Discrepancy Analysis of a New Randomized Diffusion Algorithm for Weighted Round Matrices

Takeharu Shiraga*

February 22, 2018

Abstract

For an arbitrary initial configuration of indivisible (discrete) loads over vertices of a distributed network (graph), we consider the problem of minimizing the discrepancy between the maximum and minimum load among all vertices. For this problem, diffusion-based algorithms are well studied because of its simplicity. In diffusion-based algorithms, each vertex distributes its loads as evenly as possible among its neighbors in each synchronous round.

This paper presents a new randomized diffusion-based algorithm inspired by multiple random walks. In multiple random walks, at each vertex, each token (load) generates a random variable in $[0, 1)$, and move to a vertex corresponding to the given probability distribution (transition matrix). In our algorithm, at each vertex, each token k ($k \in \{0, 1, \ldots, X-1\}$) generate a random number in $[k/X, (k+1)/X)$, and moves to a vertex corresponding to the given probability distribution. Our algorithm is adaptive to any transition transition probabilities while almost all previous works are concerned with uniform transition probabilities.

For this algorithm, we analyze the discrepancy between the token configuration and its expected value, and give an upper bound depending on the local 2-divergence of the transition matrix and $\sqrt{\log n}$, where n is the number of vertices. The local 2-divergence is a measure which often appeared in previous works. We also give an upper bound of the local-2 divergence for any reversible and lazy transition matrix.

These yield the following specific results. First, our algorithm achieves $O(\sqrt{d \log n})$ discrepancy for any d regular graph, which matches the best result on previous works of diffusion model. Note that our algorithm does not need any assumption of the number of tokens such as negative loads which are often assumed in previous works. Second, for general graphs with maximum degree $d_{\max}$, our algorithm achieves $O(\sqrt{d_{\max} \log n})$ discrepancy using the transition matrix based on the metropolis hasting algorithm. Note that this algorithm does not need information of $d_{\max}$ while almost all previous works use it.

Key words: Load balancing, Diffusion, Markov chain

1 Introduction

This paper is concerned with the load balancing algorithms on distributed networks. Let $G = (V, E)$ be an undirected and connected network (graph), and $\mathcal{X}^{(0)} \in \mathbb{Z}^n_{\geq 0}$ be an initial configuration of loads (tokens) over V, where $n = |V|$. Then, we consider a distributed and iterative algorithm to balance the tokens over vertices, i.e., each vertex moves its tokens to its neighbors iteratively at each discrete and synchronous time step to minimize the *discrepancy* between the maximum and minimum tokens among all vertices.

*Department of Information and System Engineering, Faculty of Science and Engineering, Chuo University, Tokyo, Japan
shiraga@ise.chuo-u.ac.jp

1.1 Previous works

Load balancing algorithms studied in previous works are classified as diffusion-based algorithms and matching-based algorithms. Roughly speaking, each vertex sends its tokens to neighbors as evenly as possible (according to *round matrix*) at each round in diffusion-based algorithms. Matching-based algorithms generate a (different) matching of the graph in a distributed way at each round, and the endpoints of each matching balances tokens as evenly as possible.

Rabani et al. [10] studied deterministic diffusion-based algorithms and matching-based algorithms, and gave a framework of the analysis of these models. They introduced the local 1-divergence and show that the discrepancy is upper bounded by the local 1-divergence within $\mathrm{O}(\log(Kn)/(1-\lambda))$ step, where $K = \max_{x,y}|\mathcal{X}_x^{(0)} - \mathcal{X}_y^{(0)}|$ is the initial discrepancy and λ is the second largest eigenvalue of the round matrix. They also showed that the local 1-divergence is upper bounded by $\mathrm{O}(d\log n/(1-\lambda))$ for any d regular graph.

Randomization is a natural approach to get smaller discrepancy. Friedrich and Sauerwald [6] studied randomized version of matching models. They showed that the discrepancy is upper bounded by the local 2-divergence times $\sqrt{\log n}$ with in $\mathrm{O}(\log(Kn)/(1-\lambda))$ step and the local 2-divergence is upper bounded by $\mathrm{O}(\sqrt{d/(1-\lambda)})$.

Berenbrink et al. [3] studied a randomized diffusion-based algorithm. They showed that the discrepancy is upper bounded by $\mathrm{O}(d\max\{\sqrt{\log n}, \log\log n/(1-\lambda)\})$ within $\mathrm{O}(\log(Kn)/(1-\lambda))$ step for any d regular graph.

Akbari and Berenbrink [1] studied randomized and deterministic diffusion algorithm based on rotor-router mechanism. They gave same bound of [3] for their randomized diffusion algorithm using fewer random bits compared to [3]. Furthermore, they gave upper bounds of deterministic diffusion algorithm on specific graphs, $\mathrm{O}(d^{1.5})$ for hypercube and $\mathrm{O}(1)$ for constant-dimensional tori.

The results of Sauerwald and Sun [11] is the best result of the discrepancy so far. They showed that constant discrepancy within $\mathrm{O}(\log(Kn)/(1-\lambda))$ step for a randomized matching model.

Recently, Berenbrink et al. [4] studied deterministic diffusion based algorithms. They showed that $\mathrm{O}(d)$ discrepancy within $\mathrm{O}((\log(Kn) + d\log^2 n)/(1-\lambda))$ step. This result improves [10, 1].

There are some other results concerned with the randomized diffusion-based algorithms which use negative loads [5, 2, 11]. In [11], authors showed that $\mathrm{O}(d_{\max}\log n)$ discrepancy within $\mathrm{O}(\log(Kn)/(1-\lambda))$ step for any graphs.

