
Detailed Proofs of Lemmas, Theorems, and Corollaries

Dahua Lin
CSAIL, MIT

John Fisher
CSAIL, MIT

A List of Lemmas, Theorems, and Corollaries

For being self-contained, we list here all the lemmas, theorems, and corollaries in the main paper.

Lemma 1. *The joint transition matrix \mathbf{P}_+ given by Eq.(1) in the main paper has a stationary distribution in form of $(\alpha\boldsymbol{\mu}_X, \beta\boldsymbol{\mu}_Y)$, if and only if*

$$\boldsymbol{\mu}_X \mathbf{Q}_B = \boldsymbol{\mu}_Y, \quad \text{and} \quad \boldsymbol{\mu}_Y \mathbf{Q}_F = \boldsymbol{\mu}_X. \quad (1)$$

Under this condition, we have $\alpha b = \beta f$. Further, if both \mathbf{P}_X and \mathbf{P}_Y are both reversible, then \mathbf{P}_+ is also reversible, if and only if

$$\boldsymbol{\mu}_X(x) \mathbf{Q}_B(x, y) = \boldsymbol{\mu}_Y(y) \mathbf{Q}_F(y, x), \quad (2)$$

for all $x \in X$ and $y \in Y$.

Lemma 2. *If the condition given above holds, then $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ are respectively stationary distributions of \mathbf{Q}_{BF} and \mathbf{Q}_{FB} . Moreover, if \mathbf{P}_+ is reversible, then both \mathbf{Q}_{BF} and \mathbf{Q}_{FB} are reversible.*

Theorem 1. *Given a reversible Markov chain with transition matrix \mathbf{P} , and $\varepsilon \in (0, 1/2)$, then*

$$\log(1/(2\varepsilon))(\tau - 1) \leq t_{mix}(\varepsilon) \leq \log(1/(\varepsilon\boldsymbol{\mu}_{min}))\tau. \quad (3)$$

Here, τ is called the relaxation time, given by $1/\gamma_*(\mathbf{P})$.

Theorem 2. *Let λ_2 be the second largest eigenvalue of a reversible transition matrix \mathbf{P} , then*

$$\Phi_*^2(\mathbf{P})/2 \leq 1 - \lambda_2 \leq 2\Phi_*(\mathbf{P}). \quad (4)$$

Theorem 3. *The reversible transition matrix \mathbf{P}_+ as given by Eq.(1) in the main paper has*

$$\frac{\eta(b, f)}{2} \cdot \frac{\phi}{\phi + 1} \leq \Phi_*(\mathbf{P}_+) \leq \max\{b, f\}. \quad (5)$$

Here, $\eta(b, f) = 2\alpha b = 2\beta f$ is the total probability of cross-space transition, $\phi = \min\{\Phi_*(\mathbf{Q}_{BF}), \Phi_*(\mathbf{Q}_{FB})\}$.

Corollary 1. *The joint chain \mathbf{P}_+ is ergodic when the collapsed chains (\mathbf{Q}_{BF} and \mathbf{Q}_{FB}) are both ergodic.*

Lemma 3. *Let \mathbf{P} be a reversible transition matrix over X , such that $\mathbf{P}(x, x) \geq \xi > 0$ for each $x \in X$ then its smallest eigenvalue λ_n has $\lambda_n \geq 2\xi - 1$.*

Theorem 4. *The hierarchically bridging Markov chain with $b_k < 1$ for $k = 0, \dots, K-1$, and $f_k < 1$ for $k = 1, \dots, K$ is ergodic. If we write the equilibrium distribution in form of $(\alpha\boldsymbol{\mu}_0, \beta_1\boldsymbol{\mu}_1, \dots, \beta_K\boldsymbol{\mu}_K)$, then (S1) $\boldsymbol{\mu}_0$ equals the target distribution $\boldsymbol{\mu}$; (S2) for each $k \geq 1$, and $y \in Y_k$, $\boldsymbol{\mu}_k(y)$ is proportional to the total probability of its descendant target states (the target states derived by filling all its placeholders); (S3) α , the probability of being at the target level, is given by $\alpha^{-1} = 1 + \sum_{k=1}^K (b_0 \cdots b_{k-1}) / (f_1 \cdots f_k)$.*

Corollary 2. *If $b_k/f_{k+1} \leq \kappa < 1$ for each $k = 1, \dots, K$, then $\alpha > 1 - \kappa$.*

B Proofs

Here, we provide the proofs of the lemmas and theorems presented in the paper.

