Web Appendix for "Convergence Rate of Markov Chain Methods for Genomic Motif Discovery" by D. B. Woodard and J. S. Rosenthal

C.1 List of Symbols

Here is a list of symbols used in the main manuscript and in this Web Appendix.

- w: fixed motif length.
- L: length of the observed nucleotide sequence $\mathbf{S}$.
- M: known number of nucleotide types (typically =4 in practice).
- J: number of motifs in the generative model (defined in Assumption 3.2Slow Mixing for Multiple True Motifsassumption.3.2)
- p_0: fixed motif frequency in the inference model (defined Section 2.1Statistical Motif Discoverysubsection.2.1).
- $\mathbf{S} = (S_1, \ldots, S_L)$: observed sequence of nucleotides (defined Sec. 2.1Statistical Motif Discoverysubsection.2.1).
- $\mathbf{A} = (A_1, \ldots, A_{L/w})$: unknown vector of motif indicators (defined Sec. 2.1Statistical Motif Discoverysubsection.2.1).
- $\mathcal{X} = \{0,1\}^{L/w}$: space of possible values for $\mathbf{A}$ (defined in Sec. 2.1Statistical Motif Discoverysubsection.2.1).
- $\boldsymbol{\theta}_0$: unknown length-M vector of background nucleotide frequencies (defined Sec. 2.1Statistical Motif Discoverysubsection.2.1).
- $\boldsymbol{\theta}_{1:w} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_w)$: unknown matrix of position-specific nucleotide frequencies within the motif, where $\boldsymbol{\theta}_k$ has length M (defined Sec. 2.1Statistical Motif Discoverysubsection.2.1).
- $\mathbf{N}(\mathbf{A}^c)$; $\mathbf{N}(\mathbf{A}^{(k)})$; $\mathbf{N}(\mathbf{S})$: length-M nucleotide count vectors defined in (2.1Statistical Motif Discoveryequation.2.1).
- $\mathbf{A}_{[-i]}$: vector $\mathbf{A}$ with ith element removed; $\mathbf{A}_{[i,0]}, \mathbf{A}_{[i,1]}$: vector $\mathbf{A}$ with ith element replaced by 0 or 1, respectively.
- $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_w$: fixed length-$M$ vectors of constants (hyperparameters) used in the prior distribution of $\boldsymbol{\theta}_{0:w}$ (defined Sec. 2.1Statistical Motif Discoverysubsection.2.1).

- $p_1, \ldots, p_J$: as part of the generative model, the frequencies of the different "true" motifs (defined in Assumption 3.2Slow Mixing for Multiple True Motifsassumption.3.2).

- $\boldsymbol{\theta}_0^*$: as a part of the generative model, the true value of $\boldsymbol{\theta}_0$ (defined in Assumption 3.2Slow Mixing for Multiple True Motifsassumption.3.2).

- $\boldsymbol{\theta}_{1:w}^{j*} : j \in \{1, \ldots, J\}$: as a part of the generative model, the multiple "true" values of the matrix $\boldsymbol{\theta}_{1:w}$ (defined in Assumption 3.2Slow Mixing for Multiple True Motifsassumption.3.2).

- $\mathbf{Gap}(T)$: the spectral gap of a transition matrix T (defined in Section 2.3Markov Chain Convergence Ratessubsection.2.3).

- $\pi(\ldots)$: the likelihood, the prior, or the full, marginal, or conditional posterior distributions of the parameters, as distinguished by the arguments.

- $\mathbf{C}(\mathbf{A})$; $\mathbf{C}(\mathbf{S})$: length-2^w vectors of counts (defined in (5.3Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.3) and (5.4Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.4)).

- $\bar{\mathcal{X}}$: space of possible values for $\mathbf{C}(\mathbf{A})$ (defined in (5.5Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.5)).

- $\bar{\pi}(\mathbf{c}|\mathbf{S})$: the marginal posterior distribution of $\mathbf{C}(\mathbf{A})$, sometimes written with the dependence on $\mathbf{S}$ suppressed (defined in (5.7Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.7)).

- T: the Markov transition matrix (2.6Statistical Motif Discoveryequation.2.6) associated with the Gibbs sampler; $\bar{T}$: the projection matrix (5.9Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.9) associated with the summary vector $\mathbf{C}(\mathbf{A})$.

C.2 Proof of Lemma 3.1Slow Mixing for Multiple True Motifslemma.3.1

For notational simplicity we give the proof for the case $M = 2$. With this choice, recall from (5.24Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.24) that the free parameters in $\boldsymbol{\theta}_{0:w}$ are $\theta_{k,1} \in [0,1]$ for $k \in \{0, \ldots, w\}$, so we can write $\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1}$ and $\boldsymbol{\theta}_{1:w} \in [0,1]^w$.

Let $\sum p_j$ be shorthand for $\sum_{j=1}^J p_j$. Define

$$\phi \triangleq \min\left\{\frac{(1-\sum p_j)\,\theta_{0,1}^*}{1-p_1},\;\; 1-\left[\frac{(1-\sum p_j)\,\theta_{0,1}^* + \sum_{j=2}^J p_j}{1-p_1}\right]\right\}. \tag{C.1}$$

By Assumption 3.2Slow Mixing for Multiple True Motifsassumption.3.2 $\theta_{0,1}^* \in (0,1)$, $p_j > 0$, and $\sum p_j < 1$, so

$$\phi \in \left(0,\; \min\{\theta_{0,1}^*, 1-\theta_{0,1}^*\}\right). \tag{C.2}$$

Using (3.4Slow Mixing for Multiple True Motifsequation.3.4), define

$$\zeta \triangleq (\phi/4)^{\max\{4/\phi, 2/a\}} < \phi/4 < 1/4. \tag{C.3}$$

The constants $\phi, \zeta \in (0,1)$ do not depend on w. Then, for any $w \in \{1,2,\ldots\}$ and $j \in \{1,\ldots,J\}$ define

$$H_w^j \triangleq \left\{\boldsymbol{\theta}_{1:w} \in [0,1]^w : |\theta_{k,1} - \theta_{k,1}^{j*}| \leq \zeta \ \ \forall k \in \{1,\ldots,w\}\right\}. \tag{C.4}$$

$$B_w^j \triangleq \left\{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1} : \boldsymbol{\theta}_{1:w} \in H_w^1,\ \theta_{0,1} \in [\phi - \zeta, 1 - \phi + \zeta]\right\}. \tag{C.5}$$

Since $\phi - \zeta > 0$, the interval $[\phi - \zeta, 1 - \phi + \zeta]$ is bounded away from zero and one. By Assumption 3.3Slow Mixing for Multiple True Motifsassumption.3.3, for w large enough and all $j, j' \in \{1,\ldots,J\}$ with $j \neq j'$ there is some $k \in \{1,\ldots,w\}$ such that $t_k^j \neq t_k^{j'}$. For this k we have $\theta_{k,1}^{j*} = 1 - \theta_{k,1}^{j'*}$, so $|\theta_{k,1}^{j*} - \theta_{k,1}^{j'*}| = 1 > 2\zeta$. So B_w^j and $B_w^{j'}$ are disjoint.

Next we find a point $\boldsymbol{\theta}_{0:w}^{(1)} \in B_w^1$ such that $\sup_{\partial B_w^1} \eta < \eta(\boldsymbol{\theta}_{0:w}^{(1)})$. Then for any $j \neq 1$, $\exists \boldsymbol{\theta}_{0:w}^{(j)} \in B_w^j$ with $\sup_{\partial B_w^j} \eta < \eta(\boldsymbol{\theta}_{0:w}^{(j)})$ by symmetry, showing that (3.1Slow Mixing for Multiple True Motifsequation.3.1) holds.

Also define

$$\begin{aligned} h_w(\boldsymbol{\theta}_{0:w}) \triangleq &\sum_{\mathbf{s}\in\{1,2\}^w} \left[p_1 \prod_{k=1}^{w} \theta_{k,s_k}^{1*}\right] \log\left[p_0 \prod_{k=1}^{w} \theta_{k,s_k}\right] \\ &+ \sum_{\mathbf{s}\in\{1,2\}^w} \left[\sum_{j=2}^{J} p_j \prod_{k=1}^{w} \theta_{k,s_k}^{j*} + (1 - \sum p_j) \prod_{k=1}^{w} \theta_{0,s_k}^{*}\right] \log\left[(1-p_0) \prod_{k=1}^{w} \theta_{0,s_k}\right] \end{aligned} \tag{C.6}$$

and note that

$$\partial B_w^1 = \mathrm{cl}(B_w^1) \cap \mathrm{cl}\big([0,1]^{w+1} \backslash B_w^1\big) \qquad \subset B_w^1 \tag{C.7}$$

since B_w^1 is closed. By (C.4)-(C.5),

$$\partial B_w^1 \subset \{\boldsymbol{\theta}_{0:w} : \theta_{0,1} \in \{\phi - \zeta, 1 - \phi + \zeta\}\} \ \cup\ \left\{\boldsymbol{\theta}_{0:w} : \exists k : |\theta_{k,1} - \theta_{k,1}^{1*}| = \zeta\right\}. \tag{C.8}$$

Lemma C.1 below shows that $h_w(\boldsymbol{\theta}_{0:w})$ is maximized at $(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*}) \in B_w^1$ for some $\hat{\boldsymbol{\theta}}_0$. We will show that

$$\inf_{\boldsymbol{\theta}_{0:w} \in \partial B_w^1} \left[E \log f(\mathbf{s}|(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})) - E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})\right] \qquad > 0. \tag{C.9}$$

Lemma C.1 shows that $\exists b > 0$ such that for any w,

$$\inf_{\boldsymbol{\theta}_{0:w} \in \partial B_w^1} \left[h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*}) - h_w(\boldsymbol{\theta}_{0:w})\right] \qquad > b > 0. \tag{C.10}$$

For any constants a_1, a_2, b_1, b_2 we have that $a_1 - a_2 \geq b_1 - b_2 - |a_1 - b_1| - |a_2 - b_2|$. So for any $\boldsymbol{\theta}_{0:w} \in \partial B_w^1$,

$$
\begin{aligned}
&E \log f(\mathbf{s}|(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})) - E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \\
&\geq h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*}) - h_w(\boldsymbol{\theta}_{0:w}) - \left|E \log f(\mathbf{s}|(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})) - h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})\right| \\
&\quad - \left|E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) - h_w(\boldsymbol{\theta}_{0:w})\right|.
\end{aligned}
$$

Combining this with (C.7), (C.10), and Lemma C.2 below, for w large enough and any $\boldsymbol{\theta}_{0:w} \in \partial B_w^1$

$$
E \log f(\mathbf{s}|(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})) - E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) > b - b/4 - b/4 \qquad = b/2.
$$

So (C.9) holds for w large enough, proving Lemma 3.1Slow Mixing for Multiple True Motifslemma.3.1. □

Finally, we give the results used in the proof of Lemma 3.1Slow Mixing for Multiple True Motifslemma.3.1.

Lemma C.1. *Under Assumptions 3.1Slow Mixing for Multiple True Motifsassumption.3.1-3.3Slow Mixing for Multiple True Motifsassumption.3.3, for any w the function $h_w(\boldsymbol{\theta}_{0:w})$ defined in (C.6) is maximized at $(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}^{1*})$ where*

$$
\begin{aligned}
\hat{\theta}_{0,1} &\triangleq \frac{w\,(1 - \sum p_j)\,\theta_{0,1}^* + \sum_{j=2}^J p_j \sum_{k=1}^w \theta_{k,1}^{j*}}{w(1-p_1)} \\
&\in [\phi, 1-\phi].
\end{aligned}
\tag{C.11}
$$

Also, using the definitions (C.5) and (C.7), Equation (C.10) holds for some b that does not depend on w.

