Fast and Slim Lifted Markov Chains

Kyomin Jung
Mathematics, MIT
kmjung@mit.edu

Devavrat Shah
EECS, MIT
devavrat@mit.edu

Jinwoo Shin
Mathematics, MIT
jinwoos@mit.edu

Abstract

Metropolis-Hasting method allows for designing a reversible Markov chain P on a given graph G for a target stationary distribution π. Such a Markov chain may suffer from its slow mixing time due to reversibility. Diaconis, Holmes and Neal (1997) for the ring-like chain P, and later Chen, Lovasz and Pak (2002) for an arbitrary chain P provided an explicit construction of a non-reversible Markov chain via lifting P so that its mixing time is essentially $1/\Phi(P)$, where$\Phi(P)$ is the conductance of P. Hence it leads to the reduction of the mixing time up to its square root. However, the construction of Chen et. al. results in expansion of size of the underlying graph during lifting; it affects the performance of algorithms (such as samplings or distributed linear dynamics) based on Markov chains. Therefore, motivated to reduce the size of the lifted chain, we provide a new lifting for an arbitrary chain based on expander graphs, which results in reduction of size up to $\Omega(n)$ (n is the number of states of P) with essentially the same mixing time.

The lifting, though allows for reduction in the mixing time, cannot be reduced beyond $1/\Phi(P)$, given a starting Markov chain P. Therefore, if P is ill-designed to begin with, $1/\Phi(P)$ can be large. To overcome this limitation, we introduce a new notion of lifting, which we call *pseudo-lifting*. For an arbitrary chain P and a diameter D of its underlying graph, we provide a simple pseudo-lifting with mixing time $O(D)$ and size $O(nD)$ – this is the fastest possible mixing time for any Markov chain up to constant for any given underlying graph. Next, we provide a *hierarchical pseudo-lifting* which allows for optimizing the size of the lifted chain using geometry of its underlying graph. We exhibit the strength of this construction for graphs with constant doubling dimension ρ to design a fast mixing pseudo-lifted Markov chain with size $O\left(Dn^{1-\frac{1}{\rho+1}}\right)$, which might be much larger without its geometry.

I. INTRODUCTION

Markov chain (or a random walk) is one of the fundamental tools to design algorithms for network problems such as sampling and distributed averaging. Sampling with a given distribution on nodes is used for counting, optimization, approximation etc. Distributed averaging is computing the average of some initial values given at the nodes only by communicating between neighbors. It is closely related to the distributed linear estimation ([15], [4], [16]), the distributed spectral decomposition ([10]) or the distributed load balancing ([11]). In these algorithms, the basic idea is exchanging information based on the Markov chain defined on the network (graph). In both sampling and distributed averaging, the mixing time of the Markov chain (time to approach stationarity from a worst initial state) is crucial for their running time, hence finding the fastest mixing Markov chain for a given network is a major issue in these problems.

Markov chain is classified into two types: reversible(symmetric) and non-reversible(non-symmetric). In 2004, Boyd, Diaconis and Xiao [3] showed that the fastest reversible one for a given graph can be found using the Semi-definite Program, and in 2006 they provided a distributed procedure to find it. But, there is no known method for finding fast non-reversible Markov chains, and the following question arises naturally: how much can we reduce the mixing time by considering non-reversible Markov chains?

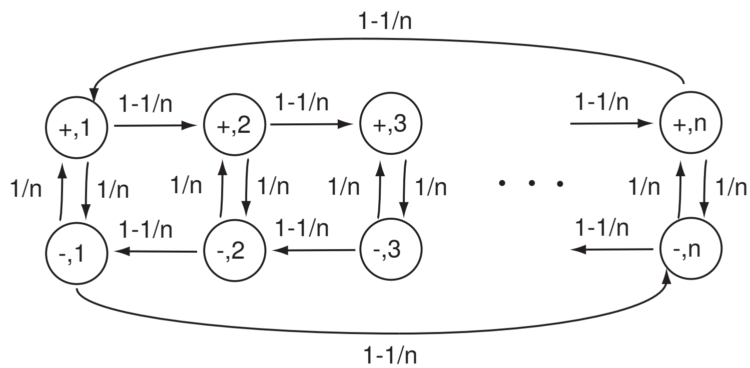

Fig. 1. A non-reversible random walk on the ring-like graph of $2n$ nodes.

Consider a non-reversible Markov chain in Figure I. It mixes in $O(n \log n)$ time, but any reversible Markov chain defined on this graph mixes in $\Omega(n^2)$ time. Note that this graph has a similar structure to the n-ring graph (i.e. the graph obtained by merging $(+, i)$ and $(-, i)$ forms the ring graph), in which any reversible Markov chain also mixes in $\Omega(n^2)$ time [14]. Diaconis, Holmes and Neal [6] first observed this example, i.e. certain reversible chains have closely-related non-reversible Markov chains (called *lifted* chains) that mix substantially faster. Lifting a Markov chain involves possibly splitting each state into multiple ones and assigning a transition probability between new states so that the projection of the new Markov chain (via mapping new states onto their original copies) results in the original Markov chain (see Section II-C for the formal definition of lifting). Therefore, the underlying graph (or the stationary distribution) of the original chain is a projected image of the underlying graph (or the stationary distribution) of the lifted one. In a subsequent work, Chen, Lovasz and Pak [5] provided an explicit construction to lift a general Markov chain with almost optimal speed up in the mixing time.

The sampling algorithm can be implemented via lifting using *memory*: for simulating the lifted chain, the current state stores what the sub-state is in the lifted chain. Therefore, it can sample much faster than algorithms using the original Markov chain P. However, Chen et al. [5] proved the lifted chain cannot mixes faster than $1/\Phi(P)$, where $\Phi(P)$ is the conductance of the original chain P. This motivates our first question:

Q1. Is it possible to overcome the limitation $1/\Phi(P)$ of the running time of the sampling algorithm using *memory*?

