Supplementary Material:

Rapid Mixing Swendsen-Wang Sampler for Stochastic Partitioned Attractive Models

A Proofs of Key Lemmas for Theorem 2

A.1 Proof of Lemma 4

Let H be an induced subgraph of $G(n,p)$ of size cn. First, consider a probability that H contains a component of size $m = o(n)$

$$\begin{aligned}
&\Pr(H \text{ contains a component of size } m)\\
&\quad\le \binom{cn}{m}(1-p)^{m(cn-m)}\\
&\quad\le (cn)^m \exp\big(-(1-o(1))pcmn\big)\\
&\quad= \exp\big(-(1-o(1))pcmn\big).
\end{aligned}$$

Since, a number of possible choices of H is bounded by 2^n, from the union bound, no choice of H contains a component of size $> d$ with probability $1-e^{-\Omega(n)}$ for some constant d, i.e. every possible choice of H does not contain a component of size between d and $o(n)$. Furthermore, a number of components of size $\le d$ is bounded by $O(1)$ with probability $1-e^{-\Omega(n)}$ as

$$\begin{aligned}
&\Pr(H \text{ contains } \ell = O(1) \text{ components of size } \le d)\\
&\quad\le \binom{cn}{d}^{\ell}(1-p)^{\ell(cn-d)}\\
&\quad\le (cn)^{\Theta(1)} \exp\big(-\Theta(n)\ell\big)\\
&\quad\le \exp\big(-\Theta(n)\ell\big)
\end{aligned}$$

and as a number of possible choices of H is bounded above by 2^n.

Now, we show that every choices of H contain a unique component of size $\ge cn-\Theta(1)$ by bounding the following probability:

$$\begin{aligned}
&\Pr(H \text{ contains } \ge 2 \text{ components of size } \Theta(n))\\
&\quad\le \sum_{i,j=\Theta(n), i+j\le cn} \binom{cn}{i}(1-p)^{-i(cn-i)}\\
&\quad= \sum_{i,j=\Theta(n), i+j\le cn} (cn)^{\Theta(n)} \exp\big(-\Theta(n^2)\big)\\
&\quad= \exp\big(-\Theta(n^2)\big).
\end{aligned}$$

where the inequality follows from the fact that no edge between two components. Using the union bound on the all possible choice of H, we conclude that every H contain a unique component of size $\ge cn-\Theta(1)$ with probability $1-e^{-\Omega(n)}$.

A.2 Proof of Lemma 5

We first show that $G(n,p)$ is disconnected with probability $e^{-\Omega(n)}$ as follows:

$$\begin{aligned}
&\Pr(G(n,p) \text{ is disconnected})\\
&\le \sum_{i=1}^{n/2} \binom{n}{i} (1-p)^{i(n-i)}\\
&\le \sum_{i=1}^{n/2} n^i \exp\big(-pi(n-i)\big)\\
&= \sum_{i=1}^{n/2} \exp\big(i\log n - pi(n-i)\big)\\
&= \exp\big(-\Omega(n)\big).
\end{aligned}$$

From the above result, one can observe that for $G(n-o(1),p)$ is disconnected with probability $e^{-\Omega(n)}$. Now, we apply the union bound to obtain that every subgraph of size $n-O(\sqrt{n})$ of $G(n,p)$ is connected with probability $1-e^{-\Omega(n)}$ as the number of possible choices of $n-O(\sqrt{n})$ component is bounded by $\sum_{i=O(\sqrt{n})} n^i = \exp\big(O(n^{1/2}\log n)\big)$.

A.3 Proof of Lemma 6

Let H be an induced subgraph of $G(n,kn,p)$ of size $(c_L n, c_R kn)$. First, consider a probability that H contains a component of size $(m_L, m_R), m_L, m_R = o(n)$

$$\begin{aligned}
&\Pr(H \text{ contains a component of size } (m_L,m_R))\\
&\le \binom{c_L n}{m_L}\binom{c_R kn}{m_R}(1-p)^{m_L(c_R kn - m_R)+m_R(c_L n - m_L)}\\
&\le (c_L n)^{m_L}(c_R kn)^{m_R}\exp\big(-(1-o(1))(c_R k m_L + c_L m_R)pn\big)\\
&= \exp\big(-(1-o(1))(c_R k m_L + c_L m_R)pn\big).
\end{aligned}$$

Since, a number of possible choices of H is bounded by $2^{(k+1)n}$, from the union bound, no choice of H contains a component of size $> (d_L, d_R)$ with probability $1-e^{-\Omega(n)}$ for some constants d_L, d_R, i.e. every possible choice of H does not contain a component of size between (d_L, d_R) and $o(n)$. Furthermore, a number of components of size $\le (d_L, d_R)$ is bounded by $O(1)$ with probability $1-e^{-\Omega(n)}$ as

$$\begin{aligned}
&\Pr(H \text{ contains } \ell = O(1) \text{ components of size } \le (d_L,d_R))\\
&\le \binom{c_L n}{d_L}^{\ell}\binom{c_R kn}{d_R}^{\ell}(1-p)^{\ell((c_L n - d_L)+(c_R kn - d_R))}\\
&\le (c_L n)^{\Theta(1)}(c_R n)^{\Theta(1)}\exp\big(-\Theta(n)\ell\big)\\
&\le \exp\big(-\Theta(n)\ell\big)
\end{aligned}$$

and as a number of possible choices of H is bounded above by 2^n.

Now, we show that every choices of H contain a unique component of size $\ge \big(c_L n - \Theta(1), c_R kn - \Theta(1)\big)$ by bounding the following probability:

$$\begin{aligned}
&\Pr(H \text{ contains } \ge 2 \text{ components of size } \Theta(n))\\
&\le \sum_{i,j=\Theta(n), i,j\le n} \binom{c_L n}{i}\binom{c_R kn}{j}(1-p)^{-i(c_R kn - j)-j(c_L n - i)}\\
&= \sum_{i,j=\Theta(n), i,j\le n} (c_L n)^{\Theta(n)}(c_R kn)^{\Theta(n)}\exp\big(-\Theta(n^2)\big)\\
&= \exp\big(-\Theta(n^2)\big).
\end{aligned}$$

where the inequality follows from the fact that no edge between two components. Using the union bound on the all possible choice of H, we conclude that every H contain a unique component of size $\geq \big(c_L n - \Theta(1), c_R kn - \Theta(1)\big)$ with probability $1 - e^{-\Omega(n)}$.

A.4 Proof of Lemma 7

We first show that $G(n, kn, p)$ has an isolated vertex with probability $e^{-\Omega(n)}$ as follows:

$$\begin{aligned}
&\Pr(G(n, kn, p) \text{ has an isolated vertex}) \\
&\leq \sum_{i=1}^{n} (1-p)^{kn} + \sum_{i=1}^{kn} (1-p)^{n} \\
&= \exp\big(-\Omega(n)\big).
\end{aligned}$$

Now, we show that $G(n, kn, p)$ is disconnected with probability $e^{-\Omega(n)}$ as follows:

$$\begin{aligned}
&\Pr(G(n, kn, p) \text{ is disconnected} \,|\, G \text{ has no isolated vertex}) \\
&\leq \sum_{i=1}^{n-1} \sum_{j=1}^{n/2} \binom{n}{i} \binom{n}{j} (1-p)^{i(n-j)+j(n-i)} \\
&\leq \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} n^{i+j} \big[\exp\big(-pi(n-j) - pj(n-i)\big) + \exp\big(-pij - p(n-i)(n-j)\big)\big] \\
&\leq \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} \left[\exp\left((i+j)\log n - \frac{i+j}{2} pn\right) + \exp\left((i+j)\log n - \frac{1}{4} pn^2\right)\right] \\
&= \exp\big(-\Omega(n)\big).
\end{aligned}$$

One can follow that

$$\begin{aligned}
&\Pr(G(n, kn, p) \text{ is disconnected}) \\
&\quad \geq \Pr(G(n, kn, p) \text{ is disconnected} \,|\, G \text{ has no isolated vertex}) \\
&\qquad \times \Pr(G \text{ has no isolated vertex}) + \Pr(G \text{ has an isolated vertex}) \\
&\quad = 1 - e^{-\Omega(n)}.
\end{aligned}$$

From the above result, observe that for $G(n - o(1), kn - o(1), p)$ is disconnected with probability $e^{-\Omega(n)}$. Now, we apply the union bound to obtain that every subgraph of size $n - O(\sqrt{n})$ of $G(n, p)$ is connected with probability $1 - e^{-\Omega(n)}$ as the number of possible choices of $\big(n - O(\sqrt{n}), kn - O(\sqrt{n})\big)$ component is bounded by $\sum_{i,j=O(\sqrt{n})} n^{i+j} = \exp\big(O(n^{1/2}\log n)\big)$.