Proof. For $\mathbf{s} \in \{1,2\}^w$ and $m \in \{1,2\}$ let $\#\{s_k = m\}$ denote the number of indices $k \in$

$\{1, \ldots, w\}$ for which $s_k = m$. Then

$$\frac{\partial}{\partial \theta_{k,1}} h_w(\boldsymbol{\theta}_{0:w}) = \sum_{\mathbf{s}} \left[p_1 \prod_{k'=1}^{w} \theta^{1*}_{k',s_{k'}} \right] \left[\frac{\mathbf{1}_{\{s_k=1\}}}{\theta_{k,1}} - \frac{\mathbf{1}_{\{s_k=2\}}}{1-\theta_{k,1}} \right] \qquad k \in \{1, \ldots, w\}$$

$$= \frac{p_1 \theta^{1*}_{k,1}}{\theta_{k,1}} - \frac{p_1(1-\theta^{1*}_{k,1})}{1-\theta_{k,1}} \tag{C.12}$$

$$\frac{\partial}{\partial \theta_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) = \sum_{\mathbf{s}} \left[\sum_{j=2}^{J} p_j \prod_{k=1}^{w} \theta^{j*}_{k,s_k} + (1 - \sum p_j) \prod_{k=1}^{w} \theta^*_{0,s_k} \right] \left[\frac{\#\{s_k = 1\}}{\theta_{0,1}} - \frac{\#\{s_k = 2\}}{1-\theta_{0,1}} \right]$$

$$= \frac{1}{\theta_{0,1}} \left(\sum_{j=2}^{J} p_j \sum_{k=1}^{w} \theta^{j*}_{k,1} + w(1 - \sum p_j) \theta^*_{0,1} \right) \tag{C.13}$$

$$- \frac{1}{1-\theta_{0,1}} \left(\sum_{j=2}^{J} p_j \sum_{k=1}^{w} (1 - \theta^{j*}_{k,1}) + w(1 - \sum p_j)(1 - \theta^*_{0,1}) \right).$$

Setting this equal to zero and solving for $\theta_{0,1}$ and $\theta_{k,1}$ shows that $h_w(\boldsymbol{\theta}_{0:w})$ has a stationary point at $(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}^{1*}_{1:w})$. Using (C.1), $\hat{\theta}_{0,1} \in [\phi, 1-\phi]$.

Note that $\frac{\partial^2}{\partial \theta_{k,1} \partial \theta_{k',1}} h_w(\boldsymbol{\theta}_{0:w}) = 0$ for any $k \neq k'$, that $\frac{\partial^2}{\partial \theta_{k,1} \partial \theta_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) = 0$ for any k, and that

$$\frac{\partial^2}{\partial \theta^2_{k,1}} h_w(\boldsymbol{\theta}_{0:w}) = -\frac{p_1 \theta^{1*}_{k,1}}{\theta^2_{k,1}} - \frac{p_1(1-\theta^{1*}_{k,1})}{(1-\theta_{k,1})^2} \leq -p_1 \theta^{1*}_{k,1} - p_1(1-\theta^{1*}_{k,1}) = -p_1 \tag{C.14}$$

$$\frac{\partial^2}{\partial \theta^2_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) = -\frac{1}{\theta^2_{0,1}} \left(\sum_{j=2}^{J} p_j \sum_{k=1}^{w} \theta^{j*}_{k,1} + w(1 - \sum p_j) \theta^*_{0,1} \right)$$

$$- \frac{1}{(1-\theta_{0,1})^2} \left(\sum_{j=2}^{J} p_j \sum_{k=1}^{w} (1 - \theta^{j*}_{k,1}) + w(1 - \sum p_j)(1 - \theta^*_{0,1}) \right)$$

$$\leq -w(1-p_1) \leq -(1-p_1). \tag{C.15}$$

So $h_w(\boldsymbol{\theta}_{0:w})$ is maximized at $(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}^{1*}_{1:w})$.

To show the second part of Lemma C.1, recall (C.8). We first address $\boldsymbol{\theta}_{0:w}$ such that $\theta_{0,1} = 1 - \phi + \zeta$. Using (C.13) we have $\frac{\partial}{\partial \theta_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) \Big|_{\theta_{0,1} = \hat{\theta}_{0,1}} = 0$. Applying (C.15), for any $\boldsymbol{\theta}_{0:w}$ such that $\theta_{0,1} = 1 - \phi + \zeta$,

$$h_w(\boldsymbol{\theta}_{0:w}) - h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}) = \int_{\hat{\theta}_{0,1}}^{1-\phi+\zeta} \frac{\partial}{\partial \theta_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) \Big|_{\theta_{0,1}=z} dz$$

$$= \int_{\hat{\theta}_{0,1}}^{1-\phi+\zeta} \int_{\hat{\theta}_{0,1}}^{z} \frac{\partial^2}{\partial \theta^2_{0,1}} h_w(\boldsymbol{\theta}_{0:w}) \Big|_{\theta_{0,1}=w} dw \, dz$$

$$\leq -(1-p_1)(1-\phi+\zeta-\hat{\theta}_{0,1})^2/2 \leq -(1-p_1)\zeta^2/2. \tag{C.16}$$

By (C.12), for any fixed value of $\boldsymbol{\theta}_0$ the function $h_w(\boldsymbol{\theta}_{0:w})$ is maximized at $(\boldsymbol{\theta}_0, \boldsymbol{\theta}^{1*}_{1:w})$. Combining with (C.16),

$$\inf_{\boldsymbol{\theta}_{0:w}:\theta_{0,1}=1-\phi+\zeta} \left[h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}^{1*}_{1:w}) - h_w(\boldsymbol{\theta}_{0:w})\right] \geq \inf_{\boldsymbol{\theta}_{0:w}:\theta_{0,1}=1-\phi+\zeta} \left[h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}^{1*}_{1:w}) - h_w(\boldsymbol{\theta}_0, \boldsymbol{\theta}^{1*}_{1:w})\right]$$
$$\geq (1-p_1)\zeta^2/2 \tag{C.17}$$

which is positive and does not depend on w.

Analogously, for $\boldsymbol{\theta}_{0:w}$ such that $\theta_{0,1} = \phi - \zeta$ we have

$$\inf_{\boldsymbol{\theta}_{0:w}:\theta_{0,1}=\phi-\zeta} \left[h_w(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}^{1*}_{1:w}) - h_w(\boldsymbol{\theta}_{0:w})\right] \geq (1-p_1)\zeta^2/2. \tag{C.18}$$

Using the analogous argument to handle the case where $\exists k : |\theta_{k,1} - \theta^{1*}_{k,1}| = \zeta$, and combining with (C.8), (C.17) and (C.18) yields (C.10). This proves Lemma C.1. □

Lemma C.2. *Under Assumptions 3.1Slow Mixing for Multiple True Motifsassumption.3.1-3.3Slow Mixing for Multiple True Motifsassumption.3.3 and using the definitions (C.5) and (C.6),*

$$\sup_{\boldsymbol{\theta}_{0:w}\in B^1_w} |E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) - h_w(\boldsymbol{\theta}_{0:w})| \overset{w\to\infty}{\longrightarrow} 0. \tag{C.19}$$

Proof. Using Assumption 3.3Slow Mixing for Multiple True Motifsassumption.3.3, $\prod_{k=1}^{w} \theta^{1*}_{k,s_k} = 1$ if $\mathbf{s} = \mathbf{t}^1_{1:w}$ and $\prod_{k=1}^{w} \theta^{1*}_{k,s_k} = 0$ for all other $\mathbf{s} \in \{1,2\}^w$. Combining with (2.8Statistical Motif Discoveryequation.2.8) and (3.3Slow Mixing for Multiple True Motifsequation.3.3), the first term of $E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) = \sum_{\mathbf{s}} g_{\boldsymbol{\theta}^*}(\mathbf{s}) \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$ is

$$\sum_{\mathbf{s}} \left[p_1 \prod_{k=1}^{w} \theta^{1*}_{k,s_k}\right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \tag{C.20}$$
$$= p_1 \log \left[p_0 \prod_{k=1}^{w} \theta_{k,t^1_k} + (1-p_0) \prod_{k=1}^{w} \theta_{0,t^1_k}\right].$$

We have that

$$\log \left[p_0 \prod_{k=1}^{w} \theta_{k,t^1_k} + (1-p_0) \prod_{k=1}^{w} \theta_{0,t^1_k}\right] - \log \left[p_0 \prod_{k=1}^{w} \theta_{k,t^1_k}\right] \geq 0. \tag{C.21}$$

Also, using (C.3)-(C.5) and the fact that $\theta^{1*}_{k,t^1_k} = 1$ for all $k \in \{1, \ldots, w\}$,

$$\sup_{\boldsymbol{\theta}_{0:w} \in B^1_w} \frac{(1-p_0)\prod_{k=1}^w \theta_{0,t^1_k}}{p_0 \prod_{k=1}^w \theta_{k,t^1_k}} \leq \frac{(1-p_0)(1-\phi+\zeta)^w}{p_0(1-\zeta)^w} \quad \stackrel{w\to\infty}{\longrightarrow} 0$$

since $1-\phi+\zeta < 1-\zeta$. So

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}_{0:w} \in B^1_w} \left(\log\left[p_0 \prod_{k=1}^w \theta_{k,t^1_k} + (1-p_0)\prod_{k=1}^w \theta_{0,t^1_k}\right] - \log\left[p_0 \prod_{k=1}^w \theta_{k,t^1_k}\right]\right) \\
&\leq \log\left[1 + \frac{(1-p_0)(1-\phi+\zeta)^w}{p_0(1-\zeta)^w}\right] \qquad \stackrel{w\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Combining with (C.21),

$$\sup_{\boldsymbol{\theta}_{0:w} \in B^1_w} \left| \log\left[p_0 \prod_{k=1}^w \theta_{k,t^1_k} + (1-p_0)\prod_{k=1}^w \theta_{0,t^1_k}\right] - \log\left[p_0 \prod_{k=1}^w \theta_{k,t^1_k}\right]\right| \stackrel{w\to\infty}{\longrightarrow} 0.$$

So, using (C.20),

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}_{0:w} \in B^1_w} \left| \sum_{\mathbf{s}} \left[p_1 \prod_{k=1}^w \theta^{1*}_{k,s_k}\right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) - \sum_{\mathbf{s}} \left[p_1 \prod_{k=1}^w \theta^{1*}_{k,s_k}\right] \log\left[p_0 \prod_{k=1}^w \theta_{k,s_k}\right]\right| \\
&\stackrel{w\to\infty}{\longrightarrow} 0.
\end{aligned} \tag{C.22}$$

Next we approximate the middle terms of $\sum_{\mathbf{s}} g_{\boldsymbol{\theta}^*}(\mathbf{s}) \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$. Using (2.8Statistical Motif Discoveryequation.2.8), (3.3Slow Mixing for Multiple True Motifsequation.3.3), and Assumption 3.3Slow Mixing for Multiple True Motifsassumption.3.3 they are of the following form for $j \in \{2, \ldots, J\}$.

$$\begin{aligned}
&\sum_{\mathbf{s}} \left[p_j \prod_{k=1}^w \theta^{j*}_{k,s_k}\right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \\
&= p_j \log\left[p_0 \prod_{k=1}^w \theta_{k,t^j_k} + (1-p_0)\prod_{k=1}^w \theta_{0,t^j_k}\right].
\end{aligned} \tag{C.23}$$

We have that

$$\log\left[p_0 \prod_{k=1}^w \theta_{k,t^j_k} + (1-p_0)\prod_{k=1}^w \theta_{0,t^j_k}\right] - \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,t^j_k}\right] \geq 0. \tag{C.24}$$

Let $\#\{t^j_k = t^1_k\}$ indicate the number of indices $k \in \{1, \ldots, w\}$ for which $t^j_k = t^1_k$. Using (C.4)-(C.5) and the fact that $\theta^{1*}_{k,t^j_k} = 0$ for all k such that $t^j_k \neq t^1_k$, we have that

$$\sup_{\boldsymbol{\theta}_{0:w} \in B^1_w} \frac{p_0 \prod_{k=1}^w \theta_{k,t^j_k}}{(1-p_0)\prod_{k=1}^w \theta_{0,t^j_k}} \leq \frac{p_0 \zeta^{\#\{t^j_k \neq t^1_k\}}}{(1-p_0)(\phi-\zeta)^w}.$$