Also, the distributed averaging algorithm can be implemented via lifting. We consider the linear distributed iterations [16] and the randomized gossip algorithm [4], in which using a faster mixing Markov chain leads to a better running time. For a given Makkov chain P on the network, the linear distributed iterations updates the current value $x_i(t)$ of node i at time t by $\sum_j P_{ji} x_j(t-1)$, which is the weighted average of the values of i's neighbors at time $(t-1)$. A lifted Markov chain can be simulated in the linear distributed iterations by running

	CLP for general graphs	*L1* for general graphs	*L2* for general graphs	*L3* for doubling graphs
Mixing Time	$O(C)$	$O(C \log \|V\|)$	$O\left(\frac{D}{\varepsilon}\right)$	$O\left(\frac{D}{\varepsilon}\right)$
Size	$\Omega(C\|V\|^2)$	$O(C\|E\|)$	$O(\|E\| + \|V\|D)$	$O(\|V\|^{\frac{\rho}{\rho+1}} D)$

TABLE I
COMPARISON OF OUR CONSTRUCTIONS WITH THAT OF CHEN ET. AL. [5]. V, E, D, AND ρ REPRESENT THE VERTEX SET, THE EDGE SET, THE DIAMETER, AND THE DOUBLING DIMENSION OF THE UNDERLYING GRAPH RESPECTIVELY. HERE $D \leq C \approx 1/\Phi(P)$, AND ε IS THE PARAMETER OF PSEUDO-LIFTING DEFINED IN SECTION IV.

		Original	*CLP*	*L1*	*L3*
Ring graph	Time	$\Omega(n^2)$	$O(n \log n)$	$O(n \log^2 n)$	$O\left(\frac{n}{\epsilon}\right)$
	Size	$\Theta(n)$	$\Omega(n^3)$	$O(n^2 \log n)$	$O\left(n^{1.5}\right)$
k-grid	Time	$\Omega\left(n^{\frac{2}{k}}\right)$	$O\left(n^{\frac{1}{k}} \log n\right)$	$O\left(n^{\frac{1}{k}} \log^2 n\right)$	$O\left(\frac{n^{\frac{1}{k}}}{\epsilon}\right)$
	Size	$\Theta(n)$	$\Omega\left(n^{2+\frac{1}{k}}\right)$	$O\left(n^{1+\frac{1}{k}} \log n\right)$	$O\left(n^{1+\frac{1}{k(k+1)}}\right)$
Barbell graph	Time	$\Omega\left(n^2\right)$	$O\left(n^2 \log n\right)$	$O\left(n^2 \log^2 n\right)$	$O\left(\frac{1}{\epsilon}\right)$
	Size	$\Theta(n^2)$	$\Omega\left(n^4\right)$	$O\left(n^4 \log n\right)$	$O\left(n^2\right)$

TABLE II
COMPARISON FOR SOME SPECIFIC GRAPHS. HERE 'ORIGINAL' MEANS THE UNIFORM REVERSIBLE RANDOM WALK ON THE GRAPH, AND OTHER BOUNDS ARE ACHIEVED FROM LIFTINGS USING '*CLP*', '*L1*', '*L2*', AND '*L3*' RESPECTIVELY. A BARBELL GRAPH WITH $2n$ NODES IS A GRAPH THAT HAS TWO K_n CONNECTED BY A SINGLE EDGE.

the multiple *threads* in each node. The total number of iterations is the mixing time T of the used Markov chain P, hence it would decrease after lifting. On the other hand, since $|E|$ exchanging happens per each iteration, the total number of operations is $T \times |E|$, where E is the edge set of the underlying graph of P. Therefore, it may not decrease because the size $|E|$ may increase after lifting. In other words, the size $|E|$ of the lifted chain is involved in the cost of this algorithm. This motivates our second question:

Q2. Can we construct a lifted chain which mixes as fast as that of [5], but has smaller size?

Now, consider the randomized gossip algorithm, which is a distributed averaging algorithm on a given network with a gossip constraint, i.e. means that no node communicates with more than one neighbor in every time slot. Unlike the linear distributed iterations, the lifted chains cannot be used in [4] because it requires the reversible one (Chen et al. [5] proved the fastest lifted chain should be non-reversible).

A. *Our contributions*

For the second question, we construct a lifted chain based on the *expander* graph, instead of the *complete* graph in [5], to reduce the size of the lifting. In this paper, the size of the Markov chain P indicates the number of edges in the underlying graph of P. In particular, our construction, which we call *L1*, lead to the reduction in the size of the lifted chain up to $\Theta(n)$ with respect to the construction in [5], where n is the number of states in P. Our construction builds on the technique of [5], but the usage of the expander graph in place of the complete graph requires new proof techniques for establishing the bound on the mixing time. First, we need to consider a different flow problem, and the flow shortening technique in [12]. Second, we analyze the mixing time of the hybrid-type random walk.

For the first question, we introduce the relaxed notion of lifting, which we call *pseudo-lifting*. We relaxed some properties of the *original lifting*, but it still preserves the topological properties as well as allows for the exact simulation of the original stationary distribution (see Section IV for the formal definition of pseudo-lifting). For an arbitrary Markov chain P, we obtain its pseudo-lifted chain which mixes in $O(D)$ time and has size $O(Dn + |E|)$, where D is the diameter of the underlying graph of P. (Somewhat surprisingly, the mixing time depends only on D, not entire information of P, hence this gives the rapidly mixing Markov chain even when P has a large $1/\Phi(P)$ value.) As you see in the discussion of the Barbell graph in Section IV-B, for any P defined on a given graph, D

can be much smaller than $1/\Phi(P)$ that is the limitation of the mixing time of the original lifted chain. Furthermore, the size guarantee $O(Dn + |E|)$ is even smaller than the size of *L1*. We will denote this construction as *L2*.

Finally, we provide a hierarchical pseudo-lifting construction, which we will denote as *L3*. It allows for optimizing the size of the pseudo-lifted chain by considering the trade-off with the mixing time. Specifically, we apply this construction for a class of Markov chains with geometry: its underlying graphs has a constant doubling dimension ρ. We show that a two-level hierarchical pseudo-lifted chain mixes in $O(D)$ time and has size $O(Dn^{1-\frac{1}{1+\rho}})$: it achieves the reduction up to $\Theta(n^{\frac{1}{1+\rho}})$ factor compared to *L2*.

We summarize the performance of our various constructions *L1*, *L2* and *L3* as discussed above in Tables I and II. We do head-to-head comparisons with the construction of Chen et al. (denoted by *CLP*). Key features of *L3* are: the order optimality of the mixing time as well as the small size of the lifted chain – specifically in the case of a Markov chain with geometry such as a random walk on the k-grid graph.

