $A$  to set  $B$  in one step of the chain when starting from the stationary distribution.

**Theorem 12.** In the low-temperature phase, the mixing time for the Glauber dynamics is  $e^{\Omega(n)}$ .

**Proof.** Notice that  $\varphi'(c) \rightarrow \infty$  as  $c \rightarrow 0$  and  $\varphi'(c) \rightarrow -\infty$  as  $c \rightarrow 1$ , so the local maximizers of  $\varphi$  are contained in  $(0, 1)$ . Let  $c_1$  be the smallest and  $c_2$  be the largest local maximizer of  $\varphi$ . We have  $c_1 \neq c_2$ . There exists  $\varepsilon > 0$  such that for all  $c < c_1$  and  $c_1 < c \leq c_1 + \varepsilon$ ,  $\varphi(c) < \varphi(c_1)$ , while for all  $c_2 - \varepsilon \leq c < c_2$  and  $c > c_2$ ,  $\varphi(c) < \varphi(c_2)$ , with  $c_1 + \varepsilon < c_2 - \varepsilon$ . Define the following two sets

$$S_1 = \{A_0, \dots, A_{\lfloor (c_1 + \varepsilon)n \rfloor}\} \tag{79}$$

and

$$S_2 = \{A_{\lfloor (c_2 - \varepsilon)n \rfloor}, \dots, A_n\}. \tag{80}$$

For  $n$  large enough,  $S_1 \cap S_2 = \emptyset$ ,  $A_{\lfloor (c_1 + \varepsilon)n \rfloor} \neq A_{\lfloor c_1 n \rfloor}$ , and  $A_{\lfloor (c_2 - \varepsilon)n \rfloor} \neq A_{\lfloor c_2 n \rfloor}$ . Since  $S_1$  and  $S_2$  are disjoint, at least one of them has probability bounded above by  $1/2$ . Without loss of generality, suppose  $S_1$  is such that  $\pi(S_1) \leq \frac{1}{2}$ . Then

$$\Phi_* \leq \frac{Q(S_1, S_1^c)}{\pi(S_1)} = \frac{1}{\pi(S_1)} \sum_{i=0}^{\lfloor (c_1 + \varepsilon)n \rfloor} \sum_{X \in A_i} \pi(X) \sum_{j=\lfloor (c_1 + \varepsilon)n \rfloor + 1}^n \sum_{Y \in A_j} P(X, Y). \tag{81}$$

Since a single step of the Glauber dynamics only changes the value of the normalized magnetization by at most  $1/n$ , the only non-zero transition probability in (81) is the transition from  $A_{\lfloor (c_1 + \varepsilon)n \rfloor}$  to  $A_{\lfloor (c_1 + \varepsilon)n \rfloor + 1}$ . It follows that

$$\begin{aligned}\Phi_* &\leq \frac{1}{\pi(S_1)} \sum_{i=0}^{\lfloor (c_1 + \varepsilon)n \rfloor} \sum_{X \in A_i} \pi(X) \leq \frac{\pi(A_{\lfloor (c_1 + \varepsilon)n \rfloor})}{\pi(A_{\lfloor c_1 n \rfloor})} \\ &= \frac{a_{\lfloor (c_1 + \varepsilon)n \rfloor}}{a_{\lfloor c_1 n \rfloor}} \asymp \frac{e^{n\varphi(c_1 + \varepsilon)}}{e^{n\varphi(c_1)}} = e^{n(\varphi(c_1 + \varepsilon) - \varphi(c_1))},\end{aligned}\tag{82}$$

where the asymptotics are derived in Lemma 9. Define  $\delta > 0$  as  $\delta = (\varphi(c_1) - \varphi(c_1 + \varepsilon)) / 2$ . Then the bottleneck ratio satisfies  $\Phi(S_1) \leq e^{-\delta n}$ . Using Theorem 7.3 of [12],

$$t_{\text{mix}} \geq \frac{1}{4\Phi_*} \geq \frac{e^{\delta n}}{4} = e^{\Omega(n)}. \quad \square \quad (83)$$

Call a Markov chain local if at most  $o(n)$  vertices are selected in each step. The argument used in the proof of Theorem 12 actually shows that in the low-temperature phase, the mixing is exponentially slow for any local Markov chain, with the (single-site) Glauber dynamics being one such instance. We remark that there is a difference in the qualitative nature of the phase transition investigated in this paper as compared with that in the standard statistical physics literature. While the asymptotic phase transitions in the rigorous statistical physics sense occur at parameter values giving non-unique global maximizers of the free energy density, the asymptotic transition from high-temperature phase to low-temperature phase arises as a consequence of the non-uniqueness of local maximizers for the free energy density. This discrepancy may not come as a surprise, since in simulations it is often hard to distinguish between a local maximizer and a global maximizer and the algorithm may become trapped at a local maximizer; one solution might be to add controlled moves based on network geometry.