B.1 Proof of Lemma 1

Recall that \mathbf{P}_+ is given by

$$\mathbf{P}_+ = \begin{bmatrix} (1-b)\mathbf{P}_X & b\mathbf{Q}_B \\ f\mathbf{Q}_F & (1-f)\mathbf{P}_Y \end{bmatrix}.$$

Suppose \mathbf{P}_+ has a stationary distribution in form of $(\alpha\boldsymbol{\mu}_X, \beta\boldsymbol{\mu}_Y)$, then

$$\begin{aligned} \alpha(1-b)\boldsymbol{\mu}_X \mathbf{P}_X + \beta f \boldsymbol{\mu}_Y \mathbf{Q}_F &= \alpha\boldsymbol{\mu}_X, \\ \alpha b \boldsymbol{\mu}_X \mathbf{Q}_B + \beta(1-f)\boldsymbol{\mu}_Y \mathbf{P}_Y &= \beta\boldsymbol{\mu}_Y. \end{aligned} \quad (6)$$

Since $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ are respectively stationary distributions of \mathbf{P}_X and \mathbf{P}_Y , i.e. $\boldsymbol{\mu}_X \mathbf{P}_X = \boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y \mathbf{P}_Y = \boldsymbol{\mu}_Y$, we have

$$\beta f \boldsymbol{\mu}_Y \mathbf{Q}_F = \alpha b \boldsymbol{\mu}_X, \quad \text{and} \quad \alpha b \boldsymbol{\mu}_X \mathbf{Q}_B = \beta f \boldsymbol{\mu}_Y. \quad (7)$$

Note that \mathbf{Q}_B and \mathbf{Q}_F were defined with the condition that each of their rows sums to 1, i.e. $\mathbf{Q}_F \mathbf{1}_{|Y|} = \mathbf{1}_{|Y|}$ and $\mathbf{Q}_B \mathbf{1}_{|X|} = \mathbf{1}_{|X|}$. Multiplying $\mathbf{1}$ to the right of both hand sides of either equation results in $\alpha b = \beta f$. It immediately follows that

$$\boldsymbol{\mu}_X \mathbf{Q}_B = \boldsymbol{\mu}_Y, \quad \text{and} \quad \boldsymbol{\mu}_Y \mathbf{Q}_F = \boldsymbol{\mu}_X. \quad (8)$$

For the other direction, we assume \mathbf{Q}_B and \mathbf{Q}_F satisfy the conditions above, and $\alpha b = \beta f$. Plugging these

conditions into the left hand sides of the equations in Eq.(6) results in $(\alpha\mu_X, \beta\mu_Y)\mathbf{P}_+ = (\alpha\mu_X, \beta\mu_Y)$, which implies that $(\alpha\mu_X, \beta\mu_Y)$ is a stationary distribution of \mathbf{P}_+ . The proof of the first part is done.

Next, we show the second part of the lemma, which is about the reversibility. Let $\mu_+ = (\alpha\mu_X, \beta\mu_Y)$. Under the condition that \mathbf{P}_X and \mathbf{P}_Y are both reversible, we have for each $x, x' \in X$,

$$\begin{aligned}\mu_+(x)\mathbf{P}_+(x, x') &= \alpha(1-b)\mu_X(x)\mathbf{P}_X(x, x'), \\ \mu_+(x')\mathbf{P}_+(x', x) &= \alpha(1-b)\mu_X(x')\mathbf{P}_X(x', x),\end{aligned}\quad (9)$$

and thus $\mu_+(x)\mathbf{P}_+(x, x') = \mu_+(x')\mathbf{P}_+(x', x)$ (due to the reversibility of \mathbf{P}_X). Likewise, we can get $\mu_+\mathbf{P}_+(y, y') = \mu_+(y')\mathbf{P}_+(y', y)$. Hence, \mathbf{P}_+ is reversible if and only if $\mu_+(x)\mathbf{P}_+(x, y) = \mu_+(y)\mathbf{P}_+(y, x)$ for each $x \in X$ and $y \in Y$. This can be expanded as

$$\alpha b \mu_X(x) \mathbf{Q}_B(x, y) = \beta f \mu_Y(y) \mathbf{Q}_F(y, x), \quad (10)$$

which holds if and only if $\mu_X(x) \mathbf{Q}_B(x, y) = \mu_Y(y) \mathbf{Q}_F(y, x)$ (under the condition $\alpha b = \beta f$). The proof is completed.