1.2 This work

In this paper, we consider the following diffusion-based algorithm. Let P be an arbitrary transition matrix over V, which is related to the round matrix in previous works. For example, $P_{x,y} = 1/d_x$ if $(x, y) \in E$, where d_x is the degree size of x. Now, let $\mathcal{X}_v^{(t)}$ denote the number of tokens on v at time t in our algorithm.

Algorithm. At each time step t and at each vertex $v \in V$, each token k ($k \in \{0, 1, \dots, \mathcal{X}_v^{(t)} - 1\}$) generates a random number in $[k/\mathcal{X}_v^{(t)}, (k+1)/\mathcal{X}_v^{(t)})$ uniformly at random. Then, each token moves to its corresponding neighbor.

See Figure 1 for an example. There are 5 tokens ($k = 0, 1, 2, 3, 4$) on v and $(P_{v,v_0}, P_{v,v_1}, P_{v,v_2}, P_{v,v_3}) = (1/10, 3/5, 1/5, 1/10)$. In this example, token 0 moves to v_0 (v_1) with probability $1/2$, token 1 (2) moves to v_1 with probability 1, token 3 moves to v_1 (v_2) with probability $1/2$, and token 4 moves to v_2 (v_3) with probability $1/2$.

More precisely, we assume that for each $v \in V$, each u is mapped into a interval in $[0, 1)$ whose length is $P_{v,u}$. Then each token k moves to u if the generated random number in $[k/\mathcal{X}_v^{(t)}, (k+1)/\mathcal{X}_v^{(t)})$ is in the interval of u. (See equation (1) in Section 2.2 for the concrete definition).

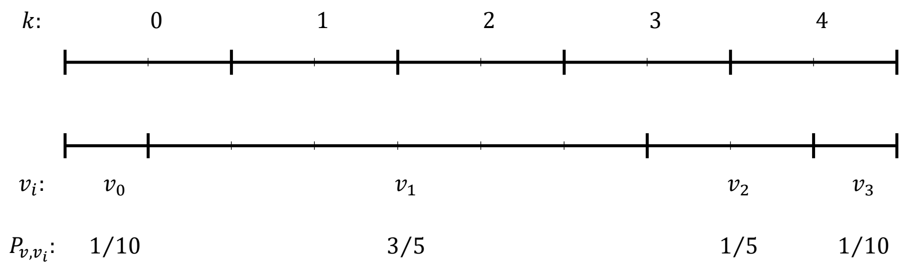

Figure 1:

This algorithm is quite simple. At each round, each vertex only sends its tokens using the number of tokens on it and the probability distribution around it (e.g., its degree size), does not need to communicate to neighbors or use the number of tokens of its neighbor.

Furthermore, this algorithm allows to use arbitrary transition matrix. Almost all previous works are corresponding to "uniform" transition matrices, i.e., $P_{v,x} = P_{v,y}$ holds for any x, y in the neighboring set of v ($x, y \neq v$).

The main result of this paper is the following general upper bound.

Main results in general form. First, we introduce the local p-divergence of P.

Definition 1.1 (local p-divergence, [10, 6]). *For any $p \in \mathbb{Z}_{\geq 0}$, the local-p divergence of P is defined by*

$$\Psi_p(P) := \max_{w \in V} \left(\sum_{t=0}^{\infty} \sum_{\substack{(v,u) \in V \times V \\ : P_{v,u} > 0}} |P^t_{v,w} - P^t_{u,w}|^p \right)^{1/p}.$$

Then, we show the following main theorem.

Theorem 1.2. *For any initial configuration $\mathcal{X}^{(0)}$ and transition matrix P,*

$$\mathbf{Pr}\left[\max_{w \in V} \left| \mathcal{X}^{(T)}_w - (\mathcal{X}^{(0)} P^T)_w \right| \leq 4\Psi_2(P)\sqrt{\ln n} \right] \geq 1 - \frac{2}{n}$$

holds for each time $T \in \mathbb{Z}_{\geq 0}$.

Note that $\mathcal{X}^{(0)} P^T$ is the expected value of $\mathcal{X}^{(T)}$ (See Section 3.2 for details). $\mathcal{X}^{(0)} P^T$ is converge to $\|\mathcal{X}^{(0)}\|_1 \pi$, where π is the stationary distribution [1] of P (See Appendix A for details).

Second, we obtain the following upper bound of the local 2-divergence for weighted P.

Theorem 1.3. *Suppose that P is reversible [2] and lazy [3]. Then,*

$$\Psi_2(P) \quad \leq \quad \sqrt{\frac{2 \max_{w \in V} \pi_w}{\min_{(v,u) \in E^P_\circ} \pi_v P_{v,u}}}$$

holds, where $E^P_\circ = \{(v,u) \in V \times V \mid P_{v,u} > 0 \text{ and } v \neq u\}$.

[1] Probability distribution such that $\pi P = \pi$ holds.

[2] P is reversible if the detailed balance equation $\pi_v P_{v,u} = \pi_u P_{u,v}$ holds for any $u, v \in V$.

[3] P is lazy if $P_{v,v} \geq 1/2$ holds for any $v \in V$.

This is a generalized version of the previous result [11], which is corresponding to uniform P.

These results in general form give the following specific result.

Contribution for regular graphs. Now, we consider the lazy transition matrix P_{L} on d-regular graph: $(P_{\mathrm{L}})_{x,x} = 1/2$, $(P_{\mathrm{L}})_{x,y} = 1/2d$ if $(x,y) \in E$ and $(P_{\mathrm{L}})_{x,y} = 0$ otherwise. For this transition matrix, we obtain the following.

Corollary 1.4. *Suppose that $G = (V,E)$ is an arbitrary connected d-regular graph and the transition matrix is P_{L}. Then, for each $T \geq \frac{\log(2Kn)}{1-\lambda}$, it holds that*

$$\mathbf{Pr}\left[\max_{x,y\in V} |\mathcal{X}_x^{(T)} - \mathcal{X}_y^{(T)}| \leq 16\sqrt{d \ln n}\right] \geq 1 - \frac{2}{n}.$$

This upper bound improves the previous result of [3]. Corollary 1.4 achieves the same upper bound of the diffusion model in [11] without negative loads.