## 5. Slower burn-in along critical curve

In this section we study the mixing time of the Glauber dynamics along the critical curve, corresponding to parameters for which  $\lambda(c) = c$  admits a unique solution  $c$  with  $\lambda'(c) = 1$ . We first identify explicitly the high-temperature vs. low-temperature phase. As explained in Section 4, the two phases may be alternatively determined by whether there is a unique local maximizer for  $\varphi$  defined in (70). The phase identification thus reduces to a 3-dimensional intricate calculus problem. Though straightforward as it sounds, various tricks are needed to solve it analytically. The crucial idea is to minimize the effect of the parameters  $p, \alpha_1, \alpha_2$  on the free energy density  $\varphi$  one by one. (See [15] for more details of the calculation in a related model.) Denote by  $l(c) := \varphi(c) - \log(1 - p)$ .

**Proposition 13.** Fix  $\alpha_2$ . Consider the maximization problem for

$$l_{\alpha_2}(c; p, \alpha_1) = \left( \log \frac{p}{1-p} \right) c + \frac{\alpha_1}{2} c^2 + \frac{\alpha_2}{6} c^3 - c \log c - (1-c) \log(1-c) \quad (84)$$

on the interval  $[0, 1]$ , where  $0 < p < 1$  and  $-\infty < \alpha_1 < \infty$  are parameters. Then there is a V-shaped region in the  $(p, \alpha_1)$ -plane with corner point  $(p^c, \alpha_1^c)$ ,

$$p^c = \frac{\bar{c} \exp\left(\frac{4\bar{c}-3}{2(1-\bar{c})^2}\right)}{\bar{c} \exp\left(\frac{4\bar{c}-3}{2(1-\bar{c})^2}\right) + (1-\bar{c})}, \quad (85)$$

$$\alpha_1^c = \frac{2-3\bar{c}}{\bar{c}(1-\bar{c})^2}, \quad (86)$$

where  $\bar{c}$  is uniquely determined by

$$\alpha_2 = \frac{2\bar{c}-1}{\bar{c}^2(1-\bar{c})^2}. \quad (87)$$

Outside this region,  $l_{\alpha_2}(c)$  has only one local maximizer  $c^*$ . Inside this region,  $l_{\alpha_2}(c)$  has exactly two local maximizers  $c_1^*$  and  $c_2^*$ .

**Proof.** The location of maximizers of  $l_{\alpha_2}(c)$  on the interval  $[0, 1]$  is closely related to the properties of its derivatives:

$$\begin{aligned} l'_{\alpha_2}(c) &= \log \frac{p}{1-p} + \alpha_1 c + \frac{\alpha_2}{2} c^2 - \log \frac{c}{1-c}, \\ l''_{\alpha_2}(c) &= \alpha_1 + \alpha_2 c - \frac{1}{c(1-c)}, \\ l'''_{\alpha_2}(c) &= \alpha_2 + \frac{1-2c}{c^2(1-c)^2}. \end{aligned} \quad (88)$$

We check that  $l'''_{\alpha_2}(c)$  is monotonically decreasing on  $[0, 1]$ ,  $l'''_{\alpha_2}(0) = \infty$ , and  $l'''_{\alpha_2}(1) = -\infty$ . Thus there is a unique  $\bar{c}$  in  $(0, 1)$  such that  $l'''_{\alpha_2}(\bar{c}) = 0$ , with  $l'''_{\alpha_2}(c) > 0$  for  $c < \bar{c}$  and  $l'''_{\alpha_2}(c) < 0$  for  $c > \bar{c}$ . Since the correspondence between  $\alpha_2$  and  $\bar{c}$  is one-to-one, we may describe  $\alpha_2$  by (87).