B.2 Proof of Lemma 2

If $(\alpha\mu_X, \beta\mu_Y)$ is a stationary distribution of \mathbf{P}_+ , then Eq.(8) holds. Thus,

$$\mu_X \mathbf{Q}_B \mathbf{Q}_F = \mu_Y \mathbf{Q}_F = \mu_X, \quad (11)$$

implying that μ_X is a stationary distribution of $\mathbf{Q}_B \mathbf{Q}_F$. Similarly, we can show that μ_Y is a stationary distribution of $\mathbf{Q}_F \mathbf{Q}_B$.

Furthermore, if \mathbf{P}_+ is reversible, according to Lemma 1, we have $\mu_X(x) \mathbf{Q}_B(x, y) = \mu_Y(y) \mathbf{Q}_F(y, x)$ for each $x \in X$ and $y \in Y$. Then for any $x, x' \in X$,

$$\begin{aligned}\mu_X(x) \mathbf{Q}_{BF}(x, x') &= \mu_X(x) \sum_{y \in Y} \mathbf{Q}_B(x, y) \mathbf{Q}_F(y, x') \\ &= \sum_{y \in Y} \mu_X(x) \mathbf{Q}_B(x, y) \mathbf{Q}_F(y, x') \\ &= \sum_{y \in Y} \mu_Y(y) \mathbf{Q}_F(y, x) \mathbf{Q}_F(y, x')\end{aligned}$$

Similarly, we can get

$$\mu_X(x') \mathbf{Q}_{BF}(x', x) = \sum_{y \in Y} \mu_Y(y) \mathbf{Q}_F(y, x') \mathbf{Q}_F(y, x).$$

Hence,

$$\mu_X(x) \mathbf{Q}_{BF}(x, x') = \mu_X(x') \mathbf{Q}_{BF}(x', x). \quad (12)$$

This implies that \mathbf{Q}_{BF} is reversible. Likewise, we can show that \mathbf{Q}_{FB} is reversible under the condition that \mathbf{P}_+ is reversible. The proof is completed.

B.3 Proof of Theorem 3

To prove theorem 3, we first establish a lemma on flow decomposition, and then accomplish the proof based on the lemma.

B.3.1 A Lemma on Flow Decomposition

For the joint chain \mathbf{P}_+ , we analyze its bottleneck ratio by decomposing the flows. Consider a partition of the union space $X \cup Y$ into two parts: $A \cup B$ (with $A \subset X$ and $B \subset Y$) and $A^c \cup B^c$ (with $A^c = X/A$ and $B^c = Y/B$). The flow between them comprises three parts:

$$\mathcal{F}(A, A^c) + \mathcal{F}(B, B^c) + (\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c)).$$

Here, $\mathcal{F}(A, A^c)$ is the flow within X , $\mathcal{F}(B, B^c)$ is the flow within Y , and $\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c)$ is the flow between X and Y . The first two are inherited from the original Markov chains. We focus on the third one, which reflects the effect of bridging. For this part of flow, we derive the following lemma by decomposing it along multiple paths.

Lemma 4. *Given arbitrary partition of $X \cup Y$ into $A \cup B$ and $A^c \cup B^c$ as described above, we have*

$$\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c) \geq \alpha b \cdot \Phi_*(\mathbf{Q}_{BF}) \mu_X(A), \quad (13)$$

when $\mu_X(A) \leq \mu_X(A^c)$, and

$$\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c) \geq \beta f \cdot \Phi_*(\mathbf{Q}_{FB}) \mu_Y(B), \quad (14)$$

when $\mu_Y(B) \leq \mu_Y(B^c)$.