B. *Organization*

In Section II, the preliminaries and notations are summarized. In Section III, we show how to build *L1*, the fast mixing lifted chain with a smaller size than that of *CLP*. *L1* has the mixing time $O\left(C \log 1/\pi_0\right)$ and size $O(C|E|)$, while those of *CLP* are $\Theta(C)$ and $\Omega(C|V|^2)$ respectively. In Section IV, we define new concept of lifting, ε *pseudo-lifting*, and describe its construction *L2*. We prove that *L2* has the mixing time $O(D\varepsilon^{-1})$ and size $O(|E| + D|V|)$. In Section V, we use the geometry of graph and obtain the better ε pseudo-lifting construction *L3* in terms of the size of the lifted chain. *L3* has the mixing time $O(D\varepsilon^{-1})$ and size $O\left(D|V|^{1-\frac{1}{1+\rho}}\right)$.

II. Notations and Preliminaries

A. *Mixing Times of Markov Chains*

Consider a finite irreducible ergodic Markov chain, with the underlying graph $G = (V, E)$, the transition matrix P, the stationary distribution π. We may use $V(G)$ to represent vertices of V of G. P is graph comformant, that is, $P_{ij} > 0$ only if $(i, j) \in E$. The ergodic flow matrix $Q = [Q_{ij}]$, where $Q_{ij} = \pi_i P_{ij}$. It satisfies: $\sum_{i,j} Q_{ij} = 1$, $\sum_i Q_{ij} = \sum_i Q_{ji}$ and $\sum_i Q_{ij} = \pi_j$. Furthermore, every non-negative matrix Q with these properties defines a Markov chain with the stationary distribution π. The reverse chain P^* of P is defined as: $P^*_{ij} = \pi_j P_{ji}/\pi_i$ for all $(i, j) \in E$. We call P *reversible* if $P = P^*$. We will use definitions $\pi_0 = \min_i \pi_i$ and $\pi_{max} = \max_i \pi_i$.

Let $P^t(x, \cdot)$ denote the distribution of Markov chain after t steps starting from $x \in V$. For the Markov chain of our interest, $P^t(x, \cdot)$ goes to π as $t \to \infty$. The notion of *Mixing time* is the measure of rate of convergence to π. There are various (mostly equivalent) definitions of Mixing time that are considered in literature based on different measures of distance in distribution, stopping time, etc. First, we present the definition based on the total variation distance and the χ^2-distance.

Definition 1: Given $\varepsilon > 0$, let $\tau(\varepsilon)$ and $\tau_2(\varepsilon)$ represent ε-Mixing time of the Markov chain with respect to the total variation distance and the χ^2-distance respectively. Then, they are

$$\tau(\varepsilon) = \min\left\{t : \forall x \in \Omega, \frac{1}{2}\sum_{y\in\Omega} \left|P^t(x, y) - \pi(y)\right| \le \varepsilon\right\},$$

$$\tau_2(\varepsilon) = \min\left\{t : \forall x \in \Omega, \sqrt{\sum_{y\in\Omega} \frac{1}{\pi(y)} \left(P^t(x, y) - \pi(y)\right)^2} \le \varepsilon\right\}.$$

However, in this paper, we primarily consider the definition of Mixing time from the stopping rule. A stopping rule Γ is a stopping time based on the random walk of the Markov chain: at any time, it decides wether to stop or not, depending on the walk seen so far and possibly additional coin flips. Suppose, the starting node w^0 is drawn from distribution σ. The distribution of the stopping node w^Γ is denoted by $\sigma^\Gamma = \tau$ and call Γ as stopping rule from σ to τ. Let $\mathcal{H}(\sigma, \tau)$ be the infimum of mean length over all such stopping rules from σ to τ. This is well-defined as there exists the following stopping rule from σ to τ: *select i with probability τ_i and walk until getting to i*. We will denote by $\mathcal{H}(i, j)$, the average number of steps taken to reach j starting from i under the random walk of Markov chain. Similarly, the access time $\mathcal{H}(\pi, S)$ is the expected number of steps before the set S is reached with starting distribution π. Now, we present the definition of the (stopping rule based) Mixing time $\mathcal{H}$.

Definition 2: $\mathcal{H} = \max_\sigma \mathcal{H}(\sigma, \pi)$.

Therefore, to bound $\mathcal{H}$, we need to design a stopping rule whose distribution of stopping nodes is π. Sometimes, due to the difficulty for design such an exact stopping rule, we use the following strategy for bounding the mixing time $\mathcal{H}$.

Step 1. For a positive constant ε and any starting distribution σ, we design a stopping rule whose stopping distribution γ is ε-far from π (i.e. $\gamma \geq (1-\varepsilon)\pi$). This gives the upper bound for $H(\sigma, \gamma)$.

Step 2. We bound $\mathcal{H}$ by $H(\sigma, \gamma)$ using the following fact known as *fill-up Lemma* in [1]:

$$\mathcal{H} \leq \frac{1}{1-\varepsilon} \mathcal{H}_{\underline{\varepsilon}},$$

where $\mathcal{H}_{\underline{\varepsilon}} = \max_\sigma \min_{\gamma \geq (1-\varepsilon)\pi} \mathcal{H}(\sigma, \gamma)$.

B. Additional Techniques to bound Mixing Times

Various techniques have been developed over past three decades or so to estimate Mixing time of a given Markov chain. We review some of the key techniques that will be relevant for this paper.