B Proofs of Key Lemmas for Theorem 3

In this section, we provide proofs of Lemmas 8-12. To this end, we first introduce a two-dimensional function F which captures the behaviour of the Swendsen-Wang dynamics and introduce the connection between F and the Ising model. Throughout this section, we only consider the Ising model on the complete bipartite graph of size (n, kn) with

$$\beta_{uv} = -\frac{1}{2}\log\left(1 - \frac{B}{n\sqrt{k}}\right), \ \gamma_v = 0 \quad \text{for all } (u, v) \in E, \ v \in V$$

where $B > 0$ is some constant.

B.1 Simplified Swendsen-Wang

We first introduce the following result [21] about the giant component of the bipartite Erdős-Rényi random graph.

Lemma 13 ([21, Theorem 6, Theorem 12]) *Consider the bipartite Erdős-Rényi random graph*

$$G = (V_L, V_R, E) = G(n, kn, p)$$

where $p = \frac{B}{n\sqrt{k}}$ *for some constant* $B > 0$ *and* $k \geq 1$ *is some constant. Then, the following statements hold a.a.s.*

a) For $B < 1$*, the largest (connected) component of* G *has size* $O(\log n)$*.*

b) For $B > 1$*, the following event happens:* G *has a unique "giant" component which consists of* $\theta_R kn(1+o(1))$ *vertices in* V_R *and* $\theta_L n(1+o(1))$ *vertices in* V_L *where* θ_R *is the unique positive solution of*

$$\theta_R + \exp\left(\frac{B}{\sqrt{k}}\left(\exp\left(-B\sqrt{k}\theta_R\right) - 1\right)\right) = 1 \tag{7}$$

and θ_L *is the unique positive solution of*

$$\theta_L + \exp\left(B\sqrt{k}\left(\exp\left(-\frac{B\theta_L}{\sqrt{k}}\right) - 1\right)\right) = 1. \tag{8}$$

The second largest component of G *has size* $O(\log^2 n)$*.*

c) For $B = 1$*, the largest component of* G *has size* $o(n)$*.*

By simple calculation, one can observe that (7), (8) reduce to

$$\exp(-B\sqrt{k}\theta_R) = 1 - \theta_L \qquad \exp\left(-\frac{B}{\sqrt{k}}\theta_L\right) = 1 - \theta_R. \tag{9}$$

Now, consider the Ising model on the complete bipartite graph $G = (V_L, V_R, E)$ of size (n, kn). We briefly explain what happens in a single iteration of the Swendsen-Wang chain on G for each step asymptotically. Given a spin configuration σ with $\alpha(\sigma) = (\alpha_L, \alpha_R)$, the step 2 of the Swendsen-Wang dynamics starting from σ is equivalent to sampling two bipartite Erdős-Rényi random graphs $G(\alpha_L n, \alpha_R kn, p)$, $G((1-\alpha_L)n, (1-\alpha_R)kn, p)$ where $p = \frac{B}{n\sqrt{k}}$.

Suppose $(1-\alpha_L)(1-\alpha_R)B \leq 1$ and $\alpha_L\alpha_R B > 1$. Then, by Lemma 13, there exists a single giant component of size $(\theta_L\alpha_L n, \theta_R\alpha_R kn)$ where (θ_L, θ_R) is a unique positive solution of

$$\exp(-B\sqrt{k}\alpha_R\theta_R) = 1 - \theta_L \qquad \exp\left(-\frac{B}{\sqrt{k}}\alpha_L\theta_L\right) = 1 - \theta_R, \tag{10}$$

and the other 'small' components have size $o(n)$ a.a.s. after the step 2 of the Swendsen-Wang dynamics. One can notice that (10) is equivalent to (9) by substituting $n \leftarrow \alpha_L n$, $k \leftarrow \frac{k\alpha_R}{\alpha_L}$ and $B \leftarrow \sqrt{\alpha_L\alpha_R}B$. At the step 3 of the Swendsen-Wang dynamics, asymptotically a half of the small components, which have size $((1-\theta_L\alpha_L)n/2, (1-\theta_R\alpha_R)kn/2)$, receive same spin with the giant component. Now suppose $(1-\alpha_L)(1-\alpha_R)B, \alpha_L\alpha_R B \leq 1$. Then after the step 2 of the Swendsen-Wang dynamics, every connected components have size $O(\log n)$. After the step 3 of the Swendsen-Wang dynamics, as each spin class asymptotically have a half of the vertices of V_L, V_R, it outputs a phase $(1/2, 1/2)$ asymptotically. We ignore the case $(1-\alpha_L)(1-\alpha_R)B > 1$ for now, i.e. we ignore the giant component of the smaller spin class, which will be handled in the proof of Lemma 8. Under these intuitions, one can expect that the following function F captures the behavior of the Swendsen-Wang chain (ignoring the giant component of the smaller spin class) on the complete bipartite graph.

$$F(\alpha_L, \alpha_R) := (F_L, F_R) = \left(\frac{1}{2}\left(1 + \theta_L\alpha_L\right), \frac{1}{2}\left(1 + \theta_R\alpha_R\right)\right) \tag{11}$$

where

$$(\theta_L, \theta_R) = \begin{cases} (0, 0) & \text{for } \sqrt{\alpha_L\alpha_R}B \leq 1 \\ \text{the unique solution of (10)} & \text{for } \sqrt{\alpha_L\alpha_R}B > 1 \end{cases}.$$

We note that F is continuous on $[0,1]^2$. Formally, one can prove the following lemma about the relation between the function F and the Swendsen-Wang chain, where we omit its proof since it is elementary under the above intuitions.

Lemma 14 *Let $\{X_t : t = 0, 1, \dots\}$ be the Swendsen-Wang chain on a complete bipartite graph of size (n, kn) with any constants $B \neq 2$ and starting phase $\alpha(X_0) = (\alpha_L, \alpha_R)$. If $\alpha_L \alpha_R B \neq 1$ and $(1-\alpha_L)(1-\alpha_R)B \leq 1$, i.e., the smaller spin class is subcritical, then $\alpha(X_1) = F(\alpha_L, \alpha_R) + (o(1), o(1))$ a.a.s.*

From the definition of F, (α_L, α_R) is a fixed point of F if and only if $\alpha_L = \frac{1}{2} + \frac{1}{2}\theta_L\alpha_L$, $\alpha_R = \frac{1}{2} + \frac{1}{2}\theta_R\alpha_R$, i.e., $\theta_L = \frac{2\alpha_L - 1}{\alpha_L}$, $\theta_R = \frac{2\alpha_R - 1}{\alpha_R}$. Substituting this relation into (10) results that every fixed points of F must satisfies the following equations

$$\exp\left(B\sqrt{k}(1-2\alpha_R)\right) = \frac{1-\alpha_L}{\alpha_L} \qquad \exp\left(\frac{B}{\sqrt{k}}(1-2\alpha_L)\right) = \frac{1-\alpha_R}{\alpha_R}. \tag{12}$$

One can expect that the Swendsen-Wang chain starting from the phase which correspond to the fixed point of F tends to stay around the fixed point of F asymptotically. Now we introduce two lemmas about the fixed point of F. Lemma 15 shows that F has a unique fixed point that is Jacobian attractive. Furthermore, Lemma 16 guarantees that for any starting point (α_L, α_R),

$$F^{(t)}(\alpha_L, \alpha_R) := \underbrace{F \circ \cdots \circ F}_{t}(\alpha_L, \alpha_R)$$

converges to the fixed point of F as $t \to \infty$.

Lemma 15 *The followings holds:*

1. *For constant $B < 2$, $(1/2, 1/2)$ is the unique fixed point of F and Jacobian attractive.*
2. *For constant $B > 2$, the solution $\alpha_L^*, \alpha_R^* \in (1/2, 1]$ of (12) is the unique fixed point of F and Jacobian attractive.*

Lemma 16 *For any point $(\alpha_L, \alpha_R) \in [0, 1]^2$, $F^{(t)}(\alpha_L, \alpha_R)$ converges to the unique fixed point of F as $t \to \infty$.*

The proofs of the above lemmas are presented in Section C.2 and Section C.3, respectively.