This implies that  $l''_{\alpha_2}(c)$  is increasing from 0 to  $\bar{c}$ , and decreasing from  $\bar{c}$  to 1, with the global maximum achieved at  $\bar{c}$ , where

$$l''_{\alpha_2}(\bar{c}) = \alpha_1 + \frac{3\bar{c} - 2}{\bar{c}(1 - \bar{c})^2}. \quad (89)$$

Let  $\alpha_1^c$  be defined as in (86) so that  $l''_{\alpha_2}(\bar{c}; \alpha_1^c) = 0$ . It follows that for  $\alpha_1 \leq \alpha_1^c$ ,  $l''_{\alpha_2}(c) \leq 0$  on the entire interval  $[0, 1]$ ; whereas for  $\alpha_1 > \alpha_1^c$ ,  $l''_{\alpha_2}(c)$  takes on both positive and negative values, and we denote the transition points by  $c_1$  and  $c_2$  ( $c_1 < \bar{c} < c_2$ ). For fixed  $\alpha_2$ ,  $c_1$  and  $c_2$  are solely determined by  $\alpha_1$ , and vice versa. Let  $m(c) = \alpha_1 - l'_{\alpha_2}(c)$  so that  $\alpha_1 = m(c_1) = m(c_2)$ . We have  $m(0) = m(1) = \infty$ ,  $m(c)$  is decreasing from 0 to  $\bar{c}$ , and increasing from  $\bar{c}$  to 1.

We proceed to analyze properties of  $l'_{\alpha_2}(c)$  and  $l_{\alpha_2}(c)$  on the interval  $[0, 1]$ . For  $\alpha_1 \leq \alpha_1^c$ ,  $l'_{\alpha_2}(c)$  is monotonically decreasing. For  $\alpha_1 > \alpha_1^c$ ,  $l'_{\alpha_2}(c)$  is decreasing from 0 to  $c_1$ , increasing from  $c_1$  to  $c_2$ , then decreasing again from  $c_2$  to 1. We write down the explicit expressions of  $l'_{\alpha_2}(c_1)$  and  $l'_{\alpha_2}(c_2)$ :

$$\begin{aligned} l'_{\alpha_2}(c_1) &= \log \frac{p}{1-p} + \frac{1}{1-c_1} - \log \frac{c_1}{1-c_1} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c_1^2, \\ l'_{\alpha_2}(c_2) &= \log \frac{p}{1-p} + \frac{1}{1-c_2} - \log \frac{c_2}{1-c_2} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c_2^2. \end{aligned} \quad (90)$$

Notice that  $l_{\alpha_2}(c)$  is a bounded continuous function,  $l'_{\alpha_2}(0) = \infty$ , and  $l'_{\alpha_2}(1) = -\infty$ , so  $l_{\alpha_2}(c)$  cannot be maximized at 0 or 1. For  $\alpha_1 \leq \alpha_1^c$ ,  $l'_{\alpha_2}(c)$  crosses the  $c$ -axis only once, going from positive to negative. Thus  $l_{\alpha_2}(c)$  has a unique local maximizer  $c^* = \bar{c}$ . For  $\alpha_1 > \alpha_1^c$ , the situation is more complicated. If  $l'_{\alpha_2}(c_1) \geq 0$  (resp.  $l'_{\alpha_2}(c_2) \leq 0$ ),  $l_{\alpha_2}(c)$  has a unique local maximizer at a point  $c^* > c_2$  (resp.  $c^* < c_1$ ). If  $l'_{\alpha_2}(c_1) < 0 < l'_{\alpha_2}(c_2)$ , then  $l_{\alpha_2}(c)$  has two local maximizers  $c_1^*$  and  $c_2^*$ , with  $c_1^* < c_1 < \bar{c} < c_2 < c_2^*$ .

Let

$$n(c) = \frac{1}{1-c} - \log \frac{c}{1-c} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c^2 \quad (91)$$

so that  $l'_{\alpha_2}(c_1) = \log(p/(1-p)) + n(c_1)$  and  $l'_{\alpha_2}(c_2) = \log(p/(1-p)) + n(c_2)$ . We have  $n(0) = \infty$ ,  $n(1) = \infty$ , with derivative  $n'(c)$  given by

$$n'(c) = c \left( \frac{1-2\bar{c}}{\bar{c}^2(1-\bar{c})^2} - \frac{1-2c}{c^2(1-c)^2} \right) = c (l'''_{\alpha_2}(\bar{c}) - l'''_{\alpha_2}(c)). \quad (92)$$