Conductance. The conductance of a Markov chain with ergodic flow matrix Q is defined by

$$\Phi = \min_{S \subset V} \frac{Q(S, V \backslash S)}{\pi(S)\pi(V \backslash S)},$$

where $Q(A, B) = \sum_{i \in A, j \in B} Q_{ij}$. The Mixing time $\mathcal{H}$ and conductance Φ are related as follows:

$$\frac{1}{\Phi} \leq \mathcal{H} \leq O\left(\frac{1}{\Phi^2} \log \frac{1}{\pi_0}\right).$$

Eigenvalue. If P is reversible, one can view P as a self-adjoint operator on a suitable inner product space and this permits us to use the well-understood spectral theory of self-adjoint operators. It is well-known that P has $n = |V|$ real eigenvalues $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} > -1$. The ε-mixing time $\tau_2(\varepsilon)$ is related as

$$\tau_2(\varepsilon) \leq \left\lceil \frac{1}{\lambda_P} \log \frac{1}{\varepsilon\sqrt{\pi_0}} \right\rceil,$$

where $\lambda_P = 1 - \max\{|\lambda_1|, |\lambda_{n-1}|\}$. The λ_P is also called the spectral gap. When, P is non-reversible we consider PP^*. It is easy to see that the Markov chain with PP^* as transition matrix is reversible. Let λ_{PP^*} be the spectral gap of this reversible Markov chain. Then, the mixing time of original Markov chain (with transition matrix P) is bounded above as:

$$\tau_2(\varepsilon) \leq \left\lceil \frac{2}{\lambda_{PP^*}} \log \frac{1}{\varepsilon\sqrt{\pi_0}} \right\rceil.$$

C. Lifting and Flows

The following definition of lifting implies that it is possible to simulate the original distribution π using the lifted Markov chain. The key property to note is that lifting preserves the original topological properties of the graph.

Definition 3 (Lifting): Consider a Markov chain with transition matrix P and stationary distribution π defined on graph $G = (V, E)$. A Markov chain with transition matrix $\widehat{P}$, stationary distribution $\widehat{\pi}$ on graph $\widehat{G} = (\widehat{V}, \widehat{E})$ is called lifting of P if there exists a many-to-one function $f : \widehat{V} \to V$ such that the following holds: (a) for any $\widehat{u}, \widehat{v} \in \widehat{V}$, $(\widehat{u}, \widehat{v}) \in \widehat{E}$ only if $(f(\widehat{u}), f(\widehat{v})) \in E$; (b) for any $u, v \in V$, $\pi(u) = \widehat{\pi}(f^{-1}(u))$, and $Q(u, v) = \widehat{Q}(f^{-1}(u), f^{-1}(v))$. Here Q and $\widehat{Q}$ are ergodic flow matrices for P and $\widehat{P}$ respectively.

In [5], authors use a multi-commodity flow to construct a specific lifting of a given Markov chain P to speed up its mixing time. Specifically, they consider a multi-commodity flow problem on G with the capacity constraint on edge $(u, v) \in E$ given by Q_{uv}. A flow from a source s to a destination t, denoted by f, is defined as a non-negative function on edges of G so that

$$\sum_j f(ji) = \sum_j f(ij)$$

for every node $i \neq s, t$. The value of the flow is defined by

$$val(f) = \sum_j f(sj) - \sum_j f(js) = \sum_j f(jt) - \sum_j f(tj)$$

, and the cost of flow f^{st} is defined as

$$cost(f) = \sum_{(i,j) \in E} f(ij).$$

A *multi-commodity flow* is a collection $f = (f^{st})$ of flows, where each f^{st} is a flow from s to t. Define the *congestion* of a multi-commodity flow f as

$$\max_{(i,j) \in E} \frac{\sum_{s,t} f^{st}(ij)}{Q_{ij}}.$$

Consider the following optimization problem, essentially trying to minimize the congestion and the cost simultaneously under the condition for the amount of flows:

$$\begin{aligned} \text{minimize} \quad & K \\ \text{subject to} \quad & val(f^{st}) = \pi_s \pi_t, \quad \forall s, t, \\ & \sum_{s,t} f^{st}(ij) \leq K Q_{ij}, \quad \forall (i,j) \in E, \\ & \sum_t cost(f^{st}) \leq K \pi_s, \ \ \sum_s cost(f^{st}) \leq K \pi_t, \ \forall s, t. \end{aligned}$$

Let C be the optimal solution of the above problem. It is easy to see that $C \geq 1/\Phi$. Further, if P is reversible, then result of Leighton and Rao [13] on the approximate multi-commodity implies that

$$C = O\left(\frac{1}{\Phi} \log \frac{1}{\pi_0}\right).$$

Let the optimal multi-commodity flow of the above problem be F_1, and we can think of F_1 as a weighted collection of directed paths. In [5], authors modified F_1, and got a new multi-commodity flow F_2 that has the same amount of $s-t$ flows as F_1, while its congestion and path length are at most $12C$. They used F_2 to construct a lifting $\widehat{P}$ with mixing time $\widehat{\mathcal{H}}$ such that

$$\widehat{\mathcal{H}} \leq 144C.$$

Also, they showed that the mixing time of any lifting $\widehat{P}$ is greater than $C/2$, hence their lifted Markov chain has almost optimal speed up within a constant factor.

To obtain a lifting with very small size, we will to study the existence of the specific k-commodity flow with short path lengths. For this, we will use a *balanced multi-commodity flow*, which is a multi-commodity flow with the following condition for the amount of flows:

$$val(f^{st}) = g(s,t), \forall s, t,$$

and $g(s,t)$ satisfies the balanced condition:

$$\sum_t g(s,t) \leq \pi_s, \ \ \sum_s g(s,t) \leq \pi_t, \quad \forall s, t.$$

Therefore, F_1 and F_2 are also balanced multi-commodity flows with $g(s,t) = \pi_s \pi_t$. Given a multi-commodity flow f, let $C(f)$ be its congestion and $D(f)$ be the length of the longest flow-path. Then, the *flow number* T is defined follows:

$$T = \min_f \left(\max\left\{C(f), D(f)\right\}\right),$$

where the minimum is taken over all balanced multi-commodity flows with $g(s,t) = \pi_s \pi_t$. Hence, as stated earlier, $T \leq 12C$ (from F_2). The following claim appears in [12]:

Claim 1: (Claim 2.2 in [12]) For any $g(s,t)$ satisfying the balanced condition, there exists a balanced multi-commodity flow f with $g(s,t)$ such that $\max\{C(f), D(f)\} \le 2T$.