Finally, we provide the connection between F and the Ising model. Suppose the probability of some phase, say (α'_L, α'_R), of the Ising model on the complete bipartite graph of size (n, kn) dominates that of other phases, i.e., $\mu\big((\alpha'_L, \alpha'_R) \pm (\Theta(1), \Theta(1))\big) = 1 - o(1)$. Then the Swendsen-Wang chain must converge to (α'_L, α'_R) a.a.s. Since F converges to its unique fixed point by Lemma 16, one can naturally expect that the fixed point of F is equivalent to (α'_L, α'_R). The following lemma establishes such intuition formally.

Lemma 17 *For the Ising model on the complete bipartite graph of size (n, kn) with $\beta_{uv} = -\frac{1}{2}\log\left(1 - \frac{B}{n\sqrt{k}}\right)$ for some constant $B > 0$ and $\gamma_v = 0$, the 'maximum a posteriori phase' is*

$$\lim_{n\to\infty} \arg\max_{(\alpha_L, \alpha_R)} \Pr(\alpha_L, \alpha_R) = \begin{cases} \left(\frac{1}{2}, \frac{1}{2}\right) & \text{for } B \leq 2 \\ (\alpha_L^*, \alpha_R^*) & \text{for } B > 2 \end{cases}$$

where $\alpha_L^*, \alpha_R^* \in (1/2, 1]$ is the unique solution of (12).

The proof of the above lemma is presented in Section C.4.

B.2 Proof of Lemma 8

We first note that it suffices to prove Lemma 8 for any small enough $\delta > 0$. We start by stating the following claim.

Claim 18 *For any constant $B > 2$ and fixed point (α_L^*, α_R^*) of F, the following inequality holds*

$$(1-\alpha_L^*)(1-\alpha_R^*)B^2 < 1,$$

i.e. the smaller spin class of the phase corresponding to the fixed point of F is subcritical.

Proof. 1 With parametrization $z_L^* = 2\alpha_L^* - 1,\ z_R^* = 2\alpha_R^* - 1$, we have

$$(1-\alpha_L^*)(1-\alpha_R^*)B^2 = \frac{1}{4}\frac{(1-z_L^*)(1-z_R^*)}{z_L^* z_R^*}\log\frac{1+z_L^*}{1-z_L^*}\log\frac{1+z_R^*}{1-z_R^*}, \tag{13}$$

where we used the fact that (α_L^*, α_R^*) satisfies (12). In the proof of Lemma 17, we show that (33) holds. This completes the proof of Claim 18.

Due to the above claim, any small enough $\delta > 0$ satisfies $(1-\alpha_L^*+\delta)(1-\alpha_R^*+\delta)B^2 < 1$. Now, for $B > 2$, Lemma 16 implies that there exists a constant T_1 such that

$$F^{(T_1)}([0,1]^2) \subset [\alpha_L^*-\delta, \alpha_L^*+\delta] \times [\alpha_R^*-\delta, \alpha_R^*+\delta].$$

First, suppose $F(1-\alpha_{L,0}, 1-\alpha_{R,0}) = (1/2, 1/2)$, i.e. the smaller spin class is subcritical. Then, in T_1 iterations, the Swendsen-Wang chain moves l_∞-distance δ from (α_L^*, α_R^*) with probability $1-o(1)$ due to Lemma 14. Now, consider the case $F(1-\alpha_{L,0}, 1-\alpha_{R,0}) > (1/2, 1/2)$, i.e. two giant components appears in both spins in the step 2 of the Swendsen-Wang dynamics. Then, giant components merge with probability 1/2 and it results $\alpha(X_{T_1}) > (\alpha_L^*-\delta, \alpha_R^*-\delta)$ with probability $\Theta(1)$. Therefore, starting from $\alpha(X_{T_1}) > (\alpha_L^*-\delta, \alpha_R^*-\delta)$, the Swendsen-Wang chain also moves within l_∞-distance δ from (α_L^*, α_R^*) in T_1 iterations with probability $1-o(1)$ due to Lemma 14. This completes the proof of Lemma 8.

B.3 Proof of Lemma 9

Due to Lemma 15, i.e., the Jacobian attractive fixed point of F, and Claim 18, there exist constants $\delta > 0, c < 1$ such that $(1-\alpha_L^*+\delta)(1-\alpha_R^*+\delta)B^2 < 1$ and

$$|F(\alpha_L, \alpha_R) - (\alpha_L^*, \alpha_R^*)| \le c|(\alpha_L, \alpha_R) - (\alpha_L^*, \alpha_R^*)|,$$

for all $\alpha_L \in [\alpha_L^*-\delta, \alpha_L^*+\delta],\ \alpha_R \in [\alpha_R^*-\delta, \alpha_R^*+\delta]$. For the proof of Lemma 9, we assume that for some t, the event $\|\alpha(X_t) - (\alpha_L^*, \alpha_R^*)\|_\infty \le \delta$ occurs (initially at $t = 0$, it occurs) and introduce the following two lemmas.

Lemma 19 *Consider the bipartite Erdős-Rényi random graph $G(n, kn, p)$ where $p = \frac{B}{n\sqrt{k}}$ for some constants $B > 0$ and $k \ge 1$. Let $C_1, C_2, \ldots$ be connected components of G in decreasing order of size. Then, there exist constants $K_1, K_2 > 0$ such that*

a) for $B < 1$, we have

$$E\left[\sum_{i\ge1}|C_i|^2\right] \le K_1 n,$$

b) for $B > 1$, we have

$$E\left[\sum_{i\ge2}|C_i|^2\right] \le K_2 n,$$

Lemma 20 *Consider the Swendsen-Wang dynamics on the complete bipartite graph of size (n, kn) with some constant $k \ge 1$, $\beta_{uv} = -\frac{1}{2}\log\left(1 - \frac{B}{n\sqrt{k}}\right)$ for some constant $B > 2$ and $\gamma_v = 0$. Let $C_1, C_2, \ldots$ be connected components of G in decreasing order of size after the step 2 of the Swendsen-Wang dynamics. Then, given the event $\sum_{i\ge2}|C_i|^2 < wKn$ for some $w \ge 1$ and $K > 0$, it follows that*

$$\Pr\left(||C_1 \cap V_L| - \theta_L n|, ||C_1 \cap V_R| - \theta_R kn| \le w\sqrt{n}\right) \ge 1 - \frac{2K}{w} - \frac{1+k}{w^2},$$

where (θ_L, θ_R) is the unique positive solution of (10).

The proofs of Lemma 19 and Lemma 20 are presented in Section C.5 and Section C.6, respectively. From $(1-\alpha_L^*+\delta)(1-\alpha_R^*+\delta)B^2 < 1$, $\|\boldsymbol{\alpha}(X_t) - (\alpha_L^*, \alpha_R^*)\|_\infty \le \delta$ and Lemma 19, after the step 2 of the Swendsen-Wang dynamics (starting from X_t), we have

$$E\left[\sum_{i\ge2}|C_i|^2\right] \le Kn$$

for some constant K. Hence, by Markov's inequality, for any $w_t \geq 1$, we have

$$\Pr\left(\sum_{i\geq 2}|C_i|^2 < w_t K n\right) \geq 1 - 1/w_t. \tag{14}$$

We will decide the value of w_t later in this proof. Let's assume the event $\sum_{i\geq 2}|C_i|^2 < w_t K n$ occurs. Then, from Azuma's inequality, the number Z_i of vertices that receive spin i in $V \setminus C_1$ in the step 3 of the Swendsen-Wang dynamics is concentrated around its expectation as

$$\Pr\left(|Z_i \cap V_L - E[Z_i \cap V_L]| \geq w_t\sqrt{Kn}\right) \leq 2\exp(-w_t/2)$$
$$\Pr\left(|Z_i \cap V_R - E[Z_i \cap V_R]| \geq w_t\sqrt{Kn}\right) \leq 2\exp(-w_t/2).$$

Using union bound, one can achieve that

$$\Pr\left(|Z_i \cap V_j - E[Z_i \cap V_j]| \geq w_t\sqrt{Kn} \;\text{ for any } i \in \{-1,1\}, j \in \{L,R\}\right) \leq 8\exp(-w_t/2). \tag{15}$$