As  $l'''_{\alpha_2}(c)$  is monotonically decreasing on  $[0, 1]$ ,  $n(c)$  is decreasing from 0 to  $\bar{c}$ , and increasing from  $\bar{c}$  to 1, with the global minimum achieved at  $\bar{c}$ ,

$$n(\bar{c}) = \frac{1}{1-\bar{c}} - \log \frac{\bar{c}}{1-\bar{c}} + \frac{1-2\bar{c}}{2(1-\bar{c})^2}. \quad (93)$$

This implies that  $l'_{\alpha_2}(c_1; p, \alpha_1^c) \geq 0$  for  $p \geq p^c$  where  $p^c$  is defined in (85). The only possible region in the  $(p, \alpha_1)$ -plane where  $l'_{\alpha_2}(c_1) < 0 < l'_{\alpha_2}(c_2)$  is thus bounded by  $p < p^c$  and  $\alpha_1 > \alpha_1^c$ .

Finally, we analyze the behavior of  $l'_{\alpha_1}(c_1)$  and  $l'_{\alpha_1}(c_2)$  more closely when  $p$  and  $\alpha_1$  are chosen from this region. Recall that  $c_1 < \bar{c} < c_2$ . By monotonicity of  $n(c)$  on the intervals  $(0, \bar{c})$  and  $(\bar{c}, 1)$ , there exist continuous functions  $a(p)$  and  $b(p)$  of  $p$ , such that  $l'_{\alpha_2}(c_1) < 0$  for  $c_1 > a(p)$  and  $l'_{\alpha_2}(c_2) > 0$  for  $c_2 > b(p)$ . As  $p \rightarrow 0$ ,  $a(p) \rightarrow 0$  and  $b(p) \rightarrow 1$ .  $a(p)$  is an increasing function of  $p$ , whereas  $b(p)$  is a decreasing function, and they satisfy  $n(a(p)) = n(b(p)) = -p$ . The restrictions on  $c_1$  and  $c_2$  yield restrictions on  $\alpha_1$ , and we have  $l'_{\alpha_2}(c_1) < 0$  for  $\alpha_1 < m(a(p))$  and  $l'_{\alpha_2}(c_2) > 0$  for  $\alpha_1 > m(b(p))$ . As  $p \rightarrow 0$ ,  $m(a(p)) \rightarrow \infty$  and  $m(b(p)) \rightarrow \infty$ .  $m(a(p))$  and  $m(b(p))$  are both decreasing functions of  $p$  and they satisfy  $l'_{\alpha_2}(c_1; p, m(a(p))) = l'_{\alpha_2}(c_2; p, m(b(p))) = 0$ . As  $l'_{\alpha_2}(c_2; p, \alpha_1) > l'_{\alpha_2}(c_1; p, \alpha_1)$  for every  $(p, \alpha_1)$ , the curve  $m(b(p))$  must lie below the curve  $m(a(p))$ , and together they generate the bounding curves for the V-shaped region in the  $(p, \alpha_1)$ -plane with corner point  $(p^c, \alpha_1^c)$  where two local maximizers exist for  $l_{\alpha_2}(c)$ .  $\square$

From Proposition 13, the critical curve is traced out by  $(p^c, \alpha_1^c, \alpha_2)$  defined in (85)–(87), where we take  $1/2 \leq \bar{c} = c^* \leq 2/3$  to meet the non-negativity constraints on  $\alpha_1^c$  and  $\alpha_2$ . See Fig. 3. We delve deeper into the behavior of the function  $\lambda$  defined in (14) along the critical curve. To lighten the notation, we denote the associated parameters by  $(p, \alpha_1, \alpha_2)$ , and the unique local maximizer by  $c^*$  with  $1/2 \leq c^* \leq 2/3$ .

**Lemma 14.** Along the critical curve, we have

- (1)  $\lambda(c^*) = c^*$ ,  $\lambda'(c^*) = 1$ ,  $\lambda''(c^*) = 0$ , and  $\lambda'''(c^*) \leq -8$ .
- (2)  $\lambda''(c) \leq 0$  for  $c \geq c^*$ , and  $\lambda''(c) \geq 0$  for  $c \leq c^*$ .