D. Expanders

The expander graphs are sparse graphs which have high connectivity properties, quantified using the edge expansion $h(G)$ as defined as

$$h(G) = \min_{1 \le |S| \le \frac{n}{2}} \frac{|\partial(S)|}{|S|},$$

where $\partial(S)$ is the set of edges with exactly one endpoint in S. For constants d and c, a family $\mathcal{G} = \{G_1, G_2, \dots\}$ of d-regular graphs is called a (d,c)-expander family if $h(G) > c$ for every $G \in \mathcal{G}$. There are many explicit constructions of a (d,c)-expander family available in recent times. We will use a (d,c)-expander graph $G^{Ex} = (V, E^{Ex})$ (i.e. $V^{Ex} = V$), and a transition matrix P^{Ex} defined on this graph. For a given π, we can define a reversible P^{Ex} so that its stationary distribution is π as follows,

$$P^{Ex}_{ij} = \begin{cases} \frac{\pi_0}{d\pi_i} & \text{if } (i,j) \in E^{Ex} \\ 1 - \frac{\pi_0}{\pi_i} & \text{if } i = j \end{cases}.$$

In the case of $\pi_{max} = O(\pi_0)$, it is easy to check that $\Phi(P^{Ex}) = \Theta(h(G)) = \Omega(1)$, where $\Phi(P^{Ex})$ is the conductance of P^{Ex}. Hence, $\lambda_{P^{Ex}} = \Omega(1)$, and the random walk defined by P^{Ex} mixes fast. In Section III, we will consider only such π.

E. Doubling Dimension Metric

The notion of doubling dimension metric space (or graphs) was introduced in [2], [8], [7]. Specifically, consider a metric space $\mathcal{M} = (\mathcal{X}, \mathbf{d})$, where $\mathcal{X}$ is the set of point endowed with a metric $\mathbf{d}$. Given $x \in \mathcal{X}$, define a ball of radius $r \in \mathbb{R}_+$ around x as $\mathbf{B}(x,r) = \{y \in \mathcal{X} : \mathbf{d}(x,y) < r\}$. Define

$$\rho(x,r) = \inf\{K \in \mathbb{N} : \exists\ y_1, \dots, y_K \in \mathcal{X}, \mathbf{B}(x,r) \subset \cup_{i=1}^K \mathbf{B}(y_i, r/2)\}.$$

Then, the $\rho(\mathcal{M}) = \sup_{x \in \mathcal{X}, r \in \mathbb{R}_+} \rho(x,r)$ is called the *doubling constant* of $\mathcal{M}$ and $\log_2 \rho(\mathcal{M})$ is called the *doubling dimension* of $\mathcal{M}$. Doubling dimension of a graph $G = (V,E)$ is defined with respect to metric induced on V by the shortest path metric.

III. Efficient Lifting Using Expanders: Construction of *L1*

In what follows, we will consider only lazy P such that $P \ge I/2$. We can assume this without loss of generality for the purpose of mixing time because if P is not such then we can modify it as $(I+P)/2$; the mixing time of $(I+P)/2$ is within constant factor of the mixing time of P. Next, we present our construction.

A. Construction

We use the multi-commodity flow based construction which was introduced in [5]. They essentially use a multi-commodity flow between source-destination pairs for all $s,t \in V$. Instead, we will use a balanced multi-commodity flow between source-destination pairs that are obtained from an expander. Thus, the essential change in our construction is the use of an expander in place of a complete graph used in [5]. A caricature of this lifting is explained in Figure 2. However, this change makes analysis of mixing time lot more challenging and requires us to use different analysis technique. Further, we use arguments based on the classical linear programming to derive the bound on the size of lifting.

To this end, we consider the following multi-commodity flow: let $G^{Ex} = (V, E^{Ex})$ be an expander with a transition matrix P^{Ex} and a stationary distribution π as required – this is feasible since we have assume $\pi_{max} = O(\pi_0)$. We note that this assumption is used only for existence of expander. Consider a multi-commodity flow $f = (f^{st})_{(s,t) \in E^{Ex}}$ so that

(a) $val(f^{st}) = \pi_s P^{Ex}_{st} = Q^{Ex}_{st}$ for all $(s,t) \in E^{Ex}$;
(b) $\sum_{s,t} f^{st}(ij) \le KQ_{ij}$ for all $(i,j) \in E$;

Lemma 2: There is a feasible multi-commodity flow F_3 in the above flow problem with congestion(K) and path-length at most $\widehat{C} = 24C$.

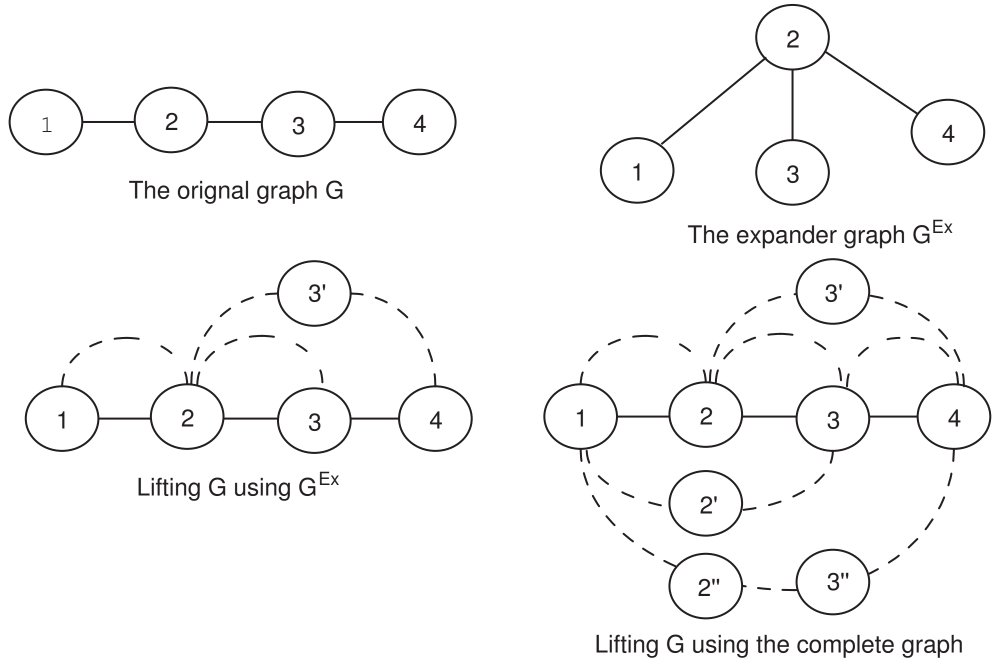

Fig. 2. A caricature of lifting using expander. Let line graph G be a line graph with 4 nodes. We wish to use an expander G^{Ex} with 4 nodes, shown on the top-right side of the figure. G is lifted by adding paths that correspond to edges of expander. For example, an edge $(2,4)$ of expander is added as path $(2,3',4)$. We also draw the lifting in [5] which uses the complete graph.