Proof. To analyze the flow $\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c)$, we further decompose it along multiple paths. Based on Eq.(1) in the main paper, we have

$$\begin{aligned}\mathcal{F}(A, B^c) &= \sum_{x \in A} \sum_{y \in B^c} \alpha b \mu_X(x) \mathbf{Q}_B(x, y) \\ &= \alpha b \sum_{x \in A} \sum_{y \in B^c} \mu_X(x) \mathbf{Q}_B(x, y) \sum_{x' \in X} \mathbf{Q}_F(y, x')\end{aligned}\quad (15)$$

In this way, we decompose the flow into a sum of the terms in form of $\mu_X(x) \mathbf{Q}_B(x, y) \mathbf{Q}_F(y, x')$, which we call the *path weight* along $x \rightarrow y \rightarrow x'$, denoted by $\omega(x, y, x')$. We can then rewrite $\mathcal{F}(A, B)$ as

$$\mathcal{F}(A, B) = \alpha b \sum_{x \in A} \sum_{y \in B} \sum_{x' \in X} \omega(x, y, x'). \quad (16)$$

For conciseness, we use $\omega(A, B, C)$ to denote the sum of paths traveling from A , via B , and ending up in C , i.e. $\sum_{x \in A} \sum_{y \in B} \sum_{x' \in C} \omega(x, y, x')$. Then, we have

$$\mathcal{F}(A, B^c) = \alpha b (\omega(A, B^c, A) + \omega(A, B^c, A^c)), \quad (17)$$

$$\mathcal{F}(A^c, B) = \alpha b (\omega(A^c, B, A) + \omega(A^c, B, A^c)). \quad (18)$$

As \mathcal{F} is symmetric for a reversible chain, we have $\mathcal{F}(B, A^c) = \mathcal{F}(A^c, B)$, and thus

$$\begin{aligned} \mathcal{F}(A, B^c) + \mathcal{F}(B, A^c) &\geq \alpha b(\omega(A, B^c, A^c) + \omega(A^c, B, A)) \\ &= \alpha b \omega(A, Y, A^c). \end{aligned} \quad (19)$$

On the other hand, we note that

$$\begin{aligned} \omega(A, Y, A^c) &= \sum_{x \in A} \sum_{y \in Y} \sum_{x' \in A^c} \omega(x, y, x') \\ &= \sum_{x \in A} \sum_{y \in Y} \sum_{x' \in A^c} \mu_X(x) \mathbf{Q}_B(x, y) \mathbf{Q}_F(y, x') \\ &= \sum_{x \in A} \sum_{x' \in A^c} \mu_X(x) \sum_{y \in Y} \mathbf{Q}_B(x, y) \mathbf{Q}_F(y, x') \\ &= \sum_{x \in A} \sum_{x' \in A^c} \mu_X(x) \mathbf{Q}_{BF}(x, x'). \end{aligned} \quad (20)$$

This is exactly the flow from A to A^c with respect to the collapsed chain \mathbf{Q}_{BF} , *i.e.*

$$\omega(A, Y, A^c) = \mathcal{F}_{\mathbf{Q}_{BF}}(A, A^c). \quad (21)$$

Assuming $\mu_X(A) \leq \mu_X(A^c)$, we have $\mathcal{F}_{\mathbf{Q}_{BF}}(A, A^c) \geq \Phi_*(\mathbf{Q}_{BF})\mu(A)$, by the definition of bottleneck ratio. Combining this with Eq.(19) results in

$$\begin{aligned} \mathcal{F}(A, B^c) + \mathcal{F}(B, A^c) &\geq \alpha b \cdot \omega(A, Y, A^c) \\ &\geq \alpha b \cdot \Phi_*(\mathbf{Q}_{BF})\mu_X(A). \end{aligned} \quad (22)$$

Likewise, with the assumption $\mu_Y(B) \leq \mu_Y(B^c)$, we have

$$\mathcal{F}(A, B^c) + \mathcal{F}(B, A^c) \geq \beta f \cdot \Phi_*(\mathbf{Q}_{FB})\mu_Y(B). \quad (23)$$