Contribution for general graphs. Since our algorithm allows to use any transition matrix, we can use the Metropolis chain P_{M} for general graphs, where $(P_{\mathrm{M}})_{x,y} = (1/2)\min\{1/d_x, 1/d_y\}$ if $(x,y) \in E$, $(P_{\mathrm{M}})_{x,y} = 0$ if $(x,y) \notin E$ and the self loop is equal to the remaining probability [9]. For this transition matrix, we obtain the following.

Corollary 1.5. *Suppose that $G = (V,E)$ is an arbitrary connected graph and the transition matrix is P_{M}. Then, for each $T \geq \frac{\log(2Kn)}{1-\lambda}$, it holds that*

$$\mathbf{Pr}\left[\max_{x,y\in V} |\mathcal{X}_x^{(T)} - \mathcal{X}_y^{(T)}| \leq 16\sqrt{d_{\max} \ln n}\right] \geq 1 - \frac{2}{n}.$$

This also matches the previous result of the diffusion model on general graph [11] without negative loads. Furthermore, each vertex only needs the degree of its neighbors to calculate $(P_{\mathrm{M}})_{x,y} = (1/2)\max\{d_x, d_y\}$, while previous work [11] needs $d_{\max}$, which is the maximum value of the degree among all vertices.

2 Notations and Model description

In this section we describe our model precisely.

2.1 Notations

Let V be a vertex set, and let $n = |V|$. Let $P \in [0,1]^{n\times n}$ be a transition matrix on V, i.e., $\sum_{u\in V} P_{v,u} = 1$ holds for any $v \in V$, where $P_{v,u}$ denotes (v,u) entry of P. $P^t_{v,u}$ denotes (v,u) entry of P^t. In this paper we assume that P^0 is the identity matrix.

Let N_v^P be the set of neighbors of $v \in V$, i.e., $N_v^P := \{u \in V \mid P_{v,u} > 0\}$. In this paper, we assume an arbitrary ordering on N_v^P, i.e., we denote $N_v^P = \{v_0, v_1, \ldots, v_{d_v^P - 1}\}$, where $d_v^P = |N_v^P|$. Let E^P be the set of edges of transition diagram of P, i.e., $E^P := \{(v,u) \in V \times V \mid P_{v,u} > 0\}$. Note that E^P may contain self-loop edges.

2.2 Model description

Let $\mathcal{X}^{(0)} \in \mathbb{Z}^n_{\geq 0}$ be a initial configuration of M tokens over V, and let $\mathcal{X}^{(t)} \in \mathbb{Z}^n_{\geq 0}$ denote the configuration of M tokens over V at time $t \in \mathbb{Z}_{\geq 0}$ in our algorithm. In an update from $\mathcal{X}^{(t)}$ to $\mathcal{X}^{(t+1)}$ of our algorithm, at each vertex $v \in V$, we generate $\mathcal{X}_v^{(t)}$ random numbers $r_v^{(t)}(0), r_v^{(t)}(1), \ldots, r_v^{(t)}(\mathcal{X}_v^{(t)} - 1)$, where each

$r_v^{(t)}(k)$ is picked from $\left[\frac{k}{\mathcal{X}_v^{(t)}}, \frac{k+1}{\mathcal{X}_v^{(t)}}\right)$ uniformly at random. From these random numbers, we define destination operators $\mathcal{D}_v^{(t)}(k)$ such that

$$\mathcal{D}_v^{(t)}(k) = v_i \text{ if } r_v^{(t)}(k) \in \left[\sum_{j=0}^{i-1} P_{v,v_j}, \sum_{j=0}^{i} P_{v,v_j}\right). \tag{1}$$

$\mathcal{D}_v^{(t)}(k)$ denotes the destination of $(k+1)$-th token on $v \in V$ at time t. Then, let

$$\mathcal{X}_v^{(t+1)} \ := \ \sum_{u \in V} \sum_{k=0}^{\mathcal{X}_u^{(t)}-1} \mathbf{1}\{\mathcal{D}_u^{(t)}(k) = v\}. \tag{2}$$

According to the above definition, destination operators $\mathcal{D}_v^{(t)}(k)$ satisfy the following properties.

Observation 2.1. *Suppose that $\mathcal{X}^{(t)}$ is fixed. Then, for any $v, x \in V$, $k \in \{0, 1, \dots \mathcal{X}_v^{(t)} - 1\}$ and $\ell \in \{0, 1, \dots \mathcal{X}_x^{(t)} - 1\}$, $\mathcal{D}_v^{(t)}(k)$ and $\mathcal{D}_x^{(t)}(\ell)$ are independent if $v \neq x$ or $k \neq \ell$.*

Observation 2.2.

$$\mathbf{Pr}\left[\mathcal{D}_v^{(t)}(k) \ = \ v_i \mid \mathcal{X}_v^{(t)}\right] \ = \ \left|\left[\sum_{j=0}^{i-1} P_{v,v_j}, \sum_{j=0}^{i} P_{v,v_j}\right) \cap \left[\frac{k}{\mathcal{X}_v^{(t)}}, \frac{k+1}{\mathcal{X}_v^{(t)}}\right)\right| \cdot \mathcal{X}_v^{(t)}$$

This paper is concerned with the behavior of $\mathcal{X}^{(t)}$ defined by destination operators $\mathcal{D}_v^{(t)}(k)$ satisfying Observation 2.1 and 2.2.

3 Proof of Theorem 1.2

This section gives the proof of Theorem 1.2.