**Proof.** The first claim follows from direct computation. We spell out some details. It is clear that  $\lambda(c^*) = c^*$ . Letting  $A = p \exp(\alpha_1 c + \frac{\alpha_2 c^2}{2}) / (1-p)$ , we write  $\lambda$ 's first few derivatives as

$$\lambda'(c) = \frac{\alpha_1 + \alpha_2 c}{1+A} \lambda(c),$$

$$\begin{aligned}\lambda''(c) &= \lambda(c) \frac{\alpha_2}{1+A} + (1-A) \frac{\lambda'(c)^2}{\lambda(c)}, \\ \lambda'''(c) &= \lambda'(c) \frac{\alpha_2}{1+A} - \alpha_2 \lambda(c) \lambda'(c) + 2(1-A) \frac{\lambda'(c) \lambda''(c)}{\lambda(c)} \\ &\quad - (1-A) \frac{\lambda'(c)^3}{\lambda(c)^2} - (\alpha_1 + \alpha_2 c) A \frac{\lambda'(c)^2}{\lambda(c)}.\end{aligned}\quad (94)$$

Substituting the parameter values in (85) (86) (87) yields  $\lambda'(c^*) = 1$ ,  $\lambda''(c^*) = 0$ , and

$$\lambda'''(c^*) = \frac{-6(c^*)^2 + 6(c^*) - 2}{(c^*)^2(1-c^*)^2} \leq -8. \quad (95)$$

For the second claim, we show that for  $c \geq c^*$ ,  $\lambda''(c) \leq 0$ . The parallel claim may be verified using a similar line of reasoning. As derived previously,

$$\lambda''(c) = \frac{A}{(1+A)^3} ((1+A)\alpha_2 + (1-A)(\alpha_1 + \alpha_2 c)^2). \quad (96)$$

Notice that  $\lambda''(c) \leq 0$  precisely when

$$\frac{\alpha_2 + (\alpha_1 + \alpha_2 c)^2}{(\alpha_1 + \alpha_2 c)^2 - \alpha_2} \leq A, \quad (97)$$

where equality holds when  $c = c^*$ . For  $c$  increasing from  $c^*$ ,  $A$  is increasing whereas the left hand of the above inequality is decreasing. Our claim thus follows.  $\square$

By Lemma 14, the expected magnetization drift  $(\lambda(c^*) - c^*)/n$  (15) drops from first order to third order along the critical curve as compared with other parameter regions. As a result, we anticipate that the burn-in will be slower. Theorem 15 establishes an upper bound. Utilizing coupling techniques from [16] with minor adaptation, an  $O(n^{2/3})$  mixing of the Glauber dynamics is further expected. For details, see Lemma 2.9 and Theorem 4.1 of [16].

**Theorem 15.** *Along the critical curve, the burn-in time for the Glauber dynamics is  $O(n^{3/2})$ .*

**Proof.** From Lemma 2,

$$\mathbb{E}(c_{t+1} - c^* \mid c_t - c^*) = \frac{1}{n} (\lambda(c_t) - \lambda(c^*)) + \left(1 - \frac{1}{n}\right) (c_t - c^*) + O\left(\frac{1}{n^2}\right). \quad (98)$$

Let  $e_t = c_t - c^*$  and define  $g(e_t) = \lambda(e_t + c^*) - \lambda(c^*)$ . Then

$$\begin{aligned}\mathbb{E}(|e_{t+1}| \mid e_t) &= \frac{1}{n} g(|e_t|) + \left(1 - \frac{1}{n}\right) |e_t| + O\left(\frac{1}{n^2}\right) \text{ for } e_t \geq 0, \\ \mathbb{E}(|e_{t+1}| \mid e_t) &= -\frac{1}{n} g(-|e_t|) + \left(1 - \frac{1}{n}\right) |e_t| + O\left(\frac{1}{n^2}\right) \text{ for } e_t < 0.\end{aligned}\quad (99)$$