The proof of the lemma is completed. \square

B.3.2 The Proof of the Main Theorem

Let $\mu_+ = (\alpha\mu_X, \beta\mu_Y)$ be the stationary distribution of \mathbf{P}_+ . For conciseness, we let $F_s(A, B) \triangleq \mathcal{F}(A \cup B, A^c \cup B^c)$. When A and B are clear from the context, we simply write F_s . Then the bottleneck ratio of \mathbf{P}_+ is the minimum of the values of $F_s(A, B)/\mu_+(A \cup B)$, among all possible choices of $A \subset X$ and $B \subset Y$ such that $\mu_+(A \cup B) \leq 1/2$ and $A \cup B \neq \emptyset$. Throughout this proof, we assume $A \subset X$, $B \subset Y$, and $\mu_+(A \cup B) \leq 1/2$, *i.e.* $\alpha\mu_X(A) + \beta\mu_Y(B) \leq 1/2$.

Under this assumption, there are three cases, which we respectively discuss as follows.

Case 1. $\mu_X(A) \leq 1/2$ and $\mu_Y(B) \leq 1/2$.

We have $\phi = \min\{\Phi_*(\mathbf{Q}_{BF}), \Phi_*(\mathbf{Q}_{FB})\}$ in the theorem. In addition, $F_s \geq \mathcal{F}(A, B^c) + \mathcal{F}(B, A^c)$. Combining this with Lemma 4, we get

$$F_s \geq \alpha b \cdot \phi \mu_X(A), \text{ and } F_s \geq \beta f \cdot \phi \mu_Y(B). \quad (24)$$

Note that $\eta/2 = \alpha b = \beta f$ and $\alpha + \beta = 1$. Thus

$$\frac{F_s}{\mu_+(A \cup B)} \geq \frac{\alpha F_s + \beta F_s}{\alpha \mu_X(A) + \beta \mu_Y(B)} \geq \eta \phi / 2. \quad (25)$$

Case 2. $\mu_X(A) < 1/2$ and $\mu_Y(B) > 1/2$.

Given arbitrary $\kappa > 2$, there are two possibilities:

Case 2.1. $1/\kappa \leq \mu_X(A) < 1/2$ and $\mu_Y(B) > 1/2$. Then

$$F_s \geq \alpha b \cdot \phi \mu_X(A) > \frac{1}{\kappa} \alpha b \phi. \quad (26)$$

Recall that $\mu_+(A \cup B) \leq 1/2$. Thus

$$\frac{F_s}{\mu_+(A \cup B)} \geq \frac{2}{\kappa} \frac{\eta \phi}{2}. \quad (27)$$

Case 2.2. $\mu_X(A) < 1/\kappa$ and $\mu_Y(B) > 1/2$. Here, we utilize the following fact: $F_s \geq \mathcal{F}(B, A^c) = \mathcal{F}(B, X) - \mathcal{F}(B, A)$. Then, by the definition of flow, we have

$$\mathcal{F}(B, X) = \beta f \mu_Y(B) > \beta f / 2, \quad (28)$$

and by the symmetry of \mathcal{F} (due to reversibility),

$$\mathcal{F}(B, A) = \mathcal{F}(A, B) \leq \mathcal{F}(A, Y) = \alpha b / \kappa. \quad (29)$$

With $\alpha b = \beta f$, combining the results above leads to

$$F_s > \beta f / 2 - \alpha b / \kappa = \alpha b (1/2 - 1/\kappa). \quad (30)$$

As a result, we get

$$\frac{F_s}{\mu_+(A \cup B)} > 2\alpha b (1/2 - 1/\kappa) = \frac{\eta}{2} (1 - 2/\kappa). \quad (31)$$

Case 3. $\mu_X(A) > 1/2$ and $\mu_Y(B) < 1/2$. Following a similar argument as we developed above for case 2, given $\kappa > 2$, we can likewise get

$$\frac{F_s}{\mu_+(A \cup B)} \geq \begin{cases} \frac{2}{\kappa} \frac{\eta \phi}{2} & (\mu_X(A) \geq 1/\kappa), \\ \frac{\eta}{2} (1 - 2/\kappa) & (\mu_X(A) < 1/\kappa). \end{cases} \quad (32)$$