This Lemma is derived directly from Claim 1, because the flow number is less than $12C$ and this flow is a balanced multi-commodity flow. Now, we can think of F_3 as a weighted collection of directed paths $\{(\mathcal{P}_r, w_r) : 1 \le r \le N\}$, where the total weight of paths from node s to t is $\pi_s P^{Ex}_{st}$, where $(s,t) \in E^{Ex}$. Let ℓ_r be the length of path $\mathcal{P}_r$. From Lemma 2, we have the following:

$$\sum_r w_r = 1, \qquad \ell_r \le \widehat{C}, \tag{1}$$

$$\sum_{r:\mathcal{P}_r \text{ starts at } i} w_r = \pi_i, \quad \sum_{r:\mathcal{P}_r \text{ ends at } i} w_r = \pi_i, \text{ for } i \in V \tag{2}$$

$$\sum_{r:(i,j)\in E(\mathcal{P}_r)} w_r \le \widehat{C} Q_{ij}, \text{ for } (i,j) \in E. \tag{3}$$

Using such a collection of weighted paths, we construct the desired lifting next. As Figure 2, we construct the lifted graph $\widehat{G} = (\widehat{V}, \widehat{E})$ from G by adding a directed path $\mathcal{P}'_r$ of length ℓ_r connecting i to j if $\mathcal{P}_r$ goes from i to j. Subsequently, $\ell_r - 1$ new nodes are added to the original graph. The ergodic flow on an edge (i,j) of the lifted chain is defined by

$$\widehat{Q}_{ij} = \begin{cases} w_r/2\widehat{C} & \text{if } (i,j) \in E(\mathcal{P}'_r) \\ Q_{ij} - \sum_{r:ij\in E(\mathcal{P}_r)} w_r/2\widehat{C} & \text{if } (i,j) \in E(G) \end{cases}$$

It is easy to check it defines a Markov chain on $\widehat{G}$, and a natural way of mapping the paths $\mathcal{P}'_r$ onto the paths $\mathcal{P}_r$ collapses the random walk on $\widehat{G}$ onto the random walk on G. The stationary distribution of the lifted chain is

$$\widehat{\pi}_i = \begin{cases} w_r/2\widehat{C} & \text{if } i \in V(\mathcal{P}'_r) \backslash V(G) \\ \pi_i - \sum_{r:\mathcal{P}_r \text{ thru } i} w_r/2\widehat{C} & \text{if } i \in V(G) \end{cases}$$

Thus, the above stated construction is a valid lifting of given Markov chain P defined on G.

B. Mixing Time of the Lifted Chain

First, we state two useful Lemmas.

Lemma 3: Let P_1, P_2 be reversible Markov chains with their stationary distributions π_1, π_2 respectively. Suppose there exist positive constants α, β, c, d such that $P_1 \geq \alpha P_2$, $P_1 \geq \beta I$, and $c\pi_2 \leq \pi_1 \leq d\pi_2$. Then,

$$\lambda_{P_1} \geq \min\left(\frac{\alpha c}{d^2}\lambda_{P_2}, 2\beta\right).$$

Proof: From the min-max characterization of the spectral gap (see, e.g., page 176 in [9]) for the reversible Markov chain, it follows that

$$\begin{aligned}\lambda_{P_1} = & \inf_{\psi:V\to\mathbb{R}}\left(\frac{\sum_{i,j\in V}(\psi(i)-\psi(j))^2(\pi_1)_i(P_1)_{ij}}{\sum_{i,j\in V}(\psi(i)-\psi(j))^2(\pi_1)_i(\pi_1)_j}\right)\\ \geq & \left(\frac{\alpha c}{d^2}\right)\inf_{\psi:V\to\mathbb{R}}\left(\frac{\sum_{i,j\in V}(\psi(i)-\psi(j))^2(\pi_2)_i(P_2)_{ij}}{\sum_{i,j\in V}(\psi(i)-\psi(j))^2(\pi_2)_i(\pi_2)_j}\right)\\ = & \left(\frac{\alpha c}{d^2}\right)\lambda_{P_2}.\end{aligned}$$

The smallest eigenvalue of P_1 is greater than $2\beta - 1$ because $P_1 \geq \beta I$. So, the distance between the smallest eigenvalue and -1 is greater than 2β. This completes the proof of Lemma 3. ■

Lemma 4: Let P_1, P_2 be Markov chains with their stationary distributions π_1, π_2 respectively. Now, suppose P_2 is reversible as well. (P_1 is not necessarily reversible.) If there exist positive constants α, β, c, d such that $P_1 \geq \alpha P_2$, $P_1 \geq \beta I$ and $c\pi_2 \leq \pi_1 \leq d\pi_2$. Then,

$$\lambda_{P_1 P_1^*} \geq \min\left(\frac{\alpha\beta c}{d^2}\lambda_{P_2}, 2\beta^2\right).$$

Proof: $P_1 P_1^*$ is a reversible Markov chain which has π_1 as its stationary distribution. Because $P_1^* \geq \beta I$, $P_1 P_1^* \geq \alpha P_2 P_1^* \geq \alpha\beta P_2$. Also, $P_1 P_1^* \geq \beta^2 I$. Now, the proof of Lemma 4 follows from Lemma 3. ■

Now, we state the main result stating the bound on mixing time of the lifted Markov chain.

Theorem 5: The mixing time, $\widehat{\mathcal{H}}$ of the lifted Markov chain represented by $\widehat{Q}$ defined on $\widehat{G}$ is bounded above as $\widehat{\mathcal{H}} = O(C\log(1/\pi_0))$.