Combining this with Assumption 3.3Slow Mixing for Multiple True Motifsassumption.3.3 and (C.3), for all w large enough

$$\sup_{\boldsymbol{\theta}_{0:w} \in B_w^1} \frac{p_0 \prod_{k=1}^w \theta_{k,t_k^j}}{(1-p_0)\prod_{k=1}^w \theta_{0,t_k^j}} \leq \frac{p_0 \zeta^{wa/2}}{(1-p_0)(\phi-\zeta)^w}$$

$$\leq \frac{p_0 (\phi/4)^w}{(1-p_0)(\phi-\zeta)^w} \qquad \stackrel{w\to\infty}{\longrightarrow} 0$$

since $\phi/4 < \phi - \zeta$. So

$$\sup_{\boldsymbol{\theta}_{0:w} \in B_w^1} \left(\log\left[p_0 \prod_{k=1}^w \theta_{k,t_k^j} + (1-p_0)\prod_{k=1}^w \theta_{0,t_k^j} \right] - \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,t_k^j} \right] \right)$$
$$\leq \log\left[\frac{p_0(\phi/4)^w}{(1-p_0)(\phi-\zeta)^w} + 1 \right] \qquad \stackrel{w\to\infty}{\longrightarrow} 0. \tag{C.25}$$

Using (C.24) and (C.25),

$$\sup_{\boldsymbol{\theta}_{0:w} \in B_w^1} \left| \log\left[p_0 \prod_{k=1}^w \theta_{k,t_k^j} + (1-p_0)\prod_{k=1}^w \theta_{0,t_k^j} \right] - \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,t_k^j} \right] \right| \stackrel{w\to\infty}{\longrightarrow} 0.$$

Combining with (C.23), for $j \in \{2,\ldots,J\}$

$$\sup_{\boldsymbol{\theta}_{0:w} \in B_w^1} \left| \sum_{\mathbf{s}} \left[p_j \prod_{k=1}^w \theta_{k,s_k}^{j*} \right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) - \sum_{\mathbf{s}} \left[p_j \prod_{k=1}^w \theta_{k,s_k}^{j*} \right] \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,s_k} \right] \right|$$
$$\stackrel{w\to\infty}{\longrightarrow} 0. \tag{C.26}$$

Finally we address the last term of term of $\sum_{\mathbf{s}} g_{\boldsymbol{\theta}^*}(\mathbf{s}) \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$. Using (2.8Statistical Motif Discoveryequation.2.8) and (3.3Slow Mixing for Multiple True Motifsequation.3.3) it is

$$\sum_{\mathbf{s}} \left[(1-\sum p_j) \prod_{k=1}^w \theta_{0,s_k}^* \right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$$
$$= \sum_{\mathbf{s}} \left[(1-\sum p_j) \prod_{k=1}^w \theta_{0,s_k}^* \right] \log\left[p_0 \prod_{k=1}^w \theta_{k,s_k} + (1-p_0)\prod_{k=1}^w \theta_{0,s_k} \right]. \tag{C.27}$$

We will show that a subset of sequences $\mathbf{s}$ can be omitted when considering (C.27). Denote by $F(x;n,q)$ the cumulative distribution function of a Binomial(n,q) random variable, evaluated at $x \in \mathbb{R}$. For $\mathbf{s} \in \{1,2\}^w$ recall that $\#\{s_k \neq t_k^1\}$ denotes the number of indices $k \in \{1,\ldots,w\}$ for which $s_k \neq t_k^1$. Define

$$D_w \triangleq \left\{ \mathbf{s} : \#\{s_k \neq t_k^1\} > w\phi/4 \right\}. \tag{C.28}$$

Then

$$\begin{aligned}
&\sum_{\mathbf{s}\in D_w}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\\
&\geq \max\left\{\sum_{\mathbf{s}:\#\{s_k\neq t^1_k,\,t^1_k=1\}>w\phi/4}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right],\ \sum_{\mathbf{s}:\#\{s_k\neq t^1_k,\,t^1_k=2\}>w\phi/4}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\right\}\\
&= \max\left\{\sum_{\mathbf{s}:\#\{s_k=2,\,t^1_k=1\}>w\phi/4}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right],\ \sum_{\mathbf{s}:\#\{s_k=1,\,t^1_k=2\}>w\phi/4}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\right\}\\
&= \max\Big\{1-F\big(w\phi/4;\,\#\{t^1_k=1\},1-\theta^*_{0,1}\big),\ \ 1-F\big(w\phi/4;\,\#\{t^1_k=2\},\theta^*_{0,1}\big)\Big\}. \qquad \text{(C.29)}
\end{aligned}$$

For fixed x, $F(x;n,q)$ is monotonic nonincreasing in n and q. Using (C.2) and (C.29), since $\phi<\min\{\theta^*_{0,1},1-\theta^*_{0,1}\}$ and $w/2\leq\max\{\#\{t^1_k=1\},\ \#\{t^1_k=2\}\}$, we have the following.

$$\begin{aligned}
\sum_{\mathbf{s}\in D_w}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right] &\geq \max\Big\{1-F\big(w\phi/4;\,\#\{t^1_k=1\},\phi\big),\ \ 1-F\big(w\phi/4;\,\#\{t^1_k=2\},\phi\big)\Big\}\\
&= 1-F\left(w\phi/4;\,\max\Big\{\#\{t^1_k=1\},\#\{t^1_k=2\}\Big\},\phi\right)\\
&\geq 1-F\left(w\phi/4;\,w/2,\phi\right). \qquad \text{(C.30)}
\end{aligned}$$

Using the normal approximation to the binomial distribution, the quantity $F\left(w\phi/4;\,w/2,\phi\right)$ decays exponentially in w. So by (C.30), the sum

$$\sum_{\mathbf{s}\notin D_w}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right]=1-\sum_{\mathbf{s}\in D_w}\left[\prod_{k=1}^{w}\theta^*_{0,s_k}\right] \qquad \text{(C.31)}$$

decays exponentially in w. Using this fact and (C.5),

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}_{0:w}\in B^1_w}\left|\sum_{\mathbf{s}\notin D_w}\left[(1-\sum p_j)\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\log\left[p_0\prod_{k=1}^{w}\theta_{k,s_k}+(1-p_0)\prod_{k=1}^{w}\theta_{0,s_k}\right]\right|\\
&\leq \sup_{\boldsymbol{\theta}_{0:w}\in B^1_w}\left[\sum_{\mathbf{s}\notin D_w}(1-\sum p_j)\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\left|\min_{\mathbf{s}}\log\left[(1-p_0)\prod_{k=1}^{w}\theta_{0,s_k}\right]\right|\\
&\leq \left[(1-\sum p_j)\sum_{\mathbf{s}\notin D_w}\prod_{k=1}^{w}\theta^*_{0,s_k}\right]\left|\,\log\left[(1-p_0)(\phi-\zeta)^w\right]\,\right|\\
&\overset{w\to\infty}{\longrightarrow} 0. \qquad \text{(C.32)}
\end{aligned}$$

Using (C.3)-(C.5) and (C.28), for $\boldsymbol{\theta}_{0:w} \in B_w^1$ and $\mathbf{s} \in D_w$,

$$\begin{aligned} \frac{p_0 \prod_{k=1}^w \theta_{k,s_k}}{(1-p_0)\prod_{k=1}^w \theta_{0,s_k}} &\le \frac{p_0 \zeta^{\#\{s_k \neq t_k^1\}}}{(1-p_0)(\phi-\zeta)^w} \\ &< \frac{p_0 \zeta^{w\phi/4}}{(1-p_0)(\phi-\zeta)^w} \\ &\le \frac{p_0 (\phi/4)^w}{(1-p_0)(\phi-\zeta)^w} \qquad \stackrel{w\to\infty}{\longrightarrow} 0 \end{aligned}$$

uniformly over $\boldsymbol{\theta}_{0:w} \in B_w^1$ and $\mathbf{s} \in D_w$, since $\phi/4 < \phi - \zeta$. So

$$\begin{aligned} &\sum_{\mathbf{s}\in D_w} \left[(1-\sum p_j)\prod_{k=1}^w \theta^*_{0,s_k}\right] \log\left[p_0 \prod_{k=1}^w \theta_{k,s_k} + (1-p_0)\prod_{k=1}^w \theta_{0,s_k}\right] \\ &\quad - \sum_{\mathbf{s}\in D_w} \left[(1-\sum p_j)\prod_{k=1}^w \theta^*_{0,s_k}\right] \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,s_k}\right] \qquad \stackrel{w\to\infty}{\longrightarrow} 0 \end{aligned} \tag{C.33}$$

uniformly over $\boldsymbol{\theta}_{0:w} \in B_w^1$. Also, using an analogous argument to (C.32),

$$\sup_{\boldsymbol{\theta}_{0:w}\in B_w^1} \left| \sum_{\mathbf{s}\notin D_w} \left[(1-\sum p_j)\prod_{k=1}^w \theta^*_{0,s_k}\right] \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,s_k}\right]\right| \stackrel{w\to\infty}{\longrightarrow} 0. \tag{C.34}$$

Combining (C.32)-(C.34),

$$\begin{aligned} \sup_{\boldsymbol{\theta}_{0:w}\in B_w^1} &\left| \sum_{\mathbf{s}} \left[(1-\sum p_j)\prod_{k=1}^w \theta^*_{0,s_k}\right] \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \right. \\ &\quad \left. - \sum_{\mathbf{s}} \left[(1-\sum p_j)\prod_{k=1}^w \theta^*_{0,s_k}\right] \log\left[(1-p_0)\prod_{k=1}^w \theta_{0,s_k}\right]\right| \stackrel{w\to\infty}{\longrightarrow} 0. \end{aligned} \tag{C.35}$$

Putting together the results (C.22), (C.26), and (C.35) for the various terms, we have that $\sum_{\mathbf{s}} g_{\boldsymbol{\theta}^*}(\mathbf{s}) \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$ converges to $h_w(\boldsymbol{\theta}_{0:w})$, uniformly over $\boldsymbol{\theta}_{0:w} \in B_w^1$. □

C.3 Proof of Theorem 3.3Rapid Mixing for ≤ 1 True Motiftheorem.3.3

For simplicity of notation we state the proof for the case $M = 2$ and $\beta_{k,m} = 1$ for all k, m, although the proof is analogous for any other choices of these constants. Recall the definitions of $\mathbf{C}(\mathbf{A})$, $\bar{\mathcal{X}}$, $\bar{\pi}$, $D_{\mathbf{c}}$, and $\bar{T}$ from Equations (5.3Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.3), (5.5Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.5), and (5.7Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.7)-(5.9Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.9). In the case $w = 1$ and $M = 2$ the vector $\mathbf{C}(\mathbf{A}) \in \bar{\mathcal{X}}$ only has two elements, $n \triangleq C(\mathbf{A})_1$ and $r \triangleq C(\mathbf{A})_2$. So we write $\bar{\pi}(n, r)$, suppressing the dependence of $\bar{\pi}$ on $\mathbf{S}$. Using (5.7Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.7), $\bar{\pi}(n, r) = \sum_{\mathbf{A}:\mathbf{C}(\mathbf{A})=(n,r)} \pi(\mathbf{A}|\mathbf{S})$. Since

$D_{(n,r)} = \{\mathbf{A} \in \mathcal{X} : \mathbf{C}(\mathbf{A}) = (n,r)\}$, let $|D_{(n,r)}|$ be the cardinality of $D_{(n,r)}$ and note that $|D_{(n,r)}| = \binom{N(\mathbf{S})_1}{n}\binom{N(\mathbf{S})_2}{r}$. Using (5.6Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.6) we have $|\mathbf{A}| = n + r$, $N(\mathbf{A}^{(1)})_1 = n$, $N(\mathbf{A}^{(1)})_2 = r$, $N(\mathbf{A}^c)_1 = N(\mathbf{S})_1 - n$, and $N(\mathbf{A}^c)_2 = N(\mathbf{S})_2 - r$. Then $\bar{\pi}$ simplifies as follows, using (2.5Statistical Motif Discoveryequation.2.5):

$$\bar{\pi}(n,r) \propto |D_{(n,r)}| p_0^{n+r}(1-p_0)^{L-n-r} \frac{\Gamma(N(\mathbf{S})_1 - n + \beta_{0,1})\Gamma(N(\mathbf{S})_2 - r + \beta_{0,2})}{\Gamma(L-n-r+|\boldsymbol{\beta}_0|)} \frac{\Gamma(n+\beta_{1,1})\Gamma(r+\beta_{1,2})}{\Gamma(n+r+|\boldsymbol{\beta}_1|)}$$

$$= |D_{(n,r)}| p_0^{n+r}(1-p_0)^{L-n-r} \frac{\Gamma(N(\mathbf{S})_1 - n + 1)\Gamma(N(\mathbf{S})_2 - r + 1)}{\Gamma(L-n-r+2)} \frac{\Gamma(n+1)\Gamma(r+1)}{\Gamma(n+r+2)}$$

$$= \frac{N(\mathbf{S})_1!}{n!(N(\mathbf{S})_1 - n)!} \left(\frac{N(\mathbf{S})_2!}{r!(N(\mathbf{S})_2 - r)!}\right) p_0^{n+r}(1-p_0)^{L-n-r} \times \frac{(N(\mathbf{S})_1 - n)!(N(\mathbf{S})_2 - r)!}{(L-n-r+1)!} \frac{n!r!}{(n+r+1)!}$$

$$\propto \frac{p_0^{n+r}(1-p_0)^{L-n-r}}{(L-n-r+1)!(n+r+1)!}. \tag{C.36}$$

This is a function of $(n+r)$ only; $\bar{\pi}(n,r)$ is also unimodal in $(n+r)$, shown as follows. The ratio

$$\frac{\bar{\pi}(n+1,r)}{\bar{\pi}(n,r)} = \frac{\bar{\pi}(n,r+1)}{\bar{\pi}(n,r)} = \frac{p_0}{1-p_0}\left(\frac{L-n-r+1}{n+r+2}\right) \tag{C.37}$$

is > 1 iff $n + r < p_0 L + 3p_0 - 2$, showing that $\bar{\pi}(n,r)$ is unimodal in $(n+r)$.