Note that $\mu_X(A) > 1/2$ and $\mu_Y(B) > 1/2$ cannot hold simultaneously under the assumption $\alpha\mu_X(A) + \beta\mu_Y(B) \leq 1/2$. Integrating the results derived for all cases, we obtain

$$\frac{F_s(A, B)}{\mu_+(A, B)} \geq \frac{\eta}{2} \min \left\{ \frac{2}{\kappa} \phi, 1 - \frac{2}{\kappa} \right\}, \quad \forall \kappa > 2. \quad (33)$$

Note that this inequality holds for all A and B with $0 < \mu_+(A \cup B) \leq 1/2$. In this way, we can get a series of lower bound of the bottleneck ratio, using different values of κ . And the supreme of these lower bounds remains a lower bound. It is easy to see that the supreme attains when $2\phi/\kappa = 1 - 2/\kappa$, leading to

$$\sup_{\kappa > 2} \min \left\{ \frac{2}{\kappa} \phi, 1 - \frac{2}{\kappa} \right\} = \frac{\phi}{1 + \phi}. \quad (34)$$

It follows that

$$\Phi_*(\mathbf{P}_+) \geq \frac{\eta}{2} \frac{\phi}{\phi + 1}. \quad (35)$$

This completes the proof of the lower bound. Next, we show the upper bound, which is easier. Due to the definition of bottleneck ratio, for any given partition of $X \cup Y$, the flow ratio derived from that partition constitutes an upper bound of $\Phi_*(\mathbf{P}_+)$.

Here, we consider the partition with one part being X and the other being Y . Then

$$F_s = \alpha b = \beta f, \quad (36)$$

and thus the flow ratio is given by

$$\frac{F_s}{\min(\alpha, \beta)} = \max(f, b). \quad (37)$$

This gives an upper bound of $\Phi_*(\mathbf{P}_+)$. The proof of the theorem is completed.

B.4 Proof of Corollary 1

When both \mathbf{Q}_{BF} and \mathbf{Q}_{FB} are ergodic, their bottleneck ratios are positive. According to Theorem 3, the bottleneck ratio of $\Phi_*(\mathbf{P}_+)$ is thus positive, implying that the spectral gap of \mathbf{P}_+ is positive (by Theorem 2), and hence \mathbf{P}_+ is irreducible.

Since \mathbf{Q}_{BF} is ergodic and thus aperiodic, the greatest common divisor of the loop lengths for each $x \in X$ is 1. Similar argument applies to each $y \in Y$. Consequently, the joint chain characterized by \mathbf{P}_+ is aperiodic.

Being an irreducible and aperiodic finite Markov chain, the joint chain given by \mathbf{P}_+ is ergodic. This completes the proof.

B.5 Proof of Lemma 3

Let $\mathbf{P}' = (\mathbf{P} - \xi \mathbf{I}) / (1 - \xi)$. Since \mathbf{P} has $\mathbf{P}(x, x) > \xi$ for each $x \in X$, the entries of the matrix \mathbf{P}' are all non-negative. In addition,

$$\mathbf{P}' \mathbf{1} = \frac{1}{1 - \xi} (\mathbf{P} - \xi \mathbf{I}) \mathbf{1} = \frac{1}{1 - \xi} (\mathbf{1} - \xi \mathbf{1}) = \mathbf{1}. \quad (38)$$

This implies that \mathbf{P}' is also a stochastic matrix.

Since \mathbf{P} is reversible, all its eigenvalues are real numbers. Without losing generality, we assume they are $\lambda_1 \geq \dots \geq \lambda_n$. As \mathbf{P} is a stochastic matrix, we have $\lambda_1 = 1$ and $\lambda_n \geq -1$. According to the spectral mapping theorem, the eigenvalues of \mathbf{P}' , denoted by $\lambda'_1, \dots, \lambda'_n$, are given by $\lambda'_i = (\lambda_i - \xi) / (1 - \xi)$, for each $i = 1, \dots, n$. As \mathbf{P}' is a stochastic matrix, we have $\lambda'_n \geq -1$, and thus $\frac{\lambda_n - \xi}{1 - \xi} \geq -1$. Therefore, $\lambda_n \geq 2\xi - 1$. The proof is completed.