Define  $\tau_0 = \min\{t \geq 0 : |e_t| \leq 1/n\}$ . Note that  $e_t$  does not change sign when  $t < \tau_0$ . Multiplying both sides of (99) by the indicator function  $\mathbf{1}_{\{\tau_0 > t\}}$  and using that  $g(0) = 0$  and  $\mathbf{1}_{\{\tau_0 > t+1\}} \leq \mathbf{1}_{\{\tau_0 > t\}}$ , we have

$$\begin{aligned}\mathbb{E}(|e_{t+1}| \mathbf{1}_{\{\tau_0 > t+1\}} \mid e_t) &\leq \frac{1}{n} g(|e_t| \mathbf{1}_{\{\tau_0 > t\}}) + \left(1 - \frac{1}{n}\right) |e_t| \mathbf{1}_{\{\tau_0 > t\}} + O\left(\frac{1}{n^2}\right), \\ \mathbb{E}(|e_{t+1}| \mathbf{1}_{\{\tau_0 > t+1\}} \mid e_t) &\leq -\frac{1}{n} g(-|e_t| \mathbf{1}_{\{\tau_0 > t\}}) + \left(1 - \frac{1}{n}\right) |e_t| \mathbf{1}_{\{\tau_0 > t\}} + O\left(\frac{1}{n^2}\right).\end{aligned}\quad (100)$$

Let  $\theta_t^+ = \mathbb{E}(|e_t| \mathbf{1}_{\{\tau_0 > t\}})$ . By Lemma 14,  $g(e) \leq 0$  for  $e \geq 0$  and  $g(e) \geq 0$  for  $e \leq 0$ ,  $g$  is concave down on the non-negative axis and concave up on the negative axis. Taking expectation of both sides of (100) and applying Jensen's inequality on  $g$ ,

$$\begin{aligned}\theta_{t+1}^+ - \theta_t^+ &\leq -\frac{1}{n} (\theta_t^+ - g(\theta_t^+)) + O\left(\frac{1}{n^2}\right), \\ \theta_{t+1}^+ - \theta_t^+ &\leq -\frac{1}{n} (g(-\theta_t^+) - (-\theta_t^+)) + O\left(\frac{1}{n^2}\right).\end{aligned}\quad (101)$$

Let  $\mu > 0$  and suppose that  $\theta_t^+ \geq \mu$ . As in the proof of Theorem 3, by the extreme value theorem, there exists  $\gamma(\mu) > 0$  such that

$$\theta_{t+1}^+ - \theta_t^+ \leq -\frac{\gamma(\mu)}{n}. \quad (102)$$

Utilizing the negative drift  $-\gamma(\mu)/n$ , there exists a time  $t^* = O(n)$  so that  $\theta_t^+ \leq 1/4$  for all  $t \geq t^*$ . Consider the Taylor series expansion of  $g(\theta_t^+)$  and  $g(-\theta_t^+)$  and using Lemma 14:

$$\begin{aligned} g(\theta_t^+) &= \lambda'(c^*)\theta_t^+ + \frac{\lambda''(c^*)}{2}(\theta_t^+)^2 + \frac{\lambda'''(d_1)}{6}(\theta_t^+)^3 \leq \theta_t^+ - \frac{4}{3}(\theta_t^+)^3, \\ g(-\theta_t^+) &= -\lambda'(c^*)\theta_t^+ + \frac{\lambda''(c^*)}{2}(\theta_t^+)^2 - \frac{\lambda'''(d_2)}{6}(\theta_t^+)^3 \geq -\theta_t^+ + \frac{4}{3}(\theta_t^+)^3, \end{aligned} \quad (103)$$

where  $d_1 \in [c^*, c^* + \theta_t^+]$  and  $d_2 \in [c^* - \theta_t^+, c^*]$ . Then it follows that

$$\theta_{t+1}^+ \leq \theta_t^+ - \frac{4}{3n}(\theta_t^+)^3 + O\left(\frac{1}{n^2}\right) \quad (104)$$

for  $t \geq t^*$ . The remainder of the proof follows analogously as in the proof of Theorem 4.1 in [16]. For some  $c$ ,

$$\lim_{c \rightarrow \infty} \mathbb{P}(\tau_0 > cn^{3/2}) = 0 \quad (105)$$

uniformly in  $n$ , which implies that the Glauber dynamics may be coupled so that the magnetizations agree in  $O(n^{3/2})$  time steps.  $\square$