Using (2.6Statistical Motif Discoveryequation.2.6) and (5.9Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.9), in each iteration of $\bar{T}$ the quantity $(n+r)$ can only be incremented or decremented by one. Using (C.37) we have that incrementing or decrementing $(n+r)$ by one changes $\bar{\pi}(n,r)$ by no more than a factor of

$$d_2 \triangleq \max\left\{\frac{L-n-r+1}{(1-p_0)}, \frac{n+r+2}{p_0}\right\} = \mathcal{O}(L). \tag{C.38}$$

We will find a lower bound for the quantity d defined in (5.11Step 1 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.11), by defining a path $\gamma_{\mathbf{c}_1,\mathbf{c}_2}$ in the graph of $\bar{T}$ for every pair of states $\mathbf{c}_1, \mathbf{c}_2 \in \bar{\mathcal{X}}$. We will construct the paths in such a way that for any state $\mathbf{c} \in \gamma_{\mathbf{c}_1,\mathbf{c}_2}$ we have $\bar{\pi}(\mathbf{c}) \geq \min\{\bar{\pi}(\mathbf{c}_1), \bar{\pi}(\mathbf{c}_2)\}/d_2$. Denote $\mathbf{c}_1 = (n_1, r_1)$ and $\mathbf{c}_2 = (n_2, r_2)$. If $n_1 \leq n_2$ and $r_1 \leq r_2$, then construct the path by first increasing the first coordinate n from n_1 to n_2, then by increasing the second coordinate r from r_1 to r_2. Along this path, $n + r$ increases at every step. Since $\bar{\pi}(n,r)$ is a function only of $n+r$ and is unimodal in $n+r$, we have that for states (n,r) along the path,

$$\bar{\pi}(n,r) \geq \min\{\bar{\pi}(n_1,r_1), \bar{\pi}(n_2,r_2)\} \geq \min\{\bar{\pi}(n_1,r_1), \bar{\pi}(n_2,r_2)\}/d_2.$$

The case where $n_1 \geq n_2$ and $r_1 \geq r_2$ is analogous, since we can construct a path in the opposite direction as above. Now consider the case where $n_1 \leq n_2$ and $r_1 > r_2$ (the case $n_1 > n_2$, $r_1 \leq r_2$ is equivalent). Starting at (n_1, r_1), first decrement r by one, then increment n by one, and repeat until either $r = r_2$ or $n = n_2$. Notice that so far $n + r$ has changed by at most one, so that $\bar{\pi}(n, r)$ has changed by at most a factor of d_2. At this point, if $r = r_2$ then increase n until $n = n_2$, or if $n = n_2$ then decrease r until $r = r_2$. Any state (n, r) along this path satisfies $\bar{\pi}(n, r) \geq \min\{\bar{\pi}(n_1, r_1), \bar{\pi}(n_2, r_2)\}/d_2$ as desired. Using (C.38), the quantity d defined in (5.11Step 1 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.11) satisfies $d^{-1} = \mathcal{O}(L)$. Combined with (5.13Step 1 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.13) and Proposition 5.2Step 1 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1prop.5.2 this proves Theorem 3.3Rapid Mixing for ≤ 1 True Motiftheorem.3.3. □

C.4 Verifying the Assumptions of Theorem A.1Bayesian Asymptoticstheorem.A.1

By (5.29Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.29) Λ is a Borel set, and $\text{Int}(B_j)$ is a Borel set for $j \in \{1, 2\}$ because it is open. So the spaces Λ_j for $j \in \{1, 2\}$ are Borel subsets of the complete, separable metric space $\mathbb{R}^{w+1}$ as required. Also, $f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$ is measurable jointly in $\mathbf{s}$ and $\boldsymbol{\theta}_{0:w}$ since it is a continuous function of $\boldsymbol{\theta}_{0:w}$ and since $\mathbf{s}$ takes a finite set of values. Of course, Λ_j might not be connected, in which case $f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$ being continuous simply means that it is continuous on each connected component of Λ_j. Assumption 4 of Theorem A.1Bayesian Asymptoticstheorem.A.1 is satisfied since $\eta(\boldsymbol{\theta}_{0:w}) = E \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})$ is continuous. To show Assumption 2, observe that for all $\boldsymbol{\theta}_{0:w} \in \Lambda_j$ where $j \in \{1, 2\}$, $f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) > 0$ for any $\mathbf{s} \in \{1, 2\}^w$, so $G\{\mathbf{s} \in \{1, 2\}^w : f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) > 0\} = 1$ as desired.

To show Assumption 3 for Λ_1, take any compact $F \subset \Lambda_1$. We claim that there is some $\zeta \in (0, \frac{1}{2})$ such that

$$F \ \subset \ ([\zeta, 1 - \zeta] \times [0, 1]^w) \ \cup \ ([0, 1] \times [\zeta, 1 - \zeta]^w) \setminus \text{Int}(B_2). \tag{C.39}$$

Otherwise, there is some sequence $\{\boldsymbol{\theta}^{(\ell)}_{0:w} : \ell \in \mathbb{N}\}$ such that $\lim_{\ell \to \infty} \theta^{(\ell)}_{0,1} \in \{0, 1\}$ and $\exists k \in \{1, \ldots, w\}$ such that $\lim_{\ell \to \infty} \theta^{(\ell)}_{k,1} \in \{0, 1\}$. Since F is compact these points must have a limit point $\tilde{\boldsymbol{\theta}}_{0:w} \in F \subset \Lambda_1$. Then $\tilde{\theta}_{0,1} \in \{0, 1\}$ and $\tilde{\theta}_{k,1} \in \{0, 1\}$ which is a contradiction.

By (C.39), for any $\boldsymbol{\theta}_{0:w} \in F$ and any $\mathbf{s}$ we have $f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \geq \min\{p_0, 1 - p_0\}\zeta^w$. Then

$$\begin{aligned} E \sup_{\boldsymbol{\theta}_{0:w} \in F} |\log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})| &\leq \sup_{\mathbf{s} \in \{1,2\}^w, \, \boldsymbol{\theta}_{0:w} \in F} |\log f(\mathbf{s}|\boldsymbol{\theta}_{0:w})| \\ &\leq -\log\left[\min\{p_0, 1 - p_0\}\zeta^w\right] \qquad < \infty. \end{aligned}$$

To show that Assumption 5 is satisfied for Λ_1, it is sufficient to consider values of $r \in \mathbb{R}$ for which $r < (\log \frac{1}{2})(\min_{\mathbf{s}} g(\mathbf{s}))$. Let $\psi = \exp\{\frac{r}{\min_{\mathbf{s}} g(\mathbf{s})}\}$, so that $\psi \in (0, \frac{1}{2})$. Then define $D = \Lambda_1 \backslash D^c$ by letting D^c be the compact subset

$$D^c = ([\psi, 1 - \psi] \times [0, 1]^w) \ \cup \ ([0, 1] \times [\psi, 1 - \psi]^w) \setminus \text{Int}(B_2) \qquad \subset \Lambda_1.$$

We will define a cover $D_1, \ldots, D_K$ of D such that (A.1Bayesian Asymptoticsequation.A.1) holds. Define

$$\begin{aligned}
D_{k00} &= \{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in [0,\psi) \wedge \theta_{k,1} \in [0,\psi)\} \qquad k \in \{1,\ldots,w\} \\
D_{k10} &= \{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in (1-\psi,1] \wedge \theta_{k,1} \in [0,\psi)\} \\
D_{k01} &= \{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in [0,\psi) \wedge \theta_{k,1} \in (1-\psi,1]\} \\
D_{k11} &= \{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in (1-\psi,1] \wedge \theta_{k,1} \in (1-\psi,1]\}.
\end{aligned}$$

For all $\boldsymbol{\theta}_{0:w} \in D$ we have $\theta_{0,1} \in [0,\psi) \cup (1-\psi,1]$ and $\exists k \in \{1,\ldots,w\} : \theta_{k,1} \in [0,\psi) \cup (1-\psi,1]$. So

$$D \subset \cup_{k=1}^{w} \left(D_{k00} \cup D_{k10} \cup D_{k01} \cup D_{k11}\right).$$

Since $\log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \leq 0$, for any $k \in \{1,\ldots,w\}$

$$\begin{aligned}
E \sup_{\boldsymbol{\theta}_{0:w} \in D_{k00}} \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) &\leq g(\mathbf{t}) \sup_{\boldsymbol{\theta}_{0:w} \in D_{k00}} \log f(\mathbf{t}|\boldsymbol{\theta}_{0:w}) \qquad \text{where } \mathbf{t} = (1,\ldots,1) \\
&\leq g(\mathbf{t}) \log\left[p_0\psi + (1-p_0)\psi\right] \leq \left[\min_{\mathbf{s}} g(\mathbf{s})\right] \log \psi \qquad = r.
\end{aligned}$$

Also,

$$\begin{aligned}
E \sup_{\boldsymbol{\theta}_{0:w} \in D_{k01}} \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) &\leq g(\mathbf{t}) \sup_{\boldsymbol{\theta}_{0:w} \in D_{k01}} \log f(\mathbf{t}|\boldsymbol{\theta}_{0:w}) \qquad \text{where } \mathbf{t} = (\underbrace{1,\ldots,1}_{k-1 \text{ ones}}, 2, 1, \ldots, 1) \\
&\leq \left[\min_{\mathbf{s}} g(\mathbf{s})\right] \log\left[p_0\psi + (1-p_0)\psi\right] \qquad = r.
\end{aligned}$$

Analogously, $E \sup_{\boldsymbol{\theta}_{0:w} \in D_{k10}} \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \leq r$ and $E \sup_{\boldsymbol{\theta}_{0:w} \in D_{k11}} \log f(\mathbf{s}|\boldsymbol{\theta}_{0:w}) \leq r$, showing that Assumption 5 holds for Λ_1. Since Assumptions 3 and 5 hold for Λ_1, they hold for Λ_2 by symmetry.