Theorem 3.1 (Theorem 1.2)**.** *For any initial configuration $\mathcal{X}^{(0)}$ and transition matrix P,*

$$\mathbf{Pr}\left[\max_{w \in V}\left|\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P^T)_w\right| \leq 4\Psi_2(P)\sqrt{\ln n}\right] \geq 1 - \frac{2}{n}$$

holds for each time $T \in \mathbb{Z}_{\geq 0}$.

3.1 Framework of the proof

First, we introduce some notations for the proof. Let $V = \{0, 1, \dots, n-1\}$. For any $t \in \{0, 1, \dots, T-1\}$, $v \in V$, and $k \in \{0, 1, \dots, M-1\}$, let

$$l_v^{(t)}(k) \ := \ (Mn)t + (M)v + k, \tag{3}$$

and let

$$\mathrm{D}_{l_v^{(t)}(k)} \ := \begin{cases} \mathcal{D}_v^{(t)}(k) & (\text{if } k \in \{0, 1, \dots, M-1\}) \\ -1 & (\text{otherwise}) \end{cases}. \tag{4}$$

This definition means that $\mathbf{Pr}[\mathrm{D}_{l_v^{(t)}(k)} = u \mid \mathcal{X}_v^{(t)}] = 0$ for any $k \geq \mathcal{X}_v^{(t)}$ and $u \in V$. For the notational convenience, let

$$\mathbf{D}_\ell \;:=\; (\mathrm{D}_l)_{l<\ell} \;=\; \mathrm{D}_0, \mathrm{D}_1, \ldots, \mathrm{D}_{\ell-1}, \tag{5}$$

which is a sequence of random variables. Then, we observe the following from the definition of the configuration of tokens (2).

Observation 3.2. *For any $t \in \mathbb{Z}_{\geq 0}$, $\mathcal{X}^{(t)}$ is determined by $\mathcal{X}^{(0)}$ and $\mathbf{D}_{Mnt} = \mathrm{D}_0, \mathrm{D}_1, \ldots, \mathrm{D}_{l_{n-1}^{(t-1)}(M-1)}$.*

Now, let

$$\mathcal{Y}_\ell \;=\; \mathcal{Y}_\ell(w, T) \;=\; \mathbf{E}\left[\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)} P^T)_w \mid \mathbf{D}_\ell\right]. \tag{6}$$

We define $\mathcal{Y}_0 := \mathbf{E}[\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)} P^T)_w]$. Note that $\mathcal{Y}_0, \mathcal{Y}_1, \ldots$ is a martingale with respect to $\mathrm{D}_0, \mathrm{D}_1, \ldots$, thus

$$\mathbf{Pr}\big[|\mathcal{Y}_{MnT} - \mathcal{Y}_0| \geq \eta\big] \leq 2\exp\left[-\eta^2/2\sum_{\ell=0}^{MnT-1}(c_\ell)^2\right] \tag{7}$$

holds from Azuma-Hoeffding inequality (See Appendix A), where c_ℓ is a value satisfies $|\mathcal{Y}_{\ell+1} - \mathcal{Y}_\ell| \leq c_\ell$.

Since $\mathcal{Y}_{MnT} = \mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)} P^T)_w$ from Observation 3.2 and $\mathcal{Y}_0 = \mathbf{E}[\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)} P^T)_w] = 0$ (See Lemma 3.4 and (17) in Section 3.2 for the detail),

$$\mathbf{Pr}\left[|\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)} P^T)_w| \geq 2\sqrt{\sum_{\ell=0}^{MnT-1}(c_\ell)^2 \ln n}\right] \leq \frac{2}{n^2} \tag{8}$$

holds by taking $\eta = \sqrt{2\sum_{\ell=0}^{MnT-1}(c_\ell)^2 \ln n^2}$. Thus, using the union bound to (8) and apply the following lemma, we obtain Theorem 1.2.

Lemma 3.3. *For any $T \in \mathbb{Z}_{\geq 0}$ and $w \in V$, it holds that*

$$\sum_{\ell=0}^{MnT-1}\big(\mathcal{Y}_{\ell+1}(w,T) - \mathcal{Y}_\ell(w,T)\big)^2 \;\leq\; 4\big(\Psi_2(P)\big)^2.$$

To complete the proof, we prove Lemma 3.3 in the following subsection.

3.2 Proof of Lemma 3.3

For the notational convenience, let

$$\widetilde{I}(X,k) \;:=\; \begin{cases} \emptyset & (\text{if } X = 0) \\ \big[\frac{k}{X}, \frac{k+1}{X}\big) & (\text{otherwise}) \end{cases} \tag{9}$$

and

$$\widetilde{P}_{v,v_i} \;:=\; \left[\sum_{j=0}^{i-1} P_{v,v_j}, \sum_{j=0}^{i} P_{v,v_j}\right). \tag{10}$$

First, we introduce the following lemma inspired by Lemma 4.1 in [12]. We use the definition of our algorithm (2) to prove Lemma 3.4.