On the other hand, using Lemma 20, we can bound the deviation of the size of the giant component as

$$||C_1 \cap V_L| - \alpha_L(X_t)\theta_L n|, ||C_1 \cap V_R| - \alpha_R(X_t)\theta_R kn| \leq w_t\sqrt{n} \tag{16}$$

with probability at least

$$1 - \frac{U_1}{w_t} - \frac{U_2}{w_t^2}$$

for some constants $U_1, U_2 > 0$, where such U_1, U_2 exist as $\frac{1}{2k} \leq \frac{\alpha_R(X_t)kn}{\alpha_L(X_t)n} \leq 2k$. By combining (14), (15) and (16), we obtain

$$\|\boldsymbol{\alpha}(X_{t+1}) - F(\boldsymbol{\alpha}(X_t))\|_\infty \leq w_t(1+\sqrt{K})n^{-1/2} \tag{17}$$

with probability at least

$$(1-1/w_t)\left(1 - 8\exp\left(-\frac{w_t}{2}\right) - \frac{U_1}{w_t} - \frac{U_2}{w_t^2}\right).$$

Furthermore, by combining (17) and $|F(\alpha_L,\alpha_R) - (\alpha_L^*,\alpha_R^*)| \leq c|(\alpha_L,\alpha_R) - (\alpha_L^*,\alpha_R^*)|$, it follows that

$$\|\boldsymbol{\alpha}(X_{t+1}) - (\alpha_L^*,\alpha_R^*)\|_\infty \leq \frac{c+1}{2}\|\boldsymbol{\alpha}(X_t) - (\alpha_L^*,\alpha_R^*)\|_\infty \leq \delta \tag{18}$$

by setting w_t as

$$w_t := \frac{1-c}{2}\frac{n^{1/2}}{1+\sqrt{K}}\|\boldsymbol{\alpha}(X_t) - (\alpha_L^*,\alpha_R^*)\|_\infty \geq \frac{1-c}{2}\frac{L}{1+\sqrt{K}}.$$

Namely, $\|\boldsymbol{\alpha}(X_t) - (\alpha_L^*,\alpha_R^*)\|_\infty$ and w_t decrease with at least multiplicative factor $(c+1)/2$. Therefore, by applying the above arguments from $t = 0, 1, \ldots$, there exists $T = O(\log n)$ such that

$$\|\boldsymbol{\alpha}(X_T) - (\alpha_L^*,\alpha_R^*)\|_\infty \leq Ln^{-1/2},$$

with probability at least

$$\begin{aligned}
&\prod_{t=0}^{T-1}\left(1-\frac{1}{w_t}\right)\left(1 - 8\exp\left(-\frac{w_t}{2}\right) - \frac{U_1}{w_t} - \frac{U_2}{w_t^2}\right)\\
&\geq \prod_{t=0}^{T-1}\exp\left(-\frac{2s}{w_t}\right)\\
&\geq \prod_{t=0}^{\infty}\exp\left(-\frac{2s}{w_t}\right)\\
&= \exp\left(-\frac{4s}{1-c}\frac{1+\sqrt{K}}{L}\sum_{t=0}^{\infty}\left(\frac{1+c}{2}\right)^t\right)\\
&= \Theta(1),
\end{aligned}$$

where the first inequality is elementary to check by defining $s := \max(U_1, U_2+1, 10)$ and assuming large enough L so that $w_t \geq \max(U_1^2, (U_2+1)^2, 100)$, without loss of generality. This completes the proof of Lemma 9.

B.4 Proof of Lemma 10

In this proof, we prove Lemma 10 for the case $B > 2$. One can apply the same argument for the case $B < 2$. Let $\{V_L, V_R\}$, $|V_L| = n, |V_R| = kn$, be a partition of V such that $(u, v) \in E$ if and only if $u \in V_L, v \in V_R$ or $v \in V_L, u \in V_R$. By following the proof arguments of Lemma 5.7 in [28], one can show that after the step 2 of the Swendsen-Wang dynamics starting from X_0 (and Y_0), there exists a constant C such that the following event occurs with probability $1 - O(1/n)$: there are more than Cn isolated vertices in both V_L, V_R. Suppose the events happen from both X_0 and Y_0. Then, we choose exactly Cn isolated vertices in both V_L, V_R (from X_0, Y_0) and we consider the following coupling: in the step 3 of the Swendsen-Wang dynamics starting from X_0 and Y_0, assign spins to components except for the chosen isolated vertices. Let $\hat{X}_1, \hat{Y}_1$ denote the spin configurations except for the chosen isolated vertices. By applying the same arguments used for deriving (14)-(16), we obtain

$$\|\boldsymbol{\alpha}(\hat{X}_1) - (\alpha_L^* - C/2, \alpha_R^* - C/2)\|_\infty, \|\boldsymbol{\alpha}(\hat{Y}_1) - (\alpha_L^* - C/2, \alpha_R^* - C/2)\|_\infty \le \frac{1}{2} L' n^{-1/2}$$

for some constant L' with probability $\Theta(1)$. Then it holds that

$$\|\boldsymbol{\alpha}(\hat{X}_1) - \boldsymbol{\alpha}(\hat{Y}_1)\|_\infty \le L' n^{-1/2} \tag{19}$$

with probability $\Theta(1)$. Assume that the event (19) occurs. Now we show that there exists a coupling such that $\alpha_L(X_1) = \alpha_L(Y_1)$, $\alpha_R(X_1) = \alpha_R(Y_1)$ with probability $\Theta(1)$. In this proof, we only provide a coupling such that $\alpha_L(X_1) = \alpha_L(Y_1)$ with probability $\Theta(1)$, where one can easily extend the proof strategy to achieve $\alpha_R(X_1) = \alpha_R(Y_1)$.

Now we provide a joint distribution on isolated vertices of V_L in the step 3 of the Swendsen-Wang dynamics starting from X_0 and Y_0 so that $\alpha_L(X_1) = \alpha_L(Y_1)$ with probability $\Theta(1)$. Let $v_1, \dots, v_{Cn}$ denote the chosen isolated vertices without spin in V_L for both chains. For $1 \le j \le Cn$, let define

$$Z_j = \begin{cases} 1 & \text{if } X_1(v_j) = 1 \\ 0 & \text{otherwise} \end{cases} \qquad Z'_j = \begin{cases} 1 & \text{if } Y_1(v_j) = 1 \\ 0 & \text{otherwise} \end{cases}.$$

Let $Z = \sum_j Z_j, Z' = \sum_j Z'_j$. Now we show that one can couple the spin configuration of X_1 and Y_1 with so that $\alpha_L(X_1) = \alpha_L(Y_1)$ (and also $\alpha_R(X_1) = \alpha_R(Y_1)$) with probability $\Theta(1)$ and complete the proof. Consider $W \sim \mathrm{Bin}(Cn, 1/2)$. Then, the distribution of W is equivalent to the distribution of Z (and Z'). Let define a coupling (joint distribution) μ on Z, Z' such that

$$\mu(Z = w, Z = w - \ell) = \min(\Pr(Z = w), \Pr(Z = w - \ell))$$

for $w \in \left[\frac{Cn}{2}, \frac{Cn}{2} + L'\sqrt{n}\right]$ where $|\ell := n(\alpha_L(\hat{X}_1) - \alpha_L(\hat{Y}_1))| \le L'\sqrt{n}$. We remark that the construction of above coupling is equivalent to the coupling appears in Section 4.2 of [26]. The coupling μ results that

$$\mu(Z = Z' - \ell) \ge \sum_{w \in \left[\frac{Cn}{2}, \frac{Cn}{2} + L'\sqrt{n}\right]} \mu(Z = w, Z' = w - \ell). \tag{20}$$

We now aim for showing that

$$\Pr(W = w) = \Omega(n^{-1/2}) \tag{21}$$

for all $w \in \left[\frac{Cn}{2} - L'\sqrt{n}, \frac{Cn}{2} + L'\sqrt{n}\right]$, which leads to $\mu(Z = Z' - \ell) = \Theta(1)$ due to (20). For $w \in$