B.6 Proof of Theorem 4

We show this theorem by progressively proving a series of claims as follows.

Claim 1. *The augmented Markov chain is ergodic.*

With $b_k > 0$ for $k = 0, \dots, K - 1$, the root is accessible from each state (including both complete and partial assignments). With $f_k > 0$ for $k = 1, \dots, K$, each state is accessible from root. These imply that any two states are accessible from each other via the root. Therefore, the chain is irreducible. In addition, $f_K < 1$ makes the chain aperiodic. As this is a finite Markov chain, we can conclude that it is ergodic.

Since the chain is ergodic, it has a unique stationary distribution, *i.e.* its equilibrium distribution. Therefore, it suffices to show that $(\alpha \boldsymbol{\mu}_0, \beta_1 \boldsymbol{\mu}_1, \dots, \beta_K \boldsymbol{\mu}_K)$ that satisfies the three statements in the theorem is a stationary distribution.

Claim 2. *Given vectors $\boldsymbol{\mu}_0, \dots, \boldsymbol{\mu}_K$ respectively over the set of states at level $0, \dots, K$, such that $\boldsymbol{\mu}_0 = \boldsymbol{\mu}$ is a distribution over X , and for each $k = 1, \dots, K$, $\boldsymbol{\mu}_k$ is defined recursively by*

$$\boldsymbol{\mu}_k(y) = \frac{1}{K - (k - 1)} \sum_{x \in Ch(y)} \boldsymbol{\mu}_{k-1}(x), \quad \text{for } y \in Y_k. \quad (39)$$

Then, $\boldsymbol{\mu}_k$ for each $k = 1, \dots, K$ represents a distribution over Y_k , and

$$\boldsymbol{\mu}_k(y) = \binom{K}{k}^{-1} \sum_{x \in X: x \succ y} \boldsymbol{\mu}(x), \quad \forall y \in Y_k. \quad (40)$$

Here, $x \succ y$ means that x is a descendant of y .

The proof of this claim is done by induction as follows.

Obviously, when $k = 1$, according to the definition above, we have

$$\boldsymbol{\mu}_1(y) = \frac{1}{K} \sum_{x \in Ch(y)} \boldsymbol{\mu}_X(x) = \frac{1}{K} \sum_{x \in Ch(y)} \boldsymbol{\mu}(x). \quad (41)$$

This satisfies Eq.(40), as $\binom{K}{1} = K$, and it is clear that $\boldsymbol{\mu}_1$ is non-negative. In addition, we have

$$\begin{aligned} \sum_{y \in Y_1} \boldsymbol{\mu}_1(y) &= \sum_{y \in Y_1} \frac{1}{K} \sum_{x \in Ch(y)} \boldsymbol{\mu}(x) \\ &= \frac{1}{K} \sum_{y \in Y_1} \sum_{x \in Ch(y)} \boldsymbol{\mu}(x) \\ &= \frac{1}{K} \sum_{x \in X} \boldsymbol{\mu}(x) \sum_{y \in Pa(x)} 1 = 1. \end{aligned} \quad (42)$$

Here, $Pa(x)$ is the set of parent states of x . In the derivation above, we use the fact that x has K parents,

i.e. $|Pa(x)| = K$. The identity above implies that μ_1 is a valid distribution over Y_1 . Therefore, the claim holds when $k = 1$.

Suppose that the claim holds for $k = 1, \dots, m$ with $m < K$, we are to show that it also holds for $k = m + 1$, so as to complete the induction. Note that μ_{m+1} is defined as

$$\mu_{m+1}(y) = \frac{1}{K-m} \sum_{x \in Ch(y)} \mu_m(x), \quad \text{for } y \in Y_{m+1}.$$

Again, μ_{m+1} is obviously non-negative, and

$$\begin{aligned} \sum_{y \in Y_{m+1}} \mu_{m+1}(y) &= \frac{1}{K-m} \sum_{y \in Y_{m+1}} \sum_{z \in Ch(y)} \mu_m(z) \\ &= \frac{1}{K-m} \sum_{z \in Y_m} \mu_m(z) \sum_{y \in Pa(x)} 1 = 1. \end{aligned} \quad (43)$$