Lemma 3.4. *For any P, $\mathcal{X}^{(0)}$, $w \in V$ and $T \in \mathbb{Z}_{\geq 0}$, it holds that*

$$\begin{aligned}&\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P^T)_w \\ &\quad = \sum_{t=0}^{T-1}\sum_{v\in V}\sum_{u\in N_v^P}\sum_{k=0}^{M-1}\left(\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} - \mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u}\cap \widetilde{I}(\mathcal{X}_v^{(t)},k)\right|\right)\left(P_{u,w}^{T-t-1} - P_{v,w}^{T-t-1}\right).\end{aligned}$$

Proof. It is not difficult to see that

$$\begin{aligned}\sum_{t=0}^{T-1}\left(\mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}P\right)P^{T-t-1} &= \sum_{t=0}^{T-1}\left(\mathcal{X}^{(t+1)}P^{T-t-1} - \mathcal{X}^{(t)}P^{T-t}\right) \\ &= \mathcal{X}^{(T)}P^0 - \mathcal{X}^{(0)}P^T \;=\; \mathcal{X}^{(T)} - \mathcal{X}^{(0)}P^T. \end{aligned}\tag{11}$$

From (11) we obtain

$$\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P^T)_w = \sum_{t=0}^{T-1}\sum_{u\in V}\left(\mathcal{X}_u^{(t+1)} - (\mathcal{X}^{(t)}P)_u\right)P_{u,w}^{T-t-1}. \tag{12}$$

Then, from the definitions (2) and (4) we have

$$\mathcal{X}_u^{(t+1)} = \sum_{v\in V}\sum_{k=0}^{\mathcal{X}_v^{(t)}-1}\mathbf{1}\{\mathcal{D}_v^{(t)}(k) = u\} = \sum_{v\in V}\sum_{k=0}^{M-1}\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\}. \tag{13}$$

We also have

$$(\mathcal{X}^{(t)}P)_u = \sum_{v\in V}\mathcal{X}_v^{(t)}P_{v,u} = \sum_{v\in V}\mathcal{X}_v^{(t)}\sum_{k=0}^{M-1}\left|\widetilde{P}_{v,u}\cap\widetilde{I}(\mathcal{X}_v^{(t)},k)\right| \tag{14}$$

from the definitions (9) and (10). Finally, we have

$$\sum_{u\in N_v^P}\left(\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} - \mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u}\cap\widetilde{I}(\mathcal{X}_v^{(t)},k)\right|\right)P_{v,w} = 0 \tag{15}$$

since

$$\sum_{u\in N_v^P}\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} = \sum_{u\in N_v^P}\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u}\cap\widetilde{I}(\mathcal{X}_v^{(t)},k)\right| = 1. \tag{16}$$

Combining (12), (13) and (14) and subtracting (15), we obtain the claim. ☐

Note that from Observation 2.2 and the chain rule of the conditional expectation,

$$\begin{aligned}\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\}\right] &= \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathcal{X}_v^{(t)}\right]\right] \\ &= \mathbf{E}\left[\mathbf{Pr}\left[\mathrm{D}_{l_v^{(t)}(k)} = u \mid \mathcal{X}_v^{(t)}\right]\right] \\ &= \mathbf{E}\left[\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u}\cap\widetilde{I}(\mathcal{X}_v^{(t)},k)\right|\right]\end{aligned}\tag{17}$$

holds, thus we have $\mathcal{Y}_0 = \mathbf{E}[\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P^T)_w] = 0$.

Next, we show the following lemma using Lemma 3.4.

Lemma 3.5. *For any $\tau \in \{0,1,\dots,T-1\}$, $\nu \in V$ and $\kappa \in \{0,1,\dots,\mathcal{X}_\nu^{(\tau)}-1\}$, it holds that*

$$\mathcal{Y}_{l_\nu^{(\tau)}(\kappa)+1} - \mathcal{Y}_{l_\nu^{(\tau)}(\kappa)} = \sum_{u \in N_\nu^P} \left(\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right) \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right).$$

Proof. From the definition of $\mathcal{Y}_\ell$ (6) and Lemma 3.4, we have

$$\mathcal{Y}_{l_\nu^{(\tau)}(\kappa)+1} - \mathcal{Y}_{l_\nu^{(\tau)}(\kappa)} = \sum_{t=0}^{T-1} \sum_{v=0}^{n-1} \sum_{k=0}^{M-1} \sum_{u \in N_v^P} (*) \left(P_{u,w}^{T-t-1} - P_{v,w}^{T-t-1} \right), \tag{18}$$

where

$$\begin{aligned} * \;=\; & \mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1} \right] \\ & - \mathbf{E}\left[\mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1} \right] \\ & - \mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] \\ & + \mathbf{E}\left[\mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right]. \end{aligned} \tag{19}$$

To obtain the claim, let we start with showing

$$\begin{aligned} & \mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] \\ = & \begin{cases} \mathbf{E}\left[\mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] & \left(\text{if } l_v^{(t)}(k) \geq l_\nu^{(\tau)}(\kappa)\right) \\ \mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} & (\text{otherwise}) \end{cases}. \end{aligned} \tag{20}$$

Obviously, $\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] = \mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\}$ holds if $l_v^{(t)}(k) < l_\nu^{(\tau)}(\kappa)$. Hence we consider if $l_v^{(t)}(k) \geq l_\nu^{(\tau)}(\kappa)$ holds. In this case, from the chain rule of the conditional expectation,

$$\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] = \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_v^{(t)}(k)} \right] \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] \tag{21}$$

holds. From Observation 2.1, we have

$$\begin{aligned} (21) &= \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_0^{(t)}(0)} \right] \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] \\ &= \mathbf{E}\left[\mathbf{Pr}\left[\mathrm{D}_{l_v^{(t)}(k)} = u \mid \mathbf{D}_{Mnt} \right] \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right]. \end{aligned} \tag{22}$$

Observation 3.2 says that $\mathbf{D}_{Mnt}$ determines $\mathcal{X}_v^{(t)}$. Using Observation 2.2, we have

$$\mathbf{Pr}\left[\mathrm{D}_{l_v^{(t)}(k)} = u \mid \mathbf{D}_{Mnt} \right] = \mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right|. \tag{23}$$

Thus combining equations (21)-(23), we obtain (20).

Note that Observation 3.2 also says that

$$\mathbf{E}\left[\mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)} \right] = \mathcal{X}_v^{(t)} \left| \widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k) \right| \; (\text{if } t \leq \tau) \tag{24}$$

holds.