C.5 Proof of Theorem 5.3Step 3 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1theorem.5.3

Assume that there exist $\epsilon > 0$ and $B_1, B_2 \subset [0,1]^{w+1}$ separated by distance ϵ such that the ratios in (5.25Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.25) decrease exponentially in L, and take F_1, F_2 as in Proposition C.1 below. Letting $\mathbf{c}_1$ be a maximizer of $\bar{\pi}(\mathbf{c}|\mathbf{S})$ over $\mathbf{c} \in F_1$, and $\mathbf{c}_2$ be a maximizer of $\bar{\pi}(\mathbf{c}|\mathbf{S})$ over $\mathbf{c} \in F_2$ and using Proposition C.1, for all L large enough

$$\begin{aligned}
\max\{\bar{\pi}(\mathbf{c}_1|\mathbf{S}), \bar{\pi}(\mathbf{c}_2|\mathbf{S})\} &\geq \frac{1}{2}\left(\bar{\pi}(\mathbf{c}_1|\mathbf{S}) + \bar{\pi}(\mathbf{c}_2|\mathbf{S})\right) \geq \frac{\bar{\pi}(F_1|\mathbf{S})}{2|F_1|} + \frac{\bar{\pi}(F_2|\mathbf{S})}{2|F_2|} \\
&\geq \frac{1}{2|\bar{\mathcal{X}}|}\left(\bar{\pi}(F_1|\mathbf{S}) + \bar{\pi}(F_2|\mathbf{S})\right) \geq \frac{1}{4|\bar{\mathcal{X}}|}. \qquad \text{(C.40)}
\end{aligned}$$

Combining with the fact that any path from $\mathbf{c}_1$ to $\mathbf{c}_2$ must include a state in $(F_1 \cup F_2)^c$,

$$\begin{aligned}
\max_{\gamma \in \Gamma_{\mathbf{c}_1,\mathbf{c}_2}} \min_{\mathbf{c}\in\gamma} \frac{\bar{\pi}(\mathbf{c}|\mathbf{S})}{\bar{\pi}(\mathbf{c}_1|\mathbf{S})\bar{\pi}(\mathbf{c}_2|\mathbf{S})} &\leq \max_{\gamma \in \Gamma_{\mathbf{c}_1,\mathbf{c}_2}} \min_{\mathbf{c}\in\gamma} \frac{4|\bar{\mathcal{X}}|\ \bar{\pi}(\mathbf{c}|\mathbf{S})}{\min\{\bar{\pi}(\mathbf{c}_1|\mathbf{S}), \bar{\pi}(\mathbf{c}_2|\mathbf{S})\}} \\
&\leq \max_{\mathbf{c}\in(F_1\cup F_2)^c} \frac{4|\bar{\mathcal{X}}|\ \bar{\pi}(\mathbf{c}|\mathbf{S})}{\min\{\bar{\pi}(\mathbf{c}_1|\mathbf{S}), \bar{\pi}(\mathbf{c}_2|\mathbf{S})\}} \\
&\leq \frac{4|\bar{\mathcal{X}}|\ \bar{\pi}\left((F_1\cup F_2)^c|\mathbf{S}\right)}{\min\{\bar{\pi}\left(\mathbf{c}_1|\mathbf{S}\right), \bar{\pi}\left(\mathbf{c}_2|\mathbf{S}\right)\}} \leq \frac{4|\bar{\mathcal{X}}|^2\ \bar{\pi}\left((F_1\cup F_2)^c|\mathbf{S}\right)}{\min\{\bar{\pi}\left(F_1|\mathbf{S}\right), \bar{\pi}\left(F_2|\mathbf{S}\right)\}}.
\end{aligned}$$

Since $|\bar{\mathcal{X}}|$ grows polynomially in L (using (5.10Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.10)), and using Proposition C.1, the quantity d decreases exponentially in L. □

Proposition C.1. *If there exist $\epsilon > 0$ and two sets $B_1, B_2 \subset [0,1]^{w+1}$ separated by Euclidean distance ϵ such that the ratios in (5.25Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.25) decrease exponentially in L, then there are two sets $F_1, F_2 \subset \bar{\mathcal{X}}$ such that:*

1. *For any $\mathbf{c}_1 \in F_1$ and $\mathbf{c}_2 \in F_2$, any path from $\mathbf{c}_1$ to $\mathbf{c}_2$ must include a state $\mathbf{c} \notin (F_1 \cup F_2)$.*

2. *The quantities*

$$\frac{\bar{\pi}\left((F_1\cup F_2)^c|\mathbf{S}\right)}{\bar{\pi}\left(F_1|\mathbf{S}\right)} \qquad \text{and} \qquad \frac{\bar{\pi}\left((F_1\cup F_2)^c|\mathbf{S}\right)}{\bar{\pi}\left(F_2|\mathbf{S}\right)} \tag{C.41}$$

   decrease exponentially in L.

Before proving Proposition C.1 we need a few preliminary results. The notation $\overset{\text{ind.}}{\sim}$ means independently distributed as.

Lemma C.3. *For any measure $\nu(dz)$ and nonnegative functions $a(z)$ and $b(z)$ on a space $z \in \mathcal{Z}$,*

$$\frac{\int a(z)\nu(dz)}{\int b(z)\nu(dz)} \geq \inf_{z\in\mathcal{Z}} \frac{a(z)}{b(z)}.$$

where the ratio inside the infimum is taken to be $=\infty$ whenever $b(z) = 0$.

Proof. We have

$$\frac{\int a(z)\nu(dz)}{\int b(z)\nu(dz)} \geq \frac{\int (\inf_w \frac{a(w)}{b(w)}) b(z)\nu(dz)}{\int b(z)\nu(dz)} = \inf_w \frac{a(w)}{b(w)}.$$

□

Lemma C.4. *Regarding the density of the Beta(a, b) distribution, where $a, b \geq 1$:*

1. *The density is unimodal if $a + b > 2$ and constant on $[0, 1]$ if $a + b = 2$.*

2. *A global maximum of the density occurs at*

$$x^* = \begin{cases} \frac{a-1}{a+b-2} & a+b > 2 \\ 0 & a+b = 2. \end{cases}$$

3. *For $X \sim$ Beta(a, b) and any $\zeta > 0$, Pr$(X \in [x^* - \zeta, x^* + \zeta]) \geq \min\{\zeta, 1\}$.*

Proof. The first two statements are well-known. To show the last, assume WLOG that $x^* \leq 1 - x^*$. We handle three cases separately: $\zeta \leq x^*$, $\zeta \in (x^*, 1 - x^*]$, and $\zeta > 1 - x^*$. For $\zeta > 1 - x^*$, $\Pr(X \in [x^* - \zeta, x^* + \zeta]) = 1$ so the result holds trivially.

For $\zeta \leq x^*$, letting $f(x)$ indicate the Beta(a, b) density and using Lemma C.3 and the fact that $f(x)$ is monotonically nondecreasing for $x < x^*$ and monotonically nonincreasing for $x > x^*$,

$$\begin{aligned}
\frac{\Pr(X \in [x^* - \zeta, x^* + \zeta])}{\Pr(X \notin [x^* - \zeta, x^* + \zeta])} &= \frac{\int_{x^*-\zeta}^{x^*} f(x)dx + \int_{x^*}^{x^*+\zeta} f(x)dx}{\int_0^{x^*-\zeta} f(x)dx + \int_{x^*+\zeta}^{1} f(x)dx} \\
&\geq \frac{f(x^* - \zeta)\zeta + f(x^* + \zeta)\zeta}{f(x^* - \zeta)(x^* - \zeta) + f(x^* + \zeta)(1 - x^* - \zeta)} \\
&\geq \min\left\{\frac{\zeta}{x^* - \zeta}, \frac{\zeta}{1 - x^* - \zeta}\right\} \geq \frac{\zeta}{1 - \zeta}.
\end{aligned}$$

So $\Pr(X \in [x^* - \zeta, x^* + \zeta]) \geq \zeta$.

Finally we address $\zeta \in (x^*, 1 - x^*]$. Then

$$\begin{aligned}
\frac{\Pr(X \in [x^* - \zeta, x^* + \zeta])}{\Pr(X \notin [x^* - \zeta, x^* + \zeta])} &\geq \frac{\int_{x^*}^{x^*+\zeta} f(x)dx}{\int_{x^*+\zeta}^{1} f(x)dx} \\
&\geq \frac{f(x^* + \zeta)\zeta}{f(x^* + \zeta)(1 - x^* - \zeta)} \geq \frac{\zeta}{1 - \zeta}
\end{aligned}$$

as desired. □

Lemma C.5. *For any $\zeta > 0$ and any $K \in \mathbb{N}$ the following holds for any $D_1, D_2 \subset [0, 1]^K$ that are separated by Euclidean distance $\geq \zeta$. Let $X_k \overset{ind.}{\sim}$ Beta(a_k, b_k) for $k \in \{1, \ldots, K\}$,* *where $a_k, b_k \geq 1$. Assume that the mode $\mathbf{x}^* = (x_1^*, \ldots, x_K^*)$ of the probability density function $f(\mathbf{x})$ of $\mathbf{X} = (X_1, \ldots, X_K)$ satisfies $\mathbf{x}^* \in D_1$, where x_k^* for $k \in \{1, \ldots, K\}$ are the modes of the univariate Beta densities as defined in Lemma C.4. Then $\frac{Pr(\mathbf{X} \notin D_1 \cup D_2)}{Pr(\mathbf{X} \in D_2)} \geq \left(\frac{\zeta}{2\sqrt{K}}\right)^{K+1}$.*

Proof. Consider the pdf $f(\mathbf{x})$ along any line segment originating at $\mathbf{x}^*$. This density is monotonically nonincreasing with distance from $\mathbf{x}^*$. For any set $D \subset [0,1]^K$ one can calculate the integral $\int_D f(\mathbf{x}) d\mathbf{x}$ by first transforming to spherical coordinates, where the origin of the coordinate system is taken to be $\mathbf{x}^*$. In this coordinate system let $\boldsymbol{\phi}$ denote the $(K-1)$-dimensional vector of angular coordinates, and $\rho \geq 0$ denote the radius, i.e. the distance from $\mathbf{x}^*$. Let $h(\rho, \boldsymbol{\phi})$ be the (invertible) function that maps from the spherical coordinates to the Euclidean coordinates. The Jacobian of the transformation h takes the form $\rho^K g(\boldsymbol{\phi})$ for some function g. So for any $D \subset [0,1]^K$ we can write

$$\int_D f(\mathbf{x}) d\mathbf{x} = \int_{h^{-1}(D)} f(h(\rho, \boldsymbol{\phi})) \rho^K g(\boldsymbol{\phi}) d\rho d\boldsymbol{\phi}.$$

In particular (using Lemma C.3),

$$\begin{aligned}
\frac{\Pr(\mathbf{X} \notin D_1 \cup D_2)}{\Pr(\mathbf{X} \in D_2)} &= \frac{\int_{h^{-1}((D_1 \cup D_2)^c)} f(h(\rho, \boldsymbol{\phi})) \rho^K g(\boldsymbol{\phi}) d\rho d\boldsymbol{\phi}}{\int_{h^{-1}(D_2)} f(h(\rho, \boldsymbol{\phi})) \rho^K g(\boldsymbol{\phi}) d\rho d\boldsymbol{\phi}} \\
&= \frac{\int \left[\int \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in (D_1 \cup D_2)^c} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho\right] g(\boldsymbol{\phi}) d\boldsymbol{\phi}}{\int \left[\int \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho\right] g(\boldsymbol{\phi}) d\boldsymbol{\phi}} \\
&\geq \inf_{\boldsymbol{\phi}} \frac{\int \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in (D_1 \cup D_2)^c} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho}{\int \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho}
\end{aligned}$$

where we consider the ratio inside the infimum to be $= \infty$ if the denominator is zero. Then

$$\begin{aligned}
\frac{\Pr(\mathbf{X} \notin D_1 \cup D_2)}{\Pr(\mathbf{X} \in D_2)} &\geq \inf_{\boldsymbol{\phi}} \frac{\int_{\zeta/2}^{\infty} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in (D_1 \cup D_2)^c} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho}{\int \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho} \\
&= \inf_{\boldsymbol{\phi}} \frac{\int_{\zeta/2}^{\sqrt{K}} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in (D_1 \cup D_2)^c} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho}{\int_0^{\sqrt{K}} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) \rho^K d\rho} \\
&\geq \left(\frac{\zeta}{2\sqrt{K}}\right)^K \inf_{\boldsymbol{\phi}} \frac{\int_{\zeta/2}^{\sqrt{K}} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in (D_1 \cup D_2)^c} f(h(\rho, \boldsymbol{\phi})) d\rho}{\int_0^{\sqrt{K}} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) d\rho}. \qquad \text{(C.42)}
\end{aligned}$$