$\left[\frac{Cn}{2} - L'\sqrt{n}, \frac{Cn}{2} + L'\sqrt{n}\right]$, it follows that

$$\begin{aligned}
&\Pr(W = w) = \binom{Cn}{w}\left(\frac{1}{2}\right)^{Cn} \\
&\geq \binom{Cn}{\frac{Cn}{2} - L'\sqrt{n}}\left(\frac{1}{2}\right)^{Cn} \\
&= \Theta(1)\frac{\sqrt{Cn}\left(\frac{Cn}{e}\right)^{Cn}}{\sqrt{Cn-2L'\sqrt{n}}\left(\frac{Cn-2L'\sqrt{n}}{2e}\right)^{\frac{Cn-2L'\sqrt{n}}{2}}\sqrt{Cn+2L'\sqrt{n}}\left(\frac{Cn+2L'\sqrt{n}}{2e}\right)^{\frac{Cn+2L'\sqrt{n}}{2}}}\left(\frac{1}{2}\right)^{Cn} \\
&= \Theta(n^{-1/2})\frac{(Cn)^n}{(Cn-2L'\sqrt{n})^{\frac{Cn-2L'\sqrt{n}}{2}}(Cn+2L'\sqrt{n})^{\frac{Cn+2L'\sqrt{n}}{2}}} \\
&= \Theta(n^{-1/2})\frac{1}{\left(1-\frac{2L'\sqrt{n}}{Cn}\right)^{\frac{Cn-2L'\sqrt{n}}{2}}\left(1+\frac{2L'\sqrt{n}}{Cn}\right)^{\frac{Cn+2L'\sqrt{n}}{2}}} \\
&\geq \Theta(n^{-1/2})\frac{1}{e^{\frac{4L'^2}{C}}} \\
&= \Theta(n^{-1/2})
\end{aligned}$$

where the second equality follows from Stirling's formula. By combining (20) and (21), we obtain

$$\mu(Z = Z' - \ell) = \Theta(1)$$

and therefore there exists a coupling on (X_t, Y_t) such that $\alpha_L(X_1) = \alpha_L(Y_1)$ with probability $\Theta(1)$. This completes the proof of Lemma 10.

B.5 Proof of Lemma 12

From Lemma 17, we know that $(\alpha_L^*, \alpha_R^*) = (1/2, 1/2)$. Throughout this proof, we use $(1/2, 1/2)$ instead of (α_L^*, α_R^*). First, choose a constant $\delta > 0$ small enough so that $F(1/2+\delta, 1/2+\delta) = (1/2, 1/2)$, i.e. $(1/2+\delta, 1/2+\delta)$ is subcritical. Then, from Lemma 16, there exists a constant T such that $F^{(T)}([0,1]) \leq (1/2 + \delta/2, 1/2 + \delta/2)$. One can directly notice that that within T iterations of the Swendsen-Wang chain, the size of the larger spin class becomes less than $(1/2+\delta, 1/2+\delta)$ with probability $1-o(1)$ by Lemma 14. Furthermore, since $(1/2+\delta, 1/2+\delta)$ is subcritical, in the step 2 of the Swendsen-Wang dynamics at the next iteration, the larger spin class becomes subcritical, i.e. $\boldsymbol{\alpha}(X_{T+1}) = (1/2 + o(1), 1/2 + o(1))$ with probability $1 - o(1)$ by Lemma 14. Given the event $\boldsymbol{\alpha}(X_{T+1}) = (1/2+o(1), 1/2+o(1))$, after the step 2 of the Swendsen-Wang dynamics starting from X_{T+1} satisfies the following:

$$E\left[\sum_{i\geq 1} |C_i|^2\right] = O(n),$$

where we use Lemma 19 a). By applying the same arguments used for deriving (14) and (15), we have

$$X_{T+2} = (1/2 + O(n^{-1/2}), 1/2 + O(n^{-1/2})), \qquad \text{with probability } \Theta(1).$$

This completes the proof of Lemma 12.

C Proofs of Technical Lemmas

C.1 Proof of Lemma 1

We will show that the Swendsen-Wang dynamics induces a reversible Markov chain and has μ as its stationary distribution. To this end, we first introduce the equivalent representation of the Ising model

$$\begin{aligned}\mu(\sigma) &= Z^{-1}\exp\left(2\sum_{(u,v)\in E}\beta_{uv}\left(\delta_{\sigma_u,\sigma_v}-1\right)+2\sum_{v\in V}\gamma_v\delta_{\sigma_v,1}\right)\\ &= Z^{-1}\exp\left(2\sum_{v\in V}\gamma_v\delta_{\sigma_v,1}\right)\prod_{(u,v)\in E}\left((1-p_{uv})+p_{uv}\delta_{\sigma_u,\sigma_v}\right) \end{aligned} \tag{22}$$

where $p_{uv} = 1-\exp(-2\beta_{uv})$, $\delta_{x,y} = \begin{cases}1 & \text{if } x=y\\ 0 & \text{otherwise}\end{cases}$, and Z is the normalizing constant, called the *partition function*. Now, consider the following random cluster model on G having 'bond occupation' variables $m = [m_{uv}] = \{0,1\}^{|E|}$:

$$\mu_{RC}(m) = Z_{RC}^{-1}\prod_{u,v:m_{uv}=1}p_{uv}\prod_{u,v:m_{uv}=0}(1-p_{uv})\prod_{C\in\mathcal{C}}\left(1+\exp\left(2\sum_{v\in V(C)}\gamma_v\right)\right)$$

where $\mathcal{C}$ is the set of all connected components with respect to m and Z_{RC} is the partition function.

Let define a joint model of the Ising model and the random cluster model, which is called the Fortuin-Kasteleyn-Swendsen-Wang (FKSW) model [8]. A probability distribution of the FKSW model is defiend as below:

$$\mu_{\text{FKSW}}(\sigma,m) = Z_{\text{FKSW}}^{-1}\exp\left(2\sum_{v\in V}\gamma_v\delta_{\sigma_v,1}\right)\prod_{(u,v)\in E}\left((1-p_{uv})\delta_{m_{uv},0}+p_{uv}\delta_{m_{uv},1}\delta_{\sigma_u,\sigma_v}\right) \tag{23}$$

where Z_{FKSW} is the partition function. By summing (23) over σ or m, one can check the following facts about the FKSW model:

- $Z = Z_{\text{RC}} = Z_{\text{FKSW}}$.
- The marginal distribution on σ is μ.
- The marginal distribution on m is μ_{RC}.
- The conditional distribution of m given σ is as follows: set $m_{uv}=0$ if $\sigma_u \neq \sigma_v$ and set $m_{uv}=0,1$ with probability $1-p_{uv}, p_{uv}$, respectively, if $\sigma_u=\sigma_v$, i.e., (u,v) is a monochromatic edge.
- The conditional distribution of σ given m is as follows: for each connected component C, set all spins $[\sigma_v : v\in C]$ to 1 and -1 with probability $\frac{\exp\left(2\sum_{v\in V(C)}\gamma_v\right)}{1+\exp\left(2\sum_{v\in V(C)}\gamma_v\right)}$ and $\frac{1}{1+\exp\left(2\sum_{v\in V(C)}\gamma_v\right)}$, respectively.

The above observations imply that the Swendsen-Wang dynamics repeatedly samples m given σ and σ given m according to the distribution of FKSW model. Furthermore, one can easily verify that the Swendsen-Wang dynamics is reversible and has μ as its stationary distribution.

C.2 Proof of Lemma 15

In this proof, we first show that F has the unique fixed point $(1/2,1/2)$ for $B<2$ and (α_L^*,α_R^*) for $B>2$. Before starting the proof, we note that $\alpha_L<1/2, \alpha_R>1/2$ (or $\alpha_L>1/2,\alpha_R<1/2$) cannot be a solution of (12). To help the proof, we use the substitution $z_L = 2\alpha_L-1$ and $z_R=2\alpha_R-1$. By substituting z_L, z_R into (12), we have

$$z_L = \frac{\sqrt{k}}{B}\log\frac{1+z_R}{1-z_R} \qquad z_R = \frac{1}{B\sqrt{k}}\log\frac{1+z_L}{1-z_L}, \tag{24}$$

i.e. any fixed point of F must satisfies (24). First, consider the case that $B < 2$. One can easily check that $(1/2, 1/2)$ is a fixed point of F and $\alpha_L, \alpha_R < 1/2$ cannot be a fixed point of F. Now, suppose that there exists a solution $z_L, z_R > 0$ of (24), i.e. there exists $\alpha_L, \alpha_R > 1/2$ satisfying (12). Using the inequality $\log \frac{1+x}{1-x} > 2x$ for $x > 0$ and (24), we have

$$z_L > \frac{4}{B^2} z_L \qquad\qquad z_R > \frac{4}{B^2} z_R.$$

Since we assumed that $B < 2$, the above inequalities leads to contradiction and results that $(1/2, 1/2)$ is the only fixed point of F for $B < 2$. Now, consider the case that $B > 2$. We first define functions $g(x), y(x)$ as below:

$$y(x) := \frac{1}{B\sqrt{k}} \log \frac{1+x}{1-x} \qquad\qquad g(x) := \frac{\sqrt{k}}{B} \log \frac{1+y(x)}{1-y(x)}.$$