Similar to the derivation for $k = 1$, here we apply the fact that $|Pa(x)| = K - m$ for each $x \in Y_m$. This shows that μ_{m+1} is a valid distribution over Y_{m+1} . Moreover, we have for each $y \in Y_{m+1}$,

$$\begin{aligned} \mu_{m+1}(y) &= \frac{1}{K-m} \sum_{z \in Ch(y)} \binom{K}{m}^{-1} \sum_{x \in X: x \succ z} \mu(x) \\ &= \frac{1}{K-m} \frac{(K-m)!m!}{K!} \sum_{z \in Ch(y)} \sum_{x \in X: x \succ z} \mu(x) \\ &= \frac{(K-m-1)!m!}{K!} (m+1) \sum_{x \in X: x \succ y} \mu(x) \\ &= \binom{K}{m+1}^{-1} \sum_{x \in X: x \succ y} \mu(x). \end{aligned} \quad (44)$$

By induction, we can conclude that the claim holds for each $k = 1, \dots, K$.

Claim 3. *When the construction is done up to the k -th level, the distribution $\mu_k^+ \triangleq (c_{k,0}\mu_0, \dots, c_{k,k}\mu_k)$ is a stationary of the augmented Markov chain. Here, μ_0, \dots, μ_k are given by Claim 2, and $c_{k,0}, \dots, c_{k,k}$ is defined such that for each $k' = 0, \dots, k$*

$$c_{k,0} = \frac{1}{Z_k}, \quad c_{k,l} = \frac{1}{Z_k} \frac{b_0 \cdots b_{l-1}}{f_1 \cdots f_l}, \quad (45)$$

with

$$Z_k = 1 + \sum_{l=1}^k \frac{b_0 \cdots b_{l-1}}{f_1 \cdots f_l}. \quad (46)$$

We are going to show this claim by induction. Note that $\mu_0 = \mu$ over X is a stationary distribution of \mathbf{P} . And $\mu_0^+ = c_{0,0}\mathbf{p}_0$, thus $c_{0,0} = 1$. It immediately follows that the claim is true for $k = 0$. Suppose this

claim holds for $k = 0, \dots, m$ with $m < K$, we are to show that it also holds for $k = m + 1$. Note from Eq.(45) that

$$\frac{c_{m+1,k}}{c_{m,k}} = \frac{Z_m}{Z_{m+1}}, \quad \forall k = 0, \dots, m, \quad (47)$$

Hence, showing the claim holds for $k = m + 1$ is equivalent to showing that $(\alpha\mu_m^+, \beta\mu_{m+1})$ is a stationary distribution of the augmented chain (up to $(m+1)$ -th level), with

$$\alpha = \frac{Z_m}{Z_{m+1}}, \quad (48)$$

and

$$\beta = \frac{Z_{m+1} - Z_m}{Z_{m+1}} = \frac{1}{Z_{m+1}} \frac{b_0 \cdots b_m}{f_1 \cdots f_{m+1}}. \quad (49)$$

According to Lemma 1, it suffices to check that this distribution satisfies the cross-space detailed balance given in Eq.(2), which is not difficult to verify based on the construction described in section 3.2.

In this proof, Claim 1 proves the ergodicity of the joint chain. Claim 2 constructs a set of vectors μ_0, \dots, μ_K , and states that they are valid distributions over X, Y_0, \dots, Y_K , and satisfy the properties given in (S2). Claim 3 (induction up to $k = K$) shows that the distributions constructed in Claim 2 is exactly a stationary distribution of the joint chain. Since the chain is ergodic, this is the equilibrium distribution. As a by product, Claim 3 also shows the the statement (S3) of the theorem. For (S1), it is automatically established by the construction described in Claim 2. Therefore, we can conclude that the proof of the theorem has been completed.

B.7 Proof of Corollary 2

Based on Theorem 4, we have

$$\frac{1}{\alpha} = 1 + \sum_{k=1}^K \frac{b_0 \cdots b_{k-1}}{f_1 \cdots f_k} \leq 1 + \sum_{k=1}^{\infty} \kappa^k = \frac{1}{1-\kappa}. \quad (50)$$

Hence, $\alpha \geq 1 - \kappa$. The proof is done.