Proof: By property of expander, we have $\lambda_{P^{Ex}} = \Omega(1)$. Therefore, to prove Theorem 5, it is sufficient to show that

$$\widehat{\mathcal{H}} = O\left(\frac{\widehat{C}}{\lambda_{P^{Ex}}}\log(1/\pi_0)\right).$$

First, note that for any node $i \in V$ (i.e. original node i in graph G),

$$\frac{1}{2}\pi_i \leq \widehat{\pi}_i \leq \pi_i. \tag{4}$$

Now, under the lifted Markov chain the probability of getting on any directed path $\mathcal{P}'_r$ starting at i is

$$\widehat{P}_{ij} = \frac{\widehat{Q}_{ij}}{\widehat{\pi}_i} = \frac{w_r}{2\widehat{C}\widehat{\pi}_i}.$$

Hence the probability of getting on any directed path starting at i is

$$\sum_{r:\mathcal{P}'_r \text{ starts at } i}\frac{w_r}{2\widehat{C}\widehat{\pi}_i} = \frac{1}{2\widehat{C}\widehat{\pi}_i}\sum_{r:\mathcal{P}'_r \text{ starts at } i} w_r = \frac{\pi_i}{2\widehat{C}\widehat{\pi}_i}.$$

From (4), this is bounded between $\frac{1}{2\widehat{C}}$, and $\frac{1}{\widehat{C}}$.

To study the $\widehat{\mathcal{H}}$, we will focus on the induced random walk (or Markov chain) on original nodes $V \subset \widehat{V}$ by the lifted Markov chain $\widehat{P}$. Let $\widehat{P}^V$ be the transition matrix of this induced random walk. Then,

$$\widehat{P}^V_{ij} = \widehat{P}_{ij} + \sum_{r:\mathcal{P}'_r \text{ goes from } i \text{ to } j}\frac{w_r}{2\widehat{C}\widehat{\pi}_i}.$$

Now, $\widehat{P}^V \geq \widehat{P} \geq I/4$, because $\widehat{P}_{ii} = \widehat{Q}_{ii}/\widehat{\pi}_i \geq Q_{ii}/2\widehat{\pi}_i = P_{ii}\pi_i/2\widehat{\pi}_i \geq P_{ii}/2 \geq I/4$. Here we have assumed that $P \geq I/2$ as discussed earlier. Now,

$$\widehat{P}^V_{ij} \geq \frac{1}{2\widehat{C}\widehat{\pi}_i} \sum_{r:\mathcal{P}'_r \text{ goes from } i \text{ to } j} w_r = \frac{\pi_i P^{Ex}_{ij}}{2\widehat{C}\widehat{\pi}_i} \geq \frac{1}{2\widehat{C}} P^{Ex}_{ij}.$$

And, its stationary distribution $\widehat{\pi}^V$ is : $\widehat{\pi}^V_i = \frac{\widehat{\pi}_i}{\widehat{\pi}(V)}$. Therefore, by (4) we have $\frac{1}{2}\pi_i \leq \widehat{\pi}^V_i \leq 2\pi_i$. Now, we can apply Lemma 4 to obtain the following:

$$\lambda_{\widehat{P}^V(\widehat{P}^V)^*} = \Omega\left(\frac{1}{\widehat{C}}\lambda_{P^{Ex}}\right). \tag{5}$$

Now, we are ready to design a stopping rule Γ, that will imply that the desired bound on $\widehat{\mathcal{H}}$ as claimed in Theorem 5. Now, the stopping rule:

(i) Walk until visiting old nodes of $V \subset \widehat{V}$ for t times, where $t = \left\lceil 2\log(2/\widehat{\pi}^V_0)/\lambda_{\widehat{P}^V(\widehat{P}^V)^*} \right\rceil$. Let this t-th old node be denoted by X.
(ii) Stop at X with probability $1/2$.
(iii) Otherwise, continue walking until getting onto any directed path $\mathcal{P}'_r$; choose an interior node Y of $\mathcal{P}'_r$ uniformly at random and stop at Y.

Given (5) and results stated in Section II-B on relation between mixing time and $\lambda_{\widehat{P}^V(\widehat{P}^V)^*}$, it follows that after time t as defined above the Markov chain, restricted to old nodes V, has distribution close to $\widehat{\pi}^V$. Therefore,

$$|\Pr(X = w) - \widehat{\pi}^V_w| \leq \widehat{\pi}^V_w/2, \quad \forall\ w \in V.$$

According to the above stopping rule, we stop at an old node w with probability $1/2$. Therefore, for any $w \in V$, we have that the stopping time Γ stops at w with probability at least $\widehat{\pi}^V_w/4 \geq \pi_w/8 \geq \widehat{\pi}_w/8$. With probability $1/2$, the rule does not stop at the node X. Let w^k be the k^{th} point in the walk starting from X. Because at any old node i, the probability of getting on any directed path is between $\frac{1}{2\widehat{C}}$ and $\frac{1}{\widehat{C}}$, a coupling argument shows that for any old node i,

$$\Pr(w^k = i | w^0, \cdot, w^k \text{ are old nodes}) \geq \left(1 - \frac{1}{\widehat{C}}\right)^k \frac{1}{2}\widehat{\pi}^V_i$$

If w is a new point on the directed path $\mathcal{P}'_r$ which connects the old node i to j. Then,

$$\begin{aligned}
\Pr(\Gamma \text{ stop at } w) &\geq \frac{1}{2}\sum_{k=0}^{\infty} \text{Prob}(w^k = i | w^0, \cdot, w^k \text{ are old points}) \times \text{Prob}(\text{at } i\text{, get on the path } \mathcal{P}'_r) \times \frac{1}{\ell_r} \\
&\geq \frac{1}{2}\sum_{k=0}^{\infty}\left(1 - \frac{1}{\widehat{C}}\right)^k \frac{1}{2}\widehat{\pi}^V_i \frac{w_r}{2\widehat{C}\widehat{\pi}_i}\frac{1}{\widehat{C}} \\
&\geq \frac{w_r}{16\widehat{C}^2}\sum_{k=0}^{\infty}\left(1 - \frac{1}{\widehat{C}}\right)^k \\
&= \frac{w_r}{16\widehat{C}} \\
&= \frac{1}{8}\widehat{\pi}_w
\end{aligned}$$