For any fixed $\boldsymbol{\phi}$ for which $0 \neq \int_0^{\sqrt{K}} \mathbf{1}_{h(\rho,\boldsymbol{\phi}) \in D_2} f(h(\rho, \boldsymbol{\phi})) d\rho$, there is some $\tilde{\rho}$ such that $h(\tilde{\rho}, \boldsymbol{\phi}) \in D_2$. Since $\mathbf{x}^* = h(0, \boldsymbol{\phi}) \in D_1$ and since D_1 and D_2 are separated by distance ζ, there must be an interval $[\rho_1(\boldsymbol{\phi}), \rho_2(\boldsymbol{\phi})] \subset [0, \tilde{\rho}]$ of width at least ζ such that any $\rho \in [0, \rho_1(\boldsymbol{\phi})]$ satisfies $h(\rho, \boldsymbol{\phi}) \notin D_2$ and any $\rho \in (\rho_1(\boldsymbol{\phi}), \rho_2(\boldsymbol{\phi}))$ satisfies $h(\rho, \boldsymbol{\phi}) \in (D_1 \cup D_2)^c$. Using (C.42) and since $f(h(\rho, \boldsymbol{\phi}))$ is monotonically nonincreasing in ρ,

$$\begin{aligned}
\frac{\Pr(\mathbf{X} \notin D_1 \cup D_2)}{\Pr(\mathbf{X} \in D_2)} &\geq \left(\frac{\zeta}{2\sqrt{K}}\right)^K \inf_{\boldsymbol{\phi}} \frac{\int_{\max\{\zeta/2, \rho_1(\boldsymbol{\phi})\}}^{\rho_2(\boldsymbol{\phi})} f(h(\rho, \boldsymbol{\phi}))d\rho}{\int_{\rho_2(\boldsymbol{\phi})}^{\sqrt{K}} f(h(\rho, \boldsymbol{\phi}))d\rho} \\
&\geq \left(\frac{\zeta}{2\sqrt{K}}\right)^K \inf_{\boldsymbol{\phi}} \frac{\int_{\max\{\zeta/2, \rho_1(\boldsymbol{\phi})\}}^{\rho_2(\boldsymbol{\phi})} f(h(\rho_2(\boldsymbol{\phi}), \boldsymbol{\phi}))d\rho}{\int_{\rho_2(\boldsymbol{\phi})}^{\sqrt{K}} f(h(\rho_2(\boldsymbol{\phi}), \boldsymbol{\phi}))d\rho} \\
&\geq \left(\frac{\zeta}{2\sqrt{K}}\right)^{K+1}.
\end{aligned}$$

□

Lemma C.6. *For $k \in \{1, \ldots, K\}$ let $X_k \overset{ind.}{\sim}$ Beta(a_k, b_k) where $a_k, b_k \geq 1$. Then for any set $D \subset [0,1]^K$ with positive Lebesgue measure ($\lambda(D) > 0$) and any $d_3 > 1$,*

$$\inf_{a_1, b_1, \ldots, a_K, b_K \in [1, d_3]} Pr(\mathbf{X} \in D) > 0$$

where $\mathbf{X} = (X_1, \ldots, X_K)$.

Proof. Since $\lambda(D) > 0$, there is some $\zeta \in (0, 1/2)$ such that the set $\tilde{D} = D \cap [\zeta, 1-\zeta]^K$ satisfies $\lambda(\tilde{D}) > 0$. Letting $f(x)$ indicate the density of any Beta(a, b) distribution where $a, b \in [1, d_3]$, and using Lemma C.4,

$$\begin{aligned}
\frac{\inf_{x \in [\zeta, 1-\zeta]} f(x)}{\sup_x f(x)} &= \frac{\min\{f(\zeta), f(1-\zeta)\}}{f(\frac{a-1}{a+b-2})} \\
&\geq \frac{\zeta^{a+b-2}(a+b-2)^{a+b-2}}{(a-1)^{a-1}(b-1)^{b-1}} \\
&\geq \zeta^{a+b-2} \geq \zeta^{2d_3-2}.
\end{aligned}$$

Now letting $f(\mathbf{x})$ indicate the function on $\mathbf{x} \in [0,1]^K$ that is the product of Beta(a_k, b_k) densities where $a_k, b_k \in [1, d_3]$,

$$\frac{\inf_{\mathbf{x} \in [\zeta, 1-\zeta]^K} f(\mathbf{x})}{\sup_{\mathbf{x}} f(\mathbf{x})} \geq \zeta^{K(2d_3-2)}.$$

So

$$\frac{\Pr(\mathbf{X} \in D)}{\Pr(\mathbf{X} \in D^c)} \geq \frac{\Pr(\mathbf{X} \in \tilde{D})}{\Pr(\mathbf{X} \in \tilde{D}^c)} \geq \frac{\lambda(\tilde{D}) \inf_{\mathbf{x} \in [\zeta, 1-\zeta]^K} f(\mathbf{x})}{(1-\lambda(\tilde{D})) \sup_{\mathbf{x}} f(\mathbf{x})} \geq \frac{\lambda(\tilde{D})\zeta^{K(2d_3-2)}}{(1-\lambda(\tilde{D}))} \tag{C.43}$$

which is strictly positive and does not depend on $\{a_k, b_k\}_{k=1}^K$. □

Lemma C.7. *Let* $X_k \overset{ind.}{\sim} Beta(a_k, b_k)$ *for* $k \in \{1, \ldots, Q\}$ *where* $Q \in \mathbb{N}$ *and* $a_k, b_k \geq 1$*. Also let* x_k^* *be the global mode of the density of* $Beta(a_k, b_k)$ *as defined in Lemma C.4. Let* $B(\mathbf{x}, \delta)$ *indicate the ball of radius* $\delta > 0$ *centered at a point* $\mathbf{x} \in [0,1]^Q$*. Then for any fixed* $\delta > 0$*,* $d_3 \geq 1$*, and* $K \in \{1, \ldots, Q\}$*,*

$$\inf_{a_k, b_k \in [1, d_3]: k=1,\ldots,K} \; \inf_{a_k, b_k \geq 1: k=K+1,\ldots,Q} \; \inf_{\mathbf{x} \in [0,1]^Q : x_k = x_k^*, k=K+1,\ldots,Q} Pr(\mathbf{X} \in B(\mathbf{x}, \delta)) > 0.$$

Proof. Take a hypercube $H(\mathbf{x}, \delta)$ centered at $\mathbf{x}$ and with some fixed side length $2\delta_1 \in (0, 1]$ for which $H(\mathbf{x}, \delta) \subset B(\mathbf{x}, \delta)$. Then

$$\begin{aligned}
&\inf_{a_k, b_k \in [1, d_3]: k=1,\ldots,K} \; \inf_{a_k, b_k \geq 1: k=K+1,\ldots,Q} \; \inf_{\mathbf{x} \in [0,1]^Q : x_k = x_k^*, k=K+1,\ldots,Q} \Pr(\mathbf{X} \in B(\mathbf{x}, \delta)) \\
&\geq \inf_{a_k, b_k \in [1, d_3]: k=1,\ldots,K} \; \inf_{a_k, b_k \geq 1: k=K+1,\ldots,Q} \; \inf_{\mathbf{x} \in [0,1]^Q : x_k = x_k^*, k=K+1,\ldots,Q} \Pr(\mathbf{X} \in H(\mathbf{x}, \delta)) \\
&= \left[\prod_{k=1}^{K} \inf_{a_k, b_k \in [1, d_3]} \; \inf_{x_k \in [0,1]} \Pr(X_k \in [x_k - \delta_1, x_k + \delta_1]) \right] \prod_{k=K+1}^{Q} \inf_{a_k, b_k \geq 1} \Pr(X_k \in [x_k^* - \delta_1, x_k^* + \delta_1]).
\end{aligned} \tag{C.44}$$

By Lemma C.4, the second product in this expression is bounded below by δ_1^{Q-K}. To bound the first product in (C.44) we will use the explicit lower bound (C.43) given in the proof of Lemma C.6, applied to the single variable X_k where $k \in \{1, \ldots, K\}$. Here we take the set $D = [x_k - \delta_1, x_k + \delta_1] \cap [0, 1]$. Let $\zeta = \frac{\delta_1}{2}$ so that $\tilde{D} = D \cap [\frac{\delta_1}{2}, 1 - \frac{\delta_1}{2}]$. Noticing that $\lambda(\tilde{D}) \geq \frac{\delta_1}{2}$, the bound (C.43) gives

$$\frac{\Pr(X_k \in D)}{\Pr(X_k \in D^c)} \geq \frac{\left(\frac{\delta_1}{2}\right)^{1+(2d_3-2)}}{1 - \frac{\delta_1}{2}} \geq \frac{\left(\frac{\delta_1}{2}\right)^{(2d_3-1)}}{1 - \left(\frac{\delta_1}{2}\right)^{(2d_3-1)}}.$$

So $\Pr(X_k \in D) \geq \left(\frac{\delta_1}{2}\right)^{(2d_3-1)}$; applying this method for each $k = 1, \ldots, K$ we have that

$$\begin{aligned}
&\inf_{a_k, b_k \in [1, d_3]: k=1,\ldots,K} \; \inf_{a_k, b_k \geq 1: k=K+1,\ldots,Q} \; \inf_{\mathbf{x} \in [0,1]^Q : x_k = x_k^*, k=K+1,\ldots,Q} \Pr(\mathbf{X} \in B(\mathbf{x}, \delta)) \\
&\geq \left(\frac{\delta_1}{2}\right)^{K(2d_3-1)} \delta_1^{Q-K} > 0.
\end{aligned}$$

□

Proof of Proposition C.1. Recall the definition (Sec. 2.1Statistical Motif Discoverysubsection.2.1) of $\boldsymbol{\beta}_k$; we will take $\beta_{k,m} = 1$ for $k \in \{0, \ldots, w\}$ and $m \in \{1, 2\}$ for simplicity of exposition, although the results do not depend on this choice. Then the prior for $\boldsymbol{\theta}_{0:w}$ is uniform: $\pi(\boldsymbol{\theta}_{0:w}) \propto \mathbf{1}_{\{\boldsymbol{\theta}_{0:w} \in [0,1]^{w+1}\}}$.

The quantities $\mathbf{N}(\mathbf{A}^{(k)})$ and $\mathbf{N}(\mathbf{A}^c)$ only depend on $\mathbf{A}$ via $\mathbf{C}(\mathbf{A})$, due to (5.6Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.6). Consider the conditional distribution $\pi(\boldsymbol{\theta}_{0:w}|\mathbf{C}(\mathbf{A}), \mathbf{S})$, which can be written as follows, using (2.3Statistical Motif Discoveryequation.2.3):

$$\begin{aligned}\pi(\boldsymbol{\theta}_{0:w}|\mathbf{C}(\mathbf{A}), \mathbf{S}) &\propto \pi(\boldsymbol{\theta}_{0:w}, \mathbf{C}(\mathbf{A}), \mathbf{S}) \;\propto \pi(\boldsymbol{\theta}_{0:w})\, \pi(\mathbf{C}(\mathbf{A}))\, \pi(\mathbf{S}|\mathbf{C}(\mathbf{A}), \boldsymbol{\theta}_{0:w}) \\ &\propto \left[\prod_{k=1}^{w} \prod_{m=1}^{2} \theta_{k,m}^{N(\mathbf{A}^{(k)})_m}\right] \prod_{m=1}^{2} \theta_{0,m}^{N(\mathbf{A}^c)_m} \\ &\propto \left[\prod_{k=1}^{w} \mathrm{Beta}(\theta_{k,1}; N(\mathbf{A}^{(k)})_1 + 1, N(\mathbf{A}^{(k)})_2 + 1)\right] \times \\ &\qquad \mathrm{Beta}(\theta_{0,1}; N(\mathbf{A}^c)_1 + 1, N(\mathbf{A}^c)_2 + 1). \end{aligned} \tag{C.45}$$

where $\mathrm{Beta}(x; a, b)$ indicates the Beta density with parameters a, b, evaluated at x. By Lemma C.4, $\pi(\boldsymbol{\theta}_{0:w}|\mathbf{C}(\mathbf{A}), \mathbf{S})$ is a density with global maximum at $\tilde{\boldsymbol{\theta}}_{0:w}$ where

$$\tilde{\theta}_{k,1} = \begin{cases} \frac{N(\mathbf{A}^{(k)})_1}{|\mathbf{N}(\mathbf{A}^{(k)})|} & |\mathbf{N}(\mathbf{A}^{(k)})| > 0 \\ 0 & \text{else} \end{cases} \qquad k \in \{1, \ldots, w\} \tag{C.46}$$

$$\tilde{\theta}_{0,1} = \begin{cases} \frac{N(\mathbf{A}^{c})_1}{|\mathbf{N}(\mathbf{A}^{c})|} & |\mathbf{N}(\mathbf{A}^{c})| > 0 \\ 0 & \text{else.} \end{cases}$$

To complete the notation define $\tilde{\theta}_{k,2} = 1 - \tilde{\theta}_{k,1}$ for $k \in \{0, \ldots, w\}$.