Now we prove the lemma using above discussion. We consider the following 3 cases.

case 1. [$t \leq \tau - 1$] or [$t = \tau$ and $v \leq \nu - 1$] or [$t = \tau$, $v = \nu$ and $k \leq \kappa - 1$]:

In this case, $l_v^{(t)}(k) \geq l_\nu^{(\tau)}(\kappa) + 1$ holds. Using (20), we have

$$(*) \quad = \quad 0 \tag{25}$$

in **case 1**.

case 2. [$t = \tau$, $v = \nu$ and $k \geq \kappa + 1$] or [$t = \tau$ and $v \geq \nu + 1$] or [$t \geq \tau + 1$]:

In this case, we have $l_v^{(t)}(k) \leq l_\nu^{(\tau)}(\kappa) - 1$. From (20), we have

$$\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1}\right] \quad = \quad \mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_v^{(t)}(k)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)}\right]. \tag{26}$$

Furthermore,

$$\mathbf{E}\left[\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k)\right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1}\right] \quad = \quad \mathbf{E}\left[\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k)\right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)}\right] \tag{27}$$

holds from (24). Thus

$$(*) \quad = \quad 0 \tag{28}$$

in **case 2**.

case 3. [$t = \tau$, $v = \nu$ and $k = \kappa$]: First, we have

$$\mathbf{E}\left[\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k)\right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1}\right] \quad = \quad \mathbf{E}\left[\mathcal{X}_v^{(t)}\left|\widetilde{P}_{v,u} \cap \widetilde{I}(\mathcal{X}_v^{(t)}, k)\right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)}\right] \tag{29}$$

from (24). We also have

$$\mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)+1}\right] = \mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} \tag{30}$$

and

$$\begin{aligned} \mathbf{E}\left[\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)}\right] &= \mathbf{E}\left[\mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| \mid \mathbf{D}_{l_\nu^{(\tau)}(\kappa)}\right] \\ &= \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| \end{aligned} \tag{31}$$

from (20). Thus

$$(*) \quad = \quad \mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| \tag{32}$$

in **case 3**.

Combining (18), (25), (28) and (32), we obtain the claim.

□

Proof of Lemma 3.3. First, we observe that

$$\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| = \begin{cases} 1 - 1 = 0 & (\text{if } \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| = 1) \\ 0 - 0 = 0 & (\text{if } \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| = 0) \end{cases} \tag{33}$$

since $\mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right| = \mathbf{Pr}\left[\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u \mid \mathcal{X}_\nu^{(\tau)}\right]$ from Observation 2.2. Furthermore, we have

$$\begin{aligned} &\sum_{u \in N_\nu^P} \left|\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right|\right| \\ &\quad \leq \sum_{u \in N_\nu^P} \left|\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\}\right| + \sum_{u \in N_\nu^P} \left|\mathcal{X}_\nu^{(\tau)}\left|\widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa)\right|\right| \\ &\quad = 1 + 1 = 2. \end{aligned} \tag{34}$$

Now, let

$$S_\nu^{(\tau)}(\kappa) = \left\{ u \in N_v^P \mid 0 < \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| < 1 \right\}. \tag{35}$$

Note that if $u \neq S_\nu^{(\tau)}(\kappa)$, then $\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| = 0$ since (33) holds. Combining Lemma 3.5, Cauchy–Schwarz and (34), we have

$$\begin{aligned}
&\left(\mathcal{Y}_{l_\nu^{(\tau)}(\kappa)+1} - \mathcal{Y}_{l_\nu^{(\tau)}(\kappa)} \right)^2 \\
&= \left(\sum_{u \in N_\nu^P} \left(\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right) \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right) \right)^2 \\
&= \left(\sum_{u \in N_\nu^P} \left(\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right) \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right) \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\} \right)^2 \\
&\leq \sum_{u \in N_\nu^P} \left(\mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right)^2 \sum_{u \in N_\nu^P} \left(\left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right) \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\} \right)^2 \\
&\leq \sum_{u \in N_\nu^P} \left| \mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right| \sum_{u \in N_\nu^P} \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right)^2 \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\} \\
&\leq 2 \sum_{u \in N_\nu^P} \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right)^2 \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\}.
\end{aligned} \tag{36}$$

The second inequality holds since $\left| \mathbf{1}\{\mathrm{D}_{l_\nu^{(\tau)}(\kappa)} = u\} - \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| \right| \leq 1$. Thus

$$\begin{aligned}
\sum_{\ell=0}^{MnT-1} (\mathcal{Y}_{\ell+1} - \mathcal{Y}_\ell)^2 &= \sum_{\tau=0}^{T-1} \sum_{\nu=0}^{n-1} \sum_{\kappa=0}^{M-1} \left(\mathcal{Y}_{l_\nu^{(\tau)}(\kappa)+1} - \mathcal{Y}_{l_\nu^{(\tau)}(\kappa)} \right)^2 \\
&\leq \sum_{\tau=0}^{T-1} \sum_{\nu=0}^{n-1} \sum_{\kappa=0}^{M-1} 2 \sum_{u \in N_\nu^P} \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right)^2 \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\} \\
&= 2 \sum_{\tau=0}^{T-1} \sum_{\nu=0}^{n-1} \sum_{u \in N_\nu^P} \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right)^2 \sum_{\kappa=0}^{M-1} \mathbf{1}\{u \in S_\nu^{(\tau)}(\kappa)\} \\
&\leq 4 \sum_{\tau=0}^{T-1} \sum_{\nu=0}^{n-1} \sum_{u \in N_\nu^P} \left(P_{u,w}^{T-\tau-1} - P_{\nu,w}^{T-\tau-1} \right)^2 \\
&= 4 \sum_{\tau=0}^{T-1} \sum_{(u,\nu) \in E^P} \left(P_{u,w}^{\tau} - P_{\nu,w}^{\tau} \right)^2 \\
&\leq 4(\Psi_2(P))^2
\end{aligned}$$

holds, and we obtained the claim. Note that the second inequality holds since for each $u \in N_\nu^P$, the number of κ such that $S_\nu^{(\tau)}(\kappa) \ni u$ ($0 < \mathcal{X}_\nu^{(\tau)} \left| \widetilde{P}_{\nu,u} \cap \widetilde{I}(\mathcal{X}_\nu^{(\tau)}, \kappa) \right| < 1$) holds is at most two. (For example, see Figure 1. If $u = v_1$, such κ is 0 and 3.)