Then x is a fixed point of g if and only if $(z_L, z_R) = (x, y(x))$ is a solution of (24). Now we show that there exists the unique fixed point $x > 0$ of g. Suppose there exist two fixed points x_1, x_2 of g. By mean value theorem, there exists x' between x_1, x_2 such that $\frac{dg}{dx}(x') = 1$. However, the derivative of $g(x)$ with respect to x is

$$\frac{dg}{dx}(x) = \frac{4k}{1-x^2} \frac{1}{B^2 k - \log^2 \frac{1+x}{1-x}}$$

and at $x = x'$ we have

$$\frac{4k}{1-x^2} = B^2 k - \log^2 \frac{1+x}{1-x}. \tag{25}$$

One can observe that LHS of (25) is increasing with x but RHS of (25) is decreasing with x, i.e. there are at most two fixed points of g and therefore there are at most two solutions of (12). $(1/2, 1/2)$ is a solution of (12) but it is not a fixed point of F. However, since $F : [0,1]^2 \to [0,1]^2$ and F is continuous, by Brouwer's fixed point theorem, F has a fixed point. Furthermore, for $(\alpha_L, \alpha_R) \le (1/2, 1/2)$, we have $F(\alpha_L, \alpha_R) \ge (1/2, 1/2)$. Using this facts, one can conclude that F has a unique fixed point $(\alpha_L^*, \alpha_R^*) > (1/2, 1/2)$ for $B > 2$.

Now, we show that the fixed point of F is Jacobian attractive. Consider the Jacobian $D(F)$ of F

$$\begin{aligned} D(F) &= \begin{pmatrix} \frac{\partial F_L}{\partial \alpha_L} & \frac{\partial F_L}{\partial \alpha_R} \\ \frac{\partial F_R}{\partial \alpha_L} & \frac{\partial F_R}{\partial \alpha_R} \end{pmatrix} \\ &= \frac{1}{2} \frac{1}{1-(1-\theta_L)(1-\theta_R)B^2 \alpha_L \alpha_R} \begin{pmatrix} \theta_L & (1-\theta_L)\theta_R B\sqrt{k}\alpha_L \\ (1-\theta_R)\theta_L B \alpha_R/\sqrt{k} & \theta_R \end{pmatrix} \end{aligned} \tag{26}$$

where (θ_L, θ_R) is a solution of (10). For $B < 2$, $D(F)$ is a zero matrix at (1/2,1/2), i.e. the largest eigen value of $D(F)$ is zero. Therefore the fixed point of F is Jacobian attractive for $B < 2$. Suppose $B > 2$. Using (10) and by direct calculation of the largest eigen value, the largest eigen value λ of $D(F)$ can be bounded as below:

$$|\lambda| < \frac{1}{2} \frac{\theta_L + \theta_R}{1 - \frac{(1-\theta_L)(1-\theta_R)}{\theta_L \theta_R} \log(1-\theta_L)\log(1-\theta_R)}. \tag{27}$$

Since we are interested in λ at the fixed point, we only need to consider $\theta_L, \theta_R > 0$. Now we show that RHS of (27) is strictly smaller than 1 to prove that F is Jacobian attractive at (α_L^*, α_R^*). Consider the following function h

$$h(\theta_L, \theta_R) := 2 - \theta_L - \theta_R - 2\frac{(1-\theta_L)(1-\theta_R)}{\theta_L \theta_R} \log(1-\theta_L)\log(1-\theta_R).$$

One can notice that $h(\theta_L, \theta_R) > 0$ if and only if RHS of (27) is strictly smaller than 1. We bound h using the following claim.

Claim 21 *For $0 < x < 1$, the following inequality holds:*

$$-\frac{1-x}{x} \log(1-x) < \sqrt{1-x}.$$

Proof. 2 Let define

$$f(x) := \frac{\sqrt{1-x}}{x}\log(1-x).$$

We show that $-1 < f(x)$ for $0 < x < 1$ and complete the proof. We have $\lim_{x\to 0^+} f(x) = -1$. Furthermore, f is strictly increasing as

$$\frac{df}{dx}(x) = -\frac{2-x}{2x^2\sqrt{1-x}}\log(1-x) - \frac{1}{x\sqrt{1-x}} > 0$$

for $0 < x < 1$ where the last inequality can be verified by using the taylor series of $\log(1-x)$. This implies that $-1 < f(x)$ for $0 < x < 1$ and completes the proof of Claim 21.

Using Claim 21, we have

$$h(\theta_L, \theta_R) > 2 - \theta_L - \theta_R - 2\sqrt{(1-\theta_L)(1-\theta_R)} \geq 0$$

for $0 < \theta_L, \theta_R < 1$. This implies that RHS of (27) is strictly smaller than 1 and therefore $|\lambda| < 1$, i.e. (α_L^*, α_R^*) is Jacobian attractive fixedpoint of F for $B > 2$. This completes the proof of Lemma 15.

C.3 Proof of Lemma 16

In this proof, we first show that F is monotonically increasing function. From the formulation (26) of the Jacobian of F, every entries of $D(F)$ is non-negative, i.e. F is monotonically increasing, if and only if the following inequality holds

$$1 - (1-\theta_L)(1-\theta_R)B^2\alpha_L\alpha_R > 0. \tag{28}$$

Since $\theta_L = \theta_R = 0$ if and only if $\sqrt{\alpha_L\alpha_R}B \leq 1$, we only need to consider the case that $\sqrt{\alpha_L\alpha_R}B > 1$ (we can ignore the case $\sqrt{\alpha_L\alpha_R} = 1$ for proving that F is monotonically increasing as F is continuous). Using (10), LHS of (28) can be represented as

$$1 - \frac{(1-\theta_L)(1-\theta_R)}{\theta_L\theta_R}\log(1-\theta_L)\log(1-\theta_R).$$

By Claim 21, we have

$$1 - \frac{(1-\theta_L)(1-\theta_R)}{\theta_L\theta_R}\log(1-\theta_L)\log(1-\theta_R) > 1 - \sqrt{(1-\theta_L)(1-\theta_R)} > 0$$

for $0 < \theta_L, \theta_R < 1$. This results that F is monotonically increasing.

Since F is monotonically increasing, $F^{(t)}(0,0) \leq F^{(t)}(\alpha_L, \alpha_R) \leq F^{(t)}(1,1)$ for any (α_L, α_R), i.e. it is enough to show that sequences $[F^{(t)}(0,0)]_t$ and $[F^{(t)}(1,1)]_t$ converge to the fixed point of F. Let (α_L^*, α_R^*) be the fixed point of F. From the definition of F, we have $F(1,1) \leq (1,1)$. Using the monotonicity, we have $F^{(2)}(1,1) \leq F(1,1)$. By applying this argument repeatedly, one can argue that $[F^{(t)}(1,1)]_t$ is a decreasing sequence and bounded below by the fixed point of F. By the monotone convergence theorem and lemma 15, $[F^{(t)}(1,1)]_t$ converges to the fixed point of F. Similarly $[F^{(t)}(0,0)]_t$ converges to the fixed point of F. This completes the proof of Lemma 16.