By (C.45) and since $|\mathbf{N}(\mathbf{A}^c)| = L - \sum_{k=1}^{w} |\mathbf{N}(\mathbf{A}^{(k)})|$, we have that $\pi(\boldsymbol{\theta}_{0:w}|\mathbf{C}(\mathbf{A}), \mathbf{S})$ only depends on $\mathbf{C}(\mathbf{A})$ via $\tilde{\boldsymbol{\theta}}_{0:w}$ and $|\mathbf{N}(\mathbf{A}^{(1)})| = |\mathbf{N}(\mathbf{A}^{(2)})| = \ldots = |\mathbf{N}(\mathbf{A}^{(w)})|$. So

$$\begin{aligned}&\pi\Big(\boldsymbol{\theta}_{0:w} \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big) \\ &= \left[\prod_{k=1}^{w} \mathrm{Beta}\Big(\theta_{k,1};\ \tilde{\theta}_{k,1}|\mathbf{N}(\mathbf{A}^{(1)})| + 1,\ \tilde{\theta}_{k,2}|\mathbf{N}(\mathbf{A}^{(1)})| + 1\Big)\right] \times \\ &\qquad \mathrm{Beta}\Big(\theta_{0,1};\ \tilde{\theta}_{0,1}(L - w|\mathbf{N}(\mathbf{A}^{(1)})|) + 1,\ \tilde{\theta}_{0,2}(L - w|\mathbf{N}(\mathbf{A}^{(1)})|) + 1\Big). \end{aligned} \tag{C.47}$$

Using Lemma C.4 and regardless of the value of $|\mathbf{N}(\mathbf{A}^{(1)})|$, $\pi\Big(\boldsymbol{\theta}_{0:w} \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)$ has a global maximum at $\tilde{\boldsymbol{\theta}}_{0:w}$.

For our analysis the only relevant quantities regarding $\mathbf{C}(\mathbf{A}) \in \bar{\mathcal{X}}$ will be $\tilde{\boldsymbol{\theta}}_{0:w}$ and $|\mathbf{N}(\mathbf{A}^{(1)})|$, so we define $F_1, F_2 \subset \bar{\mathcal{X}}$ more conveniently as sets of *possible values* of $\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big)$, i.e. values that arise from some state $\mathbf{C}(\mathbf{A}) \in \bar{\mathcal{X}}$. We will define F_1 to be a particular set for which there is some constant $d_4 > 0$ satisfying

$$\min_{(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1} \frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)} \geq d_4. \tag{C.48}$$

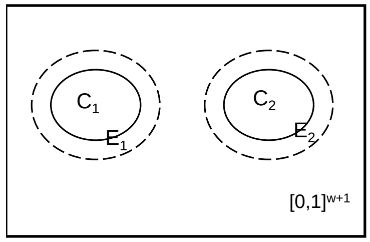

Figure 1: An illustration of the proof.

So $F_1 \subset \bar{\mathcal{X}}$ is associated with $B_1 \subset [0,1]^{w+1}$ in the sense that it (informally speaking) contains all the values of $\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big)$ for which $\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)$ is much larger than $\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)$. The set F_1 must have high probability (given $\mathbf{S}$) in order to explain the fact that the first quantity in (5.25Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.25) decreases exponentially in L.

To begin, recall the definition of $\epsilon > 0$ from Proposition C.1. Let E_1 be the set of all points $\mathbf{x} \in [0,1]^{w+1}$ that are within distance $\epsilon/3$ of the set B_1, and let E_2 be the set of all points that are within distance $\epsilon/3$ of the set B_2. This is illustrated in Web Appendix Figure 1. Then E_1 and E_2 are separated by distance $\epsilon_1 \triangleq \epsilon/3$. Let $d_5 \triangleq \frac{w+1}{\epsilon_1}$; since $B_1, B_2 \subset [0,1]^{w+1}$ are separated by distance ϵ, we have that $\epsilon \leq \sqrt{w+1}$ and so

$$d_5 = \frac{w+1}{\epsilon/3} > \frac{w+1}{\sqrt{w+1}} > 1. \tag{C.49}$$

Also define

$$\begin{aligned} V \triangleq &\Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) : \max\{|\mathbf{N}(\mathbf{A}^{(1)})|,\, |\mathbf{N}(\mathbf{A}^{c})|/w\} > d_5\Big\} \qquad \text{(C.50)}\\ &\cap \Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) :\ \text{if } \exists \boldsymbol{\theta}_0 \in [0,1] \text{ s.t. } (\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_{1:w}) \in (E_1 \cup E_2)^c \text{ then } |\mathbf{N}(\mathbf{A}^{c})|/w > d_5\Big\}\\ &\cap \Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) :\ \text{if } \exists \boldsymbol{\theta}_{1:w} \in [0,1]^w \text{ s.t. } (\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}) \in (E_1 \cup E_2)^c \text{ then } |\mathbf{N}(\mathbf{A}^{(1)})| > d_5\Big\}\\ F_j \triangleq &\Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) \in V : \tilde{\boldsymbol{\theta}}_{0:w} \in E_j\Big\} \qquad j \in \{1,2\}. \end{aligned}$$

First we show that it is not possible to move from any state $\big(\tilde{\boldsymbol{\theta}}^1_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^1\big) \in F_1$ to any state $\big(\tilde{\boldsymbol{\theta}}^2_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^2\big) \in F_2$ in one iteration of $\bar{T}$. Since $\tilde{\boldsymbol{\theta}}^1_{0:w} \in E_1$ and $\tilde{\boldsymbol{\theta}}^2_{0:w} \in E_2$ satisfy $\|\tilde{\boldsymbol{\theta}}^1_{0:w} - \tilde{\boldsymbol{\theta}}^2_{0:w}\| \geq \epsilon_1$, we have that $\exists \tilde{k} \in \{0, \ldots, w\}$ such that $|\tilde{\theta}^1_{\tilde{k},1} - \tilde{\theta}^2_{\tilde{k},1}| \geq \frac{\epsilon_1}{w+1} = \frac{1}{d_5}$. We handle the four cases: 1. where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 \leq d_5$; 2. where $|\mathbf{N}(\mathbf{A}^{c})|^1/w \leq d_5$; 3. where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 > d_5$, $|\mathbf{N}(\mathbf{A}^{c})|^1/w > d_5$ and $\tilde{k} > 0$; 4. where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 > d_5$, $|\mathbf{N}(\mathbf{A}^{c})|^1/w > d_5$ and $\tilde{k} = 0$. We assume that it is it is possible to move from $\big(\tilde{\boldsymbol{\theta}}^1_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^1\big)$ to $\big(\tilde{\boldsymbol{\theta}}^2_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^2\big)$ in one iteration of $\bar{T}$, and find a contradiction. We use the fact that, by (2.6Statistical Motif Discoveryequation.2.6) and (5.9Outline of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.9), in one iteration of $\bar{T}$ the vector $\mathbf{N}(\mathbf{A}^{(\tilde{k})})$ can only change by incrementing or decrementing a single element by one, and so $|\mathbf{N}(\mathbf{A}^{(\tilde{k})})| = |\mathbf{N}(\mathbf{A}^{(1)})|$ can only increase or decrease by one. Also, the vector $\mathbf{N}(\mathbf{A}^c)$ can only change by either incrementing its elements by a total of w, which increases $|\mathbf{N}(\mathbf{A}^c)|$ by w, or decrementing its elements by a total of w, which decreases $|\mathbf{N}(\mathbf{A}^c)|$ by w.

First take the case where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 > d_5$, $|\mathbf{N}(\mathbf{A}^c)|^1/w > d_5$ and $\tilde{k} > 0$. By (C.49), $|\mathbf{N}(\mathbf{A}^{(1)})|^1 > 1$, so $|\mathbf{N}(\mathbf{A}^{(1)})|^2 > 0$. By (C.46),

$$|\tilde{\theta}^1_{\tilde{k},1} - \tilde{\theta}^2_{\tilde{k},1}| = \left| \frac{N(\mathbf{A}^{(\tilde{k})})^1_1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1} - \frac{N(\mathbf{A}^{(\tilde{k})})^2_1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^2} \right|. \tag{C.51}$$

Also, we claim that this is bounded above by $\frac{1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1}$. In the case where $N(\mathbf{A}^{(\tilde{k})})^2_1 = N(\mathbf{A}^{(\tilde{k})})^1_1 + \delta$ and $\delta \in \{-1, 1\}$, we have $N(\mathbf{A}^{(\tilde{k})})^2_1 \geq 0$ so $N(\mathbf{A}^{(\tilde{k})})^1_1 \geq -\delta$ and thus

$$\begin{aligned} |\tilde{\theta}^1_{\tilde{k},1} - \tilde{\theta}^2_{\tilde{k},1}| = \left| \frac{N(\mathbf{A}^{(\tilde{k})})^1_1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1} - \frac{N(\mathbf{A}^{(\tilde{k})})^1_1 + \delta}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1 + \delta} \right| &= \left(\frac{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1 - N(\mathbf{A}^{(\tilde{k})})^1_1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1 + \delta} \right) \frac{|\delta|}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1} \\ &\leq \frac{|\delta|}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1} = \frac{1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1}. \end{aligned}$$

In the case where $N(\mathbf{A}^{(\tilde{k})})^2_2 = N(\mathbf{A}^{(\tilde{k})})^1_2 + \delta$ and $\delta \in \{-1, 1\}$, by using the fact that $|\tilde{\theta}^1_{\tilde{k},1} - \tilde{\theta}^2_{\tilde{k},1}| = |\tilde{\theta}^1_{\tilde{k},2} - \tilde{\theta}^2_{\tilde{k},2}|$ and applying the above argument we still obtain the upper bound $\frac{1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1}$. Combining with (C.51) we have

$$|\tilde{\theta}^1_{\tilde{k},1} - \tilde{\theta}^2_{\tilde{k},1}| \leq \frac{1}{|\mathbf{N}(\mathbf{A}^{(\tilde{k})})|^1} < \frac{1}{d_5} \tag{C.52}$$

which is a contradiction (by the definition of $\tilde{k}$).

Now take the case where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 \leq d_5$. Then by (C.50) we must have $|\mathbf{N}(\mathbf{A}^c)|^1/w > d_5$. Also, $\tilde{\boldsymbol{\theta}}^1_{0:w} \in E_1$ and there is no $\boldsymbol{\theta}_{1:w}$ such that $(\tilde{\boldsymbol{\theta}}^1_0, \boldsymbol{\theta}_{1:w}) \in (E_1 \cup E_2)^c$, so $(\tilde{\boldsymbol{\theta}}^1_0, \tilde{\boldsymbol{\theta}}^2_{1:w}) \in E_1$. Therefore the Euclidean distance between $(\tilde{\boldsymbol{\theta}}^1_0, \tilde{\boldsymbol{\theta}}^2_{1:w}) \in E_1$ and $(\tilde{\boldsymbol{\theta}}^2_0, \tilde{\boldsymbol{\theta}}^2_{1:w}) \in E_2$ is $\geq \epsilon_1$. This implies $|\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| \geq \epsilon_1 > \frac{1}{d_5}$. However, by (C.49), $|\mathbf{N}(\mathbf{A}^c)|^1 > d_5 w > w$, so $|\mathbf{N}(\mathbf{A}^c)|^2 > 0$. Then by (C.46),

$$|\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| = \left| \frac{N(\mathbf{A}^c)^1_1}{|\mathbf{N}(\mathbf{A}^c)|^1} - \frac{N(\mathbf{A}^c)^2_1}{|\mathbf{N}(\mathbf{A}^c)|^2} \right|$$