□

4 Upper bound of the local 2-divergence and specific results

This section shows Theorem 1.3 and other specific results.

Theorem 4.1 (Theorem 1.3)**.** *Suppose that P is reversible and lazy. Then,*

$$\Psi_2(P) \quad \leq \quad \sqrt{\frac{2\max_{w\in V}\pi_w}{\min_{(v,u)\in E_\circ^P}\pi_v P_{v,u}}}$$

holds, where $E_\circ^P = \{(v,u) \in V \times V \mid P_{v,u} > 0 \text{ and } v \neq u\}$.

4.1 Proof of Theorem 1.3

To prove Theorem 1.3, we introduce the following lemma. We use the reversibility of P to prove Lemma 4.2.

Lemma 4.2. *For any reversible P, it holds that*

$$\sum_{v\in V}\sum_{u\in N_v^P}\pi_v P_{v,u}(P^t_{u,w}-P^t_{v,w})^2 \quad = \quad 2\pi_w(P^{2t}_{w,w}-P^{2t+1}_{w,w}).$$

Proof. From the definition of N_v^P, we have

$$\begin{aligned}
\sum_{v\in V}\sum_{u\in N_v^P}\pi_v P_{v,u}(P^t_{u,w}-P^t_{v,w})^2 &= \sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{u,w}-P^t_{v,w})^2 \\
&= \sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}\left((P^t_{u,w})^2+(P^t_{v,w})^2-2P^t_{u,w}P^t_{v,w}\right) \\
&= \sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{u,w})^2+\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{v,w})^2 \\
&\quad -2\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}P^t_{u,w}P^t_{v,w}. \qquad (37)
\end{aligned}$$

Then, from the reversibility of P,

$$\begin{aligned}
\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{u,w})^2+\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{v,w})^2 &= \sum_{v\in V}\sum_{u\in V}\pi_u P_{u,v}(P^t_{u,w})^2+\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}(P^t_{v,w})^2 \\
&= \sum_{u\in V}\pi_u(P^t_{u,w})^2+\sum_{v\in V}\pi_v(P^t_{v,w})^2 \\
&= \sum_{u\in V}\pi_w P^t_{w,u}P^t_{u,w}+\sum_{v\in V}\pi_w P^t_{w,v}P^t_{v,w} \\
&= 2\pi_w P^{2t}_{w,w}, \qquad (38)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{v\in V}\sum_{u\in V}\pi_v P_{v,u}P^t_{u,w}P^t_{v,w} &= \pi_w\sum_{v\in V}\sum_{u\in V}P^t_{w,v}P_{v,u}P^t_{u,w} \\
&= \pi_w P^{2t+1}_{w,w} \qquad (39)
\end{aligned}$$

holds. Thus we obtain the claim. □

Proof of Theorem 1.3. From Lemma 4.2, we have

$$\sum_{t=0}^{\infty}\sum_{v\in V}\sum_{u\in N_v^P}\pi_v P_{v,u}(P_{u,w}^t - P_{v,w}^t)^2 = 2\pi_w\sum_{t=0}^{\infty}(P_{w,w}^{2t} - P_{w,w}^{2t+1}) \tag{40}$$

$$\leq 2\pi_w\sum_{t=0}^{\infty}(P_{w,w}^{2t} - P_{w,w}^{2t+2}) \tag{41}$$

$$\leq 2\pi_w. \tag{42}$$

Note that $P_{w,w}^t \geq P_{w,w}^{t+1}$ holds for lazy P. Thus we obtain the claim from (42). □

4.2 Specific results

4.2.1 Lazy chain on regular graphs

Let $G = (V, E)$ be an undirected and connected graph. Additionally, we assume G is d-regular graph. Then, we consider the following transition matrix

$$(P_{\mathrm{L}})_{v,u} = \begin{cases} \frac{1}{2d} & (\text{if } (v,u) \in E) \\ \frac{1}{2} & (\text{if } v = u) \\ 0 & (\text{othrwise}) \end{cases}. \tag{43}$$

Then, the following corollary is obtained from Theorem 1.2 and Theorem 1.3.

Corollary 4.3. *Suppose that $G = (V, E)$ is an arbitrary d-regular graph and the transition matrix is P_{L}. Then, for each $T \in \mathbb{Z}_{\geq 0}$, it holds that*

$$\mathbf{Pr}\left[\max_{w\in V}|\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P_{\mathrm{L}}^T)_w| \leq 8\sqrt{d\ln n}\right] \geq 1 - \frac{2}{n}.$$

Combining Corollary 4.3 and Proposition A.2, we obtain Corollary 1.4.

4.2.2 Metropolis chain on general graphs

Now, we consider an arbitrary undirected and connected graph $G = (V, E)$. We do not assume the regularity of G. Let d_v be the degree of $v \in V$, i.e., $d_v = |\{u \in V \mid (v,u) \in E\}|$. Then, we consider the following transition matrix on V

$$(P_{\mathrm{M}})_{v,u} = \begin{cases} \frac{1}{2}\min\left\{\frac{1}{d_v}, \frac{1}{d_u}\right\} & (\text{if } (v,u) \in E) \\ 1 - \sum_{u:(v,u)\in E}(P_{\mathrm{M}})_{v,u} & (\text{if } v = u) \\ 0 & (\text{othrwise}) \end{cases}. \tag{44}$$

(P_{M}) is known as the Metropolis chain [9]. Let $d_{\max} = \max_{v\in V} d_v$. Then, the following corollary is obtained from Theorem 1.2 and Theorem 1.3.