C.4 Proof of Lemma 17

We first formulate the probability that a phase (α_L, α_R) occurs. This probability can be formulated as follows:

$$\begin{aligned}
\Pr(\alpha_L, \alpha_R) &\propto \binom{n}{\alpha_L n}\binom{kn}{\alpha_R kn}\left(1 - \frac{B}{n\sqrt{k}}\right)^{kn^2(\alpha_L(1-\alpha_R)+\alpha_R(1-\alpha_L))} \\
&\approx \frac{1}{2\pi n\sqrt{\alpha_L(1-\alpha_L)\alpha_R(1-\alpha_R)k}}\alpha_L^{-\alpha_L n}(1-\alpha_L)^{-(1-\alpha_L)n} \\
&\qquad \times \alpha_R^{-\alpha_R kn}(1-\alpha_R)^{-(1-\alpha_R)kn}\exp\left(-Bn\sqrt{k}(\alpha_L(1-\alpha_R)+\alpha_R(1-\alpha_L))\right) \\
&= \frac{1}{2\pi n\sqrt{\alpha_L(1-\alpha_L)\alpha_R(1-\alpha_R)k}}\exp\left(n\sqrt{k}\psi(\alpha_L, \alpha_R)\right)
\end{aligned}$$

where we use Stirling's formula for the second line and ψ is defined as

$$\psi(\alpha_L, \alpha_R) := - B(\alpha_L + \alpha_R - 2\alpha_L\alpha_R) - \frac{\alpha_L}{\sqrt{k}}\log \alpha_L - \frac{1-\alpha_L}{\sqrt{k}}\log(1-\alpha_L) \\ - \sqrt{k}\alpha_R \log \alpha_R - \sqrt{k}(1-\alpha_R)\log(1-\alpha_R).$$

Since ψ is in the exponent of e, the phase achieves the maximum value of ψ should be the maximum a posteriori phase of the Ising model asymptotically. Now we analyze the phase (α_L, α_R) maximizing ψ. By taking partial derivative of ψ with respect to α_L and α_R, we have

$$\frac{\partial\psi(\alpha_L,\alpha_R)}{\partial\alpha_L} = -B(1-2\alpha_R) - \frac{1}{\sqrt{k}}\log\alpha_L + \frac{1}{\sqrt{k}}\log(1-\alpha_L) \\ \frac{\partial\psi(\alpha_L,\alpha_R)}{\partial\alpha_R} = -B(1-2\alpha_L) - \sqrt{k}\log\alpha_R + \sqrt{k}\log(1-\alpha_R).$$

By simple calculation, one can check that $\frac{\partial\psi(\alpha_L,\alpha_R)}{\partial\alpha_L} = \frac{\partial\psi(\alpha_L,\alpha_R)}{\partial\alpha_R} = 0$ if and only if the following relation holds

$$\exp\left(B\sqrt{k}(1-2\alpha_R)\right) = \frac{1-\alpha_L}{\alpha_L} \qquad \exp\left(\frac{B}{\sqrt{k}}(1-2\alpha_L)\right) = \frac{1-\alpha_R}{\alpha_R} \tag{29}$$

which is equivalent to (12). One can easily check that $\alpha_L = \alpha_R = 1/2$ is a solution of (29). If (α_L, α_R) is a solution of (29), then $(1-\alpha_L, 1-\alpha_R)$ is a solution of (29). Furthermore, LHS and RHS of the first (and the second) equation of (29) are decresing with respect to α_R, α_L (and α_L, α_R) respectively. Since $(1/2, 1/2)$ is a solution of (29), any solution (α_L, α_R) of (29) satisfies $\alpha_L, \alpha_R \geq 1/2$ or $\alpha_L, \alpha_R \leq 1/2$. Therefore, we only consider critical points of ψ in $[1/2, 1]^2$. In the proof of Lemma 15 we have shown that (29) has the only solution $(1/2, 1/2)$ for $B \leq 2$ and (29) has only two solutions $(1/2, 1/2)$, (α_L^*, α_R^*) for $B > 2$. Now we show that $(1/2, 1/2)$, (α_L^*, α_R^*) achieve the maximum value of ψ for $B \leq 2, B > 2$ respectively by showing that the Hessian of ψ is negative semidefinite at $(1/2, 1/2)$, (α_L^*, α_R^*) for $B \leq 2, B > 2$ respectively. The hessian $H(\psi)$ of ψ is as follows

$$H(\psi) = \begin{pmatrix} -\frac{1}{\alpha_L(1-\alpha_L)\sqrt{k}} & 2B \\ 2B & -\frac{\sqrt{k}}{\alpha_R(1-\alpha_R)} \end{pmatrix}.$$

By simple calculations, one can check that $H(\psi)$ is negative semidefinite if and only if

$$2B \leq \sqrt{\frac{1}{\alpha_L(1-\alpha_L)\alpha_R(1-\alpha_R)}}. \tag{30}$$

Since (30) holds for any $B \leq 2$, $(1/2, 1/2)$ maximizes ψ.

Now we show that (α_L^*, α_R^*) maximizes ψ for $B > 2$. Consider $H(\psi)$ at $(1/2, 1/2)$ and (α_L^*, α_R^*). $H(\psi)$ is negative semidefinite if and only if (30) holds. However, $(1/2, 1/2)$ does not satisfies (30) and therefore $(1/2, 1/2)$ is not a local maximum of F. Let $z_L^* = 2\alpha_L^* - 1$ and $z_R^* = 2\alpha_R^* - 1$. Then (30) at (α_L^*, α_R^*) is equivalent to

$$\frac{1}{4}\frac{(1-z_L^{*2})(1-z_R^{*2})}{z_L^* z_R^*}\log\frac{1+z_L^*}{1-z_L^*}\log\frac{1+z_R^*}{1-z_R^*} \leq 1 \tag{31}$$

where we additionally use the fact that (z_L^*, z_R^*) is a solution of (24). Let define $h(x) := \frac{1-x^2}{x}\log\frac{1+x}{1-x}$. We have $\lim_{x\to 0^+} h(x) = 2$. The derivative of h is strictly negative as

$$\frac{dh}{dx}(x) = -\frac{1+x^2}{x^2}\log\frac{1+x}{1-x}\frac{2}{x} < -2x \leq 0 \tag{32}$$

where we use an inequality $\log\frac{1+x}{1-x} > 2x$ for $x > 0$. Since (32) and $\lim_{x\to 0^+} h(x) = 2$ implies

$$\frac{1}{4}\frac{(1-z_L^{*2})(1-z_R^{*2})}{z_L^* z_R^*}\log\frac{1+z_L^*}{1-z_L^*}\log\frac{1+z_R^*}{1-z_R^*} < 1 \tag{33}$$

and this implies (31), (α_L^*, α_R^*) is the only local maximum of ψ on $[1/2, 1]^2$ for $B > 2$. Recall that every local maximum point (α_L, α_R) of ψ satisfies that $\alpha_L, \alpha_R \geq 1/2$ or $\alpha_L, \alpha_R \leq 1/2$. This implies that $(\alpha_L^*, \alpha_R^*), (1 - \alpha_L^*, 1 - \alpha_R^*)$ are only local maxima of ψ, i.e. $(1-\alpha_L^*, 1-\alpha_R^*)$ achieves maximum of ψ in $[0, 1/2] \times [0, 1]$ and (α_L^*, α_R^*) achieves maximum of ψ in $[1/2, 1] \times [0, 1]$ for $B > 2$. By Using this, one can conclude that $(\alpha_L^*, \alpha_R^*), (1-\alpha_L^*, 1-\alpha_R^*)$ achieve the maximum of ψ for $B > 2$. This completes the proof of Lemma 17.

C.5 Proof of Lemma 19

We first prove Lemma 19 a). In order to bound component sizes of G, we concern a random process called a 'branching process' on a graph. The branching process is a sampling procedure which samples a connected component of a bipartite Erdős-Rényi random graph $G(n, kn, p)$ where $p = \frac{B}{n\sqrt{k}}$. The branching process on the complete bipartite graph (V_L, V_R, E) with $|V_L| = n, |V_R| = kn$ can be described as follows:

1. Choose $u_0 \in V_L$ and initialize $S_L = S_R = \emptyset$, $W_L = u_0, W_R = \emptyset$ and an iteration number $t = 0$.
2. Set $t \leftarrow t + 1$. Choose $u_i \in W_L$ and choose random neighbors $v_1, \dots v_{r_i}$ of u_i from $V_R - S_R - W_R$ where each neighbor of u_i is chosen with probability $\frac{B}{n\sqrt{k}}$. Set $W_R = W_R \cup \{v_1, \dots v_{r_i}\}$, $W_L = W_L - \{u_i\}$ and $S_L = S_L \cup \{u_i\}$.
3. For each $v_j \in W_R$, choose random neighbors $u_{j1}, \dots, u_{js_j}$ of v_j from $V_L - S_L - W_L$ where each neighbor of v_j is chosen with probability $\frac{B}{n\sqrt{k}}$. Set $W_L = W_L \cup \{u_{j1}, \dots, u_{js_j}\}$, $W_R = W_R - \{v_j\}$ and $S_R = S_R \cup \{v_j\}$. Repeat the step 3 until $W_R = \emptyset$.
4. Repeat step 2-3 until $W_L \cup W_R = \emptyset$.