Also, we claim that this is bounded above by $\frac{w}{|\mathbf{N}(\mathbf{A}^c)|^1}$. In the case where $N(\mathbf{A}^c)^2_1 = N(\mathbf{A}^c)^1_1 + \delta$ and $N(\mathbf{A}^c)^2_2 = N(\mathbf{A}^c)^1_2 + w - \delta$ for $\delta \in \{0, \ldots, w\}$,

$$\begin{aligned} |\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| &= \left| \frac{N(\mathbf{A}^c)^1_1}{|\mathbf{N}(\mathbf{A}^c)|^1} - \frac{N(\mathbf{A}^c)^1_1 + \delta}{|\mathbf{N}(\mathbf{A}^c)|^1 + w} \right| \\ &= \left| \frac{w N(\mathbf{A}^c)^1_1 - \delta |\mathbf{N}(\mathbf{A}^c)|^1}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 + w\right)} \right| \\ &\leq \frac{\max\left\{w\left(|\mathbf{N}(\mathbf{A}^c)|^1 - N(\mathbf{A}^c)^1_1\right), \, w N(\mathbf{A}^c)^1_1\right\}}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 + w\right)} \qquad \leq \frac{w}{|\mathbf{N}(\mathbf{A}^c)|^1}. \end{aligned}$$

In the case where $N(\mathbf{A}^c)^2_1 = N(\mathbf{A}^c)^1_1 - \delta$ and $N(\mathbf{A}^c)^2_2 = N(\mathbf{A}^c)^1_2 - w + \delta$ for $\delta \in \{0, \dots, w\}$,

$$\begin{aligned} |\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| &= \left| \frac{N(\mathbf{A}^c)^1_1}{|\mathbf{N}(\mathbf{A}^c)|^1} - \frac{N(\mathbf{A}^c)^1_1 - \delta}{|\mathbf{N}(\mathbf{A}^c)|^1 - w} \right| \\ &= \left| \frac{-wN(\mathbf{A}^c)^1_1 + \delta |\mathbf{N}(\mathbf{A}^c)|^1}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 - w\right)} \right| \end{aligned} \tag{C.53}$$

This is largest when $\delta \in \{0, w\}$. Note that $N(\mathbf{A}^c)^2_1 \geq 0$ and $N(\mathbf{A}^c)^2_2 \geq 0$ so $N(\mathbf{A}^c)^1_1 \geq \delta$ and $N(\mathbf{A}^c)^1_2 \geq w - \delta$. Using (C.53), when $\delta = 0$ we have $N(\mathbf{A}^c)^1_2 \geq w$ and

$$\begin{aligned} |\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| &= \frac{wN(\mathbf{A}^c)^1_1}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 - w\right)} \\ &= \frac{w\left(|\mathbf{N}(\mathbf{A}^c)|^1 - N(\mathbf{A}^c)^1_2\right)}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 - w\right)} \leq \frac{w}{|\mathbf{N}(\mathbf{A}^c)|^1}. \end{aligned}$$

When $\delta = w$ we have $N(\mathbf{A}^c)^1_1 \geq w$ and (using (C.53))

$$|\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| = \frac{w\left(|\mathbf{N}(\mathbf{A}^c)|^1 - N(\mathbf{A}^c)^1_1\right)}{|\mathbf{N}(\mathbf{A}^c)|^1 \left(|\mathbf{N}(\mathbf{A}^c)|^1 - w\right)} \leq \frac{w}{|\mathbf{N}(\mathbf{A}^c)|^1}.$$

as claimed. So $|\tilde{\theta}^1_{0,1} - \tilde{\theta}^2_{0,1}| \leq \frac{w}{|\mathbf{N}(\mathbf{A}^c)|^1} < \frac{1}{d_5}$, which is a contradiction. The case where $|\mathbf{N}(\mathbf{A}^{(1)})|^1 > d_5$, $|\mathbf{N}(\mathbf{A}^c)|^1/w > d_5$ and $\tilde{k} = 0$, and the case where $|\mathbf{N}(\mathbf{A}^c)|^1 \leq d_5 w$, lead to contradictions analogously to the two cases handled above. So it is not possible to move from $(\tilde{\boldsymbol{\theta}}^1_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^1)$ to $(\tilde{\boldsymbol{\theta}}^2_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|^2)$ in one iteration of $\bar{T}$.

Next we show (C.48). By Lemma C.5, (C.47), (C.50), and $B_2 \subset E_2$, there is some $d_6 > 0$ that depends only on w such that

$$\begin{aligned} &\min_{\left(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\right) \in F_2} \frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)} \\ &\geq \min_{\tilde{\boldsymbol{\theta}}_{0:w} \in E_2} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|} \frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)} \\ &\geq \min_{\tilde{\boldsymbol{\theta}}_{0:w} \in E_2} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|} \frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup E_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)} \geq d_6. \end{aligned} \tag{C.54}$$

Also, by Lemma C.5 and $E_1 \backslash B_1 \subset (B_1 \cup B_2)^c$, there exists $d_7 > 0$ that depends only on w such that

$$\begin{aligned} &\min_{\tilde{\boldsymbol{\theta}}_{0:w} \in (E_1 \cup E_2)^c} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \\ &\geq \min_{\tilde{\boldsymbol{\theta}}_{0:w} \in E_1^c} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \in E_1 \backslash B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \geq d_7. \end{aligned} \tag{C.55}$$

Additionally, by Lemma C.6, $\exists d_8 > 0$ such that

$$\begin{aligned}
&\min_{\tilde{\boldsymbol{\theta}}_{0:w}} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|:|\mathbf{N}(\mathbf{A}^{(1)})|,|\mathbf{N}(\mathbf{A}^c)|/w \le d_5} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \\
&\ge \min_{\tilde{\boldsymbol{\theta}}_{0:w}} \min_{|\mathbf{N}(\mathbf{A}^{(1)})|:|\mathbf{N}(\mathbf{A}^{(1)})|,|\mathbf{N}(\mathbf{A}^c)|/w \le d_5} \Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big) > d_8. \qquad \text{(C.56)}
\end{aligned}$$

Also, for any $\boldsymbol{\theta}_{1:w}$ such that $(\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}) \in (E_1 \cup E_2)^c$, a ball of radius $\epsilon_1/2 = \epsilon/6$ centered at $(\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w})$ is entirely contained in $(B_1 \cup B_2)^c$. By Lemma C.7, $\exists d_9 > 0$

$$\begin{aligned}
&\min_{\tilde{\boldsymbol{\theta}}_{0:w}:\exists(\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}'_{1:w}) \in (E_1 \cup E_2)^c} \min_{|\mathbf{N}(\mathbf{A}^{(1)})| \le d_5} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \qquad \text{(C.57)} \\
&\ge \min_{\tilde{\boldsymbol{\theta}}_{0:w}:\exists(\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}'_{1:w}) \in (E_1 \cup E_2)^c} \min_{|\mathbf{N}(\mathbf{A}^{(1)})| \le d_5} \Pr\Big(\boldsymbol{\theta}_{0:w} \in B\big((\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}'_{1:w}), \epsilon_1/2\big) \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big) \ge d_9.
\end{aligned}$$

By the analogous argument, $\exists d_{10} > 0$

$$\min_{\tilde{\boldsymbol{\theta}}_{0:w}:\exists(\boldsymbol{\theta}'_0, \tilde{\boldsymbol{\theta}}_{1:w}) \in (E_1 \cup E_2)^c} \min_{|\mathbf{N}(\mathbf{A}^c)|/w \le d_5} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \ge d_{10}. \qquad \text{(C.58)}$$

By (C.50),

$$\begin{aligned}
(F_1 \cup F_2)^c = &\Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) : \tilde{\boldsymbol{\theta}}_{0:w} \in (E_1 \cup E_2)^c \;\vee\; \max\{|\mathbf{N}(\mathbf{A}^{(1)})|, |\mathbf{N}(\mathbf{A}^c)|/w\} \le d_5\Big\} \\
&\cup \Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) : |\mathbf{N}(\mathbf{A}^c)|/w \le d_5 \;\wedge\; \exists \boldsymbol{\theta}_0 \text{ s.t. } (\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_{1:w}) \in (E_1 \cup E_2)^c\Big\} \\
&\cup \Big\{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) : |\mathbf{N}(\mathbf{A}^{(1)})| \le d_5 \;\wedge\; \exists \boldsymbol{\theta}_{1:w} \text{ s.t. } (\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_{1:w}) \in (E_1 \cup E_2)^c\Big\}
\end{aligned}$$

and due to (C.55)-(C.58) we have

$$\min_{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|\big) \in (F_1 \cup F_2)^c} \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\big)} \ge \min\{d_7, d_8, d_9, d_{10}\} > 0.$$

Combining this result with (C.54) yields (C.48).

Now we prove the second part of Proposition C.1. Using Lemma C.3 and (C.48),

$$\begin{aligned}
&\frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2\Big)} \\
&= \frac{\sum_{\big(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|\big) \in F_2} \Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big) \pi\big(|\mathbf{N}(\mathbf{A}^{(1)})|, \tilde{\boldsymbol{\theta}}_{0:w} \,\big|\, \mathbf{S}\big)}{\sum_{(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2} \Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big) \pi\big(|\mathbf{N}(\mathbf{A}^{(1)})|, \tilde{\boldsymbol{\theta}}_{0:w} \,\big|\, \mathbf{S}\big)} \\
&\ge \min_{(\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2} \frac{\Pr\Big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)}{\Pr\Big(\boldsymbol{\theta}_{0:w} \in B_1 \,\Big|\, \tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|, \mathbf{S}\Big)} \ge d_4. \qquad \text{(C.59)}
\end{aligned}$$

Analogously,

$$\frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2\big)} \geq d_4. \tag{C.60}$$

Then by symmetry we have

$$\frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2\big)} \geq d_4$$

which combined with (C.60) yields

$$\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2\big) \geq \frac{d_4}{2+d_4} > 0. \tag{C.61}$$

Again using Lemma C.3,

$$\frac{\Pr(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 | \mathbf{S})}{\Pr(\boldsymbol{\theta}_{0:w} \in B_1 | \mathbf{S})} \geq \min\left\{ \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2\big)}, \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)} \right\}.$$

Using this fact and (C.59) and since the ratios in (5.25Step 2 of Proof of Thm. 3.1Slow Mixing for Multiple True Motifstheorem.3.1equation.5.25) are exponentially decreasing in L,

$$\frac{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)} \tag{C.62}$$

is also exponentially decreasing in L. Also, using (C.60)-(C.61),

$$\begin{aligned}
&\frac{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,\big|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2\big)} \\
&= \frac{\Pr\big(\boldsymbol{\theta}_{0:w} \in B_1, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2 \,\big|\, \mathbf{S}\big)}{\Pr\big(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_2 \,\big|\, \mathbf{S}\big)} \\
&= \frac{\Pr(\boldsymbol{\theta}_{0:w} \in B_1 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_1)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_1 \,|\, \mathbf{S}) + \Pr(\boldsymbol{\theta}_{0:w} \in B_1 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2 \,|\, \mathbf{S})}{\Pr(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_1)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_1 \,|\, \mathbf{S}) + \Pr(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2 \,|\, \mathbf{S})} \\
&\leq \frac{\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_1 \,|\, \mathbf{S}) + \Pr(\boldsymbol{\theta}_{0:w} \in B_1 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2 \,|\, \mathbf{S})}{\Pr(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2)\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2 \,|\, \mathbf{S})} \\
&= \frac{\frac{\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \in F_1 \,|\, \mathbf{S})}{\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1 \cup F_2 \,|\, \mathbf{S})} + \Pr(\boldsymbol{\theta}_{0:w} \in B_1 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1 \cup F_2)}{\Pr(\boldsymbol{\theta}_{0:w} \notin B_1 \cup B_2 \,|\, \mathbf{S}, (\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1 \cup F_2)} \\
&\leq \left(\frac{2+d_4}{d_4}\right) \frac{\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \in F_1 \,|\, \mathbf{S})}{\Pr((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1 \cup F_2 \,|\, \mathbf{S})} + \frac{1}{d_4}.
\end{aligned}$$

Combining with the fact that (C.62) is exponentially decreasing in L, $\frac{\Pr\left((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \notin F_1 \cup F_2 \,\middle|\, \mathbf{S}\right)}{\Pr\left((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)})|) \in F_1 \,\middle|\, \mathbf{S}\right)}$ is also exponentially decreasing in L. By symmetry, $\frac{\Pr\left((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \notin F_1 \cup F_2 \,\middle|\, \mathbf{S}\right)}{\Pr\left((\tilde{\boldsymbol{\theta}}_{0:w}, |\mathbf{N}(\mathbf{A}^{(1)}|) \in F_2 \,\middle|\, \mathbf{S}\right)}$ decreases exponentially in L, proving Proposition C.1. $\square$