Corollary 4.4. *Suppose that $G = (V, E)$ is an arbitrary graph and the transition matrix is P_{M}. Then, for each $T \in \mathbb{Z}_{\geq 0}$, it holds that*

$$\mathbf{Pr}\left[\max_{w\in V}|\mathcal{X}_w^{(T)} - (\mathcal{X}^{(0)}P_{\mathrm{M}}^T)_w| \leq 8\sqrt{d_{\max}\ln n}\right] \geq 1 - \frac{2}{n}.$$

Combining Corollary 4.4 and Proposition A.2, we obtain Corollary 1.5.

Acknowledgements

This work is supported by JSPS KAKENHI Grant Number 17H07116.

References

[1] H. Akbari and P. Berenbrink, Parallel rotor walks on finite graphs and applications in discrete load balancing, Proc. SPAA 2013, 186–195.

[2] H. Akbari, P. Berenbrink and T. Sauerwald, A simple approach for adapting continuous load balancing processes to discrete settings, Distributed Computing, **29**(2) (2016), 143–161.

[3] P. Berenbrink, C. Cooper, T. Friedetzky, T. Friedrich, T. Sauerwald, Randomized diffusion for indivisible loads, Journal of Computer and System Sciences **81**(1) (2015), 159–185.

[4] P. Berenbrink, R. Klasing, A. Kosowski, F. Mallmann-Trenn, and P. Uznanski, Improved analysis of deterministic load-balancing schemes, Proc. PODC 2015, 301–310.

[5] T. Friedrich, M. Gairing, and T. Sauerwald, Quasirandom load balancing, SIAM Journal on Computing, **41** (2012), 747–771.

[6] T. Friedrich and T. Sauerwald, Near-perfect load balancing by randomized rounding, Proc. STOC 2009, 121–130.

[7] D. A. Levin and Y. Peres, Markov Chain and Mixing Times: Second Edition, The American Mathematical Society, 2017.

[8] M. Mitzenmacher and E. Upfal, Probability and Computing Randomization and Probabilistic Techniques in Algorithms and Data Analysis 2nd edition, Cambridge University Press, 2017.

[9] Y. Nonaka, H. Ono, K. Sadakane, M. Yamashita, The hitting and cover times of Metropolis walks, Theoretical Computer Science, **411** (2010), 1889–1894.

[10] Y. Rabani, A. Sinclair, and R. Wanka, Local divergence of Markov chains and analysis of iterative load balancing schemes, Proc. FOCS 1998, 694–705.

[11] T. Sauerwald and H. Sun, Tight bounds for randomized load balancing on arbitrary network topologies, Proc. FOCS 2012, 341–350.

[12] T. Shiraga, Y. Yamauchi, S. Kijima, and M. Yamashita, Deterministic random walks for rapidly mixing chains, arXiv:1311.3749.

A APPENDIX

A.1 Preliminaries of Markov chains

Proposition A.1. *Suppose that P is reversible. Then, for any $\mathcal{X}^{(0)}$, $w \in V$ and*

$$T \geq \left\lceil \frac{1}{1-\lambda} \log\left(\frac{M}{\pi_{\min}\varepsilon}\right)\right\rceil,$$

it holds that

$$\left|(\mathcal{X}^{(0)}P^T)_w - M\pi_w\right| \leq \varepsilon.$$

Proof. We have

$$\left|(\mathcal{X}^{(0)}P^T)_w - M\pi_w\right| \;=\; \left|\sum_{v\in V}\mathcal{X}_v^{(0)}(P^T_{v,w}-\pi_w)\right| \;\le\; M\max_{v\in V}|P^T_{v,w}-\pi_w|.$$

Then, from Theorem 12.4 in [7], we obtain the claim. □

Proposition A.2. *Suppose that P is symmetric. Then, for any $\mathcal{X}^{(0)}$, $w \in V$ and*

$$T \ge \frac{1}{1-\lambda}\log\left(\frac{2Kn}{\varepsilon}\right),$$

it holds that

$$\left|(\mathcal{X}^{(0)}P^T)_w - \frac{M}{n}\right| \le \varepsilon,$$

where $K := \max_{x,y\in V}|\mathcal{X}_x^{(0)} - \mathcal{X}_y^{(0)}|$.

Proof. We have

$$\begin{aligned}
\left|(\mathcal{X}^{(0)}P^T)_w - M/n\right| &= \left|\sum_{v\in V}\mathcal{X}_v^{(0)}(P^T_{v,w}-1/n)\right| \\
&= \left|\sum_{v\in V}\mathcal{X}_v^{(0)}(P^T_{w,v}-1/n)\right| \\
&= \left|\sum_{v\in V}(\mathcal{X}_v^{(0)}-\mathcal{X}_x^{(0)})(P^T_{w,v}-1/n)\right| \\
&\le 2K\cdot\frac{1}{2}\sum_{v\in V}|P^T_{w,v}-1/n|.
\end{aligned}$$

Then, from Theorem 12.4 in [7], we obtain the claim. □

A.2 Concentration inequality

Theorem A.3 (Asuma-Hoeffding Inequality, [8])**.** *Let $X_0,\dots,X_n$ be a martingale such that*

$$|X_k - X_{k-1}| \le c_k.$$

Then, for all $t \ge 1$ and any $\lambda > 0$,

$$\mathbf{Pr}\left[|X_t - X_0| \ge \lambda\right] \le 2\exp\left[-\frac{\lambda^2}{2\sum_{k=1}^{t}(c_k)^2}\right].$$