For each t-th iteration, let define a random variable $K_t := |W_L|$ at the beginning of the step 4 of the branching process. Then the stopping time $\arg\min_t (K_t = 0)$ decides the number of vertices in V_L in the component of $G(n, kn, p)$ containing u_0. One can observe that K_t is bounded above by the random variable $(\sum_{i=0}^{t} R_i) - t$ where $R_0 = 1$ and $R_i \sim \text{Bin}\left(\text{Bin}\left(n, \frac{B}{n\sqrt{k}}\right) kn, \frac{B}{n\sqrt{k}}\right)$. Similarly, one can construct the branching process starting from $u_0 \in V_R$ and define K_t' as K_t. Then K_t' is bounded above by $(\sum_{i=0}^{t} R_i') - t$ where $R_0' = 1$, $R_i' \sim \text{Bin}\left(\text{Bin}\left(kn, \frac{B}{n\sqrt{k}}\right) n, \frac{B}{n\sqrt{k}}\right)$.

Let $C(v)$ of G be a component containing v. Observe that

$$E\left[\sum_{i \geq 1} |C_i|^2\right] = E\left[\sum_{v \in V_L \cup V_R} |C(v)|\right] = (1+k)nE\left[|C(v)|\right].$$

Now we show that $E\left[|C(v)|\right] = O(1)$ and complete the proof. Let define stopping times τ, τ' as below:

$$\tau := \arg\min_t \left(\left(\sum_{i=0}^{t} R_i\right) - t = 0\right) \qquad \tau' := \arg\min_t \left(\left(\sum_{i=0}^{t} R_i'\right) - t = 0\right).$$

Since K_t, K_t' are bounded above by $(\sum_{i=0}^{t} R_i) - t$, $(\sum_{i=0}^{t} R_i') - t$ respectively, we have

$$E\left[|C(v)|\right] \leq E[\tau] + E[\tau'].$$

However, by the direct application of Wald's lemma, we can conclude that

$$E[\tau], E[\tau'] = O(1)$$

and this completes the proof of Lemma 19 a).

Now we prove Lemma 19 b). We first introduce the following claim.

Claim 22 *For $B > 1$, $(1-\theta_L)(1-\theta_R)B^2 < 1$ where θ_L, θ_R are solution of (9), i.e. the rest part except for the giant component is subcritical.*

Proof. 3 Using (9), $(1-\theta_L)(1-\theta_R)B^2$ reduces to

$$(1-\theta_L)(1-\theta_R)B^2 = \frac{(1-\theta_L)(1-\theta_R)}{\theta_L\theta_R}\log(1-\theta_L)\log(1-\theta_R).$$

By applying Claim 21 to the RHS of the above identity, we completes the proof of Claim 22.

By Claim 22, we know that the induced subgraph of vertices which are not in the giant component, C_1, is subcritical. Let $\varepsilon > 0$ be a small enough constant which satisfies that

$$(1-\theta_L+\varepsilon)(1-\theta_R+\varepsilon)B^2 < 1. \tag{34}$$

By following the proof of Theorem 9 of [21] and applying Azuma's inequality, one can conclude that

$$\Pr\left(|C_1 \cap V_L| < (\theta_L - \varepsilon)n\right) < e^{-\Omega(n)}$$
$$\Pr\left(|C_1 \cap V_R| < (\theta_R - \varepsilon)kn\right) < e^{-\Omega(n)}$$

for some constant c. Let $\mathcal{E}$ be an event that $|C_1 \cap V_L| > (\theta_L - \varepsilon)n, |C_1 \cap V_R| > (\theta_R - \varepsilon)kn$. Then $\Pr(\mathcal{E}) = 1 - e^{-\Omega(n)}$. Now we utilize the following identity

$$\begin{aligned} E\left[\sum_{i\geq 2}|C_i|^2\right] &= E\left[\sum_{i\geq 2}|C_i|^2 \,\middle|\, \mathcal{E}\right]\Pr(\mathcal{E}) + E\left[\sum_{i\geq 2}|C_i|^2 \,\middle|\, \bar{\mathcal{E}}\right]\Pr(\bar{\mathcal{E}}) \\ &= E\left[\sum_{i\geq 2}|C_i|^2 \,\middle|\, \mathcal{E}\right](1-o(1)) + o(1). \end{aligned}$$

By (34) and a) of Lemma 19, we have

$$E\left[\sum_{i\geq 2}|C_i|^2 \,\middle|\, \mathcal{E}\right] = O(n).$$

This directly implies the statement of Lemma 19 b) and completes the proof of Lemma 19.

C.6 Proof of Lemma 20

Call $v \in V_L \cup V_R$ 'small' if v is not in the giant component. For each $v \in V_L \cup V_R$, let define binary random variable S_v as $S_v = 1$ if v is small, $S_v = 0$ otherwise. Let define $S_L := \sum_{v\in V_L} S_v$ and $S_R := \sum_{v \in V_R} S_v$. From Lemma 13, we know that

$$\Pr(S_v = 1) = 1 - \theta_L \qquad v \in V_L.$$

Our goal is to bound the variance of S_L, S_R to bound the variance of the giant component. We bound the second moment of S_L as below:

$$\begin{aligned} E\left[S_L^2\right] &= E\left[\sum_{v\in V_L} S_v^2\right] \\ &= \sum_{v\in V_L} E\left[S_v^2\right] + \sum_{\substack{u\neq v \\ u,v\in V_L}} E[S_uS_v] \\ &= E[S_L] + \sum_{\substack{u\neq v \\ u,v\in V_L}} \Pr(u, v \text{ are small}) \\ &= E[S_L] + \sum_{v\in V_L}\Pr(v \text{ is small}) \sum_{\substack{u\neq v \\ u\in V_L}} \Pr(u \text{ is small}|v \text{ is small}) \end{aligned}$$

One can reinterpret $\sum_{\substack{u\neq v\\ u,v\in V_L}} \Pr(u \text{ is small}|v \text{ is small})$ as

$$\begin{aligned}
&\sum_{\substack{u\neq v\\ u\in V_L}} \Pr(u \text{ is small}|v \text{ is small}) \\
&= \sum_{\substack{u\neq v:\ u\in V_L\\ u,v \text{ are in same}\\ \text{component}}} \Pr(u \text{ is small}|v \text{ is small}) + \sum_{\substack{u\neq v:\ u\in V_L\\ u,v \text{ are in different}\\ \text{component}}} \Pr(u \text{ is small}|v \text{ is small}).
\end{aligned}$$

However, we have

$$\begin{aligned}
\sum_{v\in V_L} \Pr(v \text{ is small}) \sum_{\substack{u\neq v:\ u\in V_L\\ u,v \text{ are in same}\\ \text{component}}} \Pr(u \text{ is small}|v \text{ is small}) &\leq \sum_{v\in V_L} (|C(v)|-1) \\
&\leq \sum_{i\geq 2} |C_i|(|C_i|-1) \qquad (35)\\
&\leq wKn
\end{aligned}$$

where $C(v)$ is a component containing a small vertex v. The last inequality of (35) follows from the assumption $\sum_{i\geq 2}|C_i|^2 < wKn$. For u, v are in different components, asymptotically we have

$$\Pr(u \text{ is small}|v \text{ is small}) = \Pr(u \text{ is small}) = 1-\theta_L \qquad (36)$$

as $|C(v)| = O(\log^2 n)$ for small vertex v by Lemma 13. Combining (35) and (36) results

$$\begin{aligned}
E[S_L^2] &= E[S_L] + \sum_{v\in V_L} \Pr(v \text{ is small}) \sum_{u\neq v} \Pr(u \text{ is small}|v \text{ is small}) \\
&\leq (1-\theta_L)n + wKn + (1-\theta_L)^2 n^2. \qquad (37)
\end{aligned}$$

(37) directly leads to

$$\mathrm{Var}(S_L) \leq (1-\theta_L+wK)n.$$

Using Chebyshev's inequality, we bound the deviation of S_L from its expectation as

$$\Pr(|S_L-(1-\theta_L)n| \geq w\sqrt{n}) \leq \frac{1-\theta_L+wK}{w^2}. \qquad (38)$$

One can apply the similar argument for V_R and achieve

$$\Pr(|S_R-(1-\theta_R)kn| \geq w\sqrt{kn}) \leq \frac{(1-\theta_R)k+wK}{kw^2}. \qquad (39)$$

Combining (38), (39) results

$$\Pr(\{|S_L-(1-\theta_L)n| \geq w\sqrt{n}\} \cup \{|S_R-(1-\theta_R)kn| \geq w\sqrt{n}\}) \leq \frac{2K}{w} + \frac{1+k}{w^2}.$$

This completes the proof of Lemma 20