## 6. Generalizations and future work

Numerous extensions can be made about these vertex-weighted exponential models. We started the discussion with the edge-triangle lattice gas (Ising) model, but clearly more complicated subgraph densities can be considered. Denote by  $K_k$  a complete graph on  $k$  vertices so that an edge is  $K_2$  and a triangle is  $K_3$ . For consistency, denote by  $C_0 = 1$ ,  $C_1 = S(X, i)$ , and  $C_2 = T(X, i)$ . As in (7), the crucial quantity

$$C_m(X, i) = \sum_{i \neq i_1 \neq i_2 \neq \dots \neq i_m} X_{i_1} X_{i_2} \cdots X_{i_m} \quad (106)$$

satisfies  $C_m = \binom{C_1}{m}$ , which may be justified by noting that a term in the sum defining  $C_m$  has value 1 if and only if every vertex in the product has spin 1. Moving on, we have employed a discrete-time update of the network, but the network may be updated on a continuous-time basis, and this may be realized by posing iid Poisson clocks and examining the corresponding heat kernel. More significantly, rather than the simplifying assumption that a person is either interested or not in building a friendship, in reality a person probably has different levels of interest in forming a connection, and an edge is placed between two people when the joint interest exceeds a certain threshold value. Also, in social networks people have diverse attributes; only people with the same attribute or those with more than a specified number of attributes will establish a tie, which will fall within the regime of the random cluster model and multilayer networks. All these extensions are quite challenging both theoretically and computationally, especially when network geometry comes into play, but we hope to address at least some of them in future work.

After we gain an understanding of the small-world observed structure of big network data using Markov chain dynamics, we may use this knowledge for the prediction and control of general spreading processes on large-scale networks. These processes include the social influence of opinions, users' decisions to adopt products, and epidemic intervention strategies, etc. To illustrate, we return to information diffusion over Twitter mentioned at the beginning of this paper. Updating the weight of a vertex corresponding to a celebrity or a news source will definitely have more impact than that for ordinary people. So instead of running the Glauber dynamics that chooses a vertex at random, we choose the "hubs" of the network to update. This selective procedure decreases the mixing time and drives the spreading dynamics more efficiently towards equilibrium. Other properties of the chains may be studied simultaneously. For example, the cover time of the network may be interpreted as a realization of a "web crawl", and the hitting time may be interpreted as the necessary local queries to determine the global connectivity.

## Acknowledgments

Mei Yin's research was partially supported by NSF, USA grant DMS-1308333. She thanks Richard Kenyon for helpful conversations.

## References

- [1] B. Bollobás, *Random Graphs*, Academic Press, New York, 1985.
- [2] R. Durrett, *Random Graph Dynamics*, Cambridge University Press, New York, 2007.
- [3] R. van der Hofstad, *Random Graphs and Complex Networks*. <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf> (2014).
- [4] M. Newman, *Networks: An Introduction*, Oxford University Press, New York, 2010.
- [5] S.J. Cranmer, B.A. Desmarais, *Inferential network analysis with exponential random graph models*, *Pol. Anal.* 19 (2011) 66–86.



- [6] D. Krioukov, Clustering implies geometry in networks, *Phys. Rev. Lett.* 116 (2016) 208302.
- [7] A.D. Barbour, A. Röllin, Central limit theorems in the configuration model. [arXiv:1710.02644](https://arxiv.org/abs/1710.02644) (2017).
- [8] A.C.D. van Enter, R. Fernández, F. den Hollander, F. Redig, Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures, *Comm. Math. Phys.* 226 (2002) 101–130.
- [9] S. Bhamidi, G. Bresler, A. Sly, Mixing time of exponential random graphs, *Ann. Appl. Probab.* 21 (2011) 2146–2170.
- [10] S. Chatterjee, P. Diaconis, Estimating and understanding exponential random graph models, *Ann. Statist.* 41 (2013) 2428–2461.
- [11] D. Aldous, J. Fill, Reversible Markov Chains and Random Walks on Graphs. <http://www.stat.berkeley.edu/~aldous/RWG/book.html>, (2009).
- [12] D.A. Levin, Y. Peres, E.L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, 2008.
- [13] R. Bubley, M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, in: FOCS, 1997, pp. 223–231.
- [14] R.B. Griffiths, C.-Y. Weng, J.S. Langer, Relaxation times for metastable states in the mean-field model of a ferromagnet, *Phys. Rev.* 149 (1966) 301–305.
- [15] M. Yin, Critical phenomena in exponential random graphs, *J. Stat. Phys.* 153 (2013) 1008–1021.
- [16] D.A. Levin, M.J. Luczak, Y. Peres, Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability, *Probab. Theory Related Fields* 146 (2010) 223–265.