The average length of this stopping rule is $O(t + \widehat{C})$. By (5),

$$O(t + C) = O\left(\left\lceil \frac{2}{\lambda_{\widehat{P}^V(\widehat{P}^V)^*}}\log(2/\pi_0)\right\rceil + \widehat{C}\right) = O\left(\frac{\widehat{C}}{\lambda_{P^{Ex}}}\log(1/\pi_0)\right).$$

Thus, we have established that the stopping rule Γ has average length $O(C\log 1/\pi_0)$ and the distribution of the stopping nodes is $\Omega(\widehat{\pi})$. Therefore, using the *fill-up* lemma stated in the Section II-A, it follows that $\widehat{\mathcal{H}} = O(C\log 1/\pi_0)$. ■

C. Size of the Lifted Chain

Here, we establish the bound on the size of lifted chain as described above. We want to establish that the size of the lifted chain in terms of number of edges, i.e. $|\widehat{E}| = O(C|E|)$. Note that, the lifted graph $\widehat{G}$ is obtained by adding the paths that appeared in the solution of the multi-commodity flow problem. Therefore, to establish the desired bound we need to establish bound on the number of distinct paths as well as their lengths.

To this end, let re-formulate the multi-commodity flow based on expander G^{Ex} as follows. For each $(s,t) \in E^{Ex}$, we add a flow between s and t. Let this flow be routed along possibly multiple paths. Let P_{stj} denote the j^{th} path from s to t, it's length ℓ_{stj} is at most $\widehat{C}$ as per discussion in Lemma 2 and x_{stj} be the amount of flow sent along this path. Let the overall solution, denoted by $\{(\mathcal{P}_r, w_r)\}$ gives a feasible solution in the following polytope with x_{stj} as its variables:

$$\sum_j x_{stj} = \pi_s P^{Ex}_{st}, \quad \forall (s,t) \in E^{Ex}$$
$$\sum_{st \in E^{Ex}} \sum_{j: e \in \mathcal{P}_{stj}} x_{stj} \leq \widehat{C} Q_e, \quad \forall e \in E$$
$$x_{stj} \geq 0 \qquad \forall s,t,j.$$

Clearly, any feasible solution in this polytope, say $\{(\mathcal{P}_r, w_r)\}$, will work for our lifting construction. Now, the size of its support set is $|\{(\mathcal{P}_r, w_r)\}|$. If we consider the extreme point of this polytope, the size of its support set is at most $|E^{Ex}| + |E| = O(|E|)$ because the extreme point is an unique solution of sub-collection of linear constraints in this polytope. Hence, if we choose such an extreme point $\{(\mathcal{P}_r, w_r)\}$ for our lifting, the size of our lifted chain $|\widehat{E}|$ is at most $O(C|E|)$ since each path is of length $O(C)$. Thus, we have established that the size of the lifted Markov chain is at most $O(C|E|)$.

IV. The Pseudo-Lifting: Construction of *L2*

We will use the following notion of pseudo-lifting to design fast mixing Markov chains while maintaining the small size of the lifted chain.

Definition 4 (Pseudo-Lifting): Consider a Markov chain with transition matrix P and stationary distribution π defined on graph $G = (V,E)$. Given $\varepsilon \in (0, 1/2)$, a Markov chain with transition matrix $\widehat{P}$, stationary distribution $\widehat{\pi}$ on graph $\widehat{G} = (\widehat{V}, \widehat{E})$ is called a pseudo-lifting of P with parameter ε (or an ε pseudo-lifting of P) if there exists a many-to-one function $f : \widehat{V} \to V$, $T \subset \widehat{V}$ with $|T| = |V|$ such that the following holds: (a) for any $\widehat{u}, \widehat{v} \in \widehat{V}$, $(\widehat{u}, \widehat{v}) \in \widehat{E}$ only if $(f(\widehat{u}), f(\widehat{v})) \in E$, and (b) for any $u \in V$, $(1-\varepsilon)\pi(u) = \widehat{\pi}(f^{-1}(u) \cap T)$.

The above definition suggests that (by concentrating on set T), it is possible to simulate the stationary distribution π exactly (irrespective of ε) using the pseudo-lifting. Here, we present a basic construction. The primary reason to present it is to develop an intuition for more complicated hierarchical construction in Section V.

A. Construction

Let D be the diameter of G. First, select an arbitrary node v. Now, for each $w \in V$, there exist paths $\mathcal{P}_{wv}$ and $\mathcal{P}_{vw}$, from $w \to v$ and $v \to w$ respectively. We will assume that all the paths are of length D: this can be achieved by repeating same nodes. Now, we construct a pseudo-lifted graph $\widehat{G}$ starting from G.

First, create a new node v' which is a copy of the chosen vertex v. Then, add directed paths $\mathcal{P}'_{wv}$, a copy of $\mathcal{P}_{wv}$ from $w \to v'$; add $\mathcal{P}'_{vw}$ (copy of $\mathcal{P}_{vw}$) from v' to w. Each such addition creates $D-1$ new interior nodes. Thus, we have essentially created a *virtual star topology* using the paths of the old graph and totally added $O(nD)$ new nodes. Now, we define the ergodic flow $\widehat{Q}$ for this graph $\widehat{G}$ as follows: for an edge (i,j),

$$\widehat{Q}_{ij} = \begin{cases} \frac{\varepsilon}{2D}\pi_w & \text{if } (i,j) \in E(\mathcal{P}'_{wv}) \text{ or } E(\mathcal{P}'_{vw}) \\ (1-\varepsilon)Q_{ij} & \text{if } (i,j) \in E(G). \end{cases}$$

It is easy to check that $\sum_{ij} \widehat{Q}_{ij} = 1, \sum_j \widehat{Q}_{ij} = \sum_j \widehat{Q}_{ji}$. Hence it defines a Markov chain on $\widehat{G}$. The stationary distribution of this lifting is

$$\widehat{\pi}_i = \begin{cases} \frac{\varepsilon}{2D}\pi_w & \text{if } i \in (V(\mathcal{P}'_{wv}) \cup V(\mathcal{P}'_{vw}))\backslash\{w, v'\} \\ \left(1 - \varepsilon + \frac{\varepsilon}{2D}\right)\pi_i & \text{if } i \in V(G) \\ \frac{\varepsilon}{2D} & \text{if } i = v' \end{cases}$$

Given the above definition of $\widehat{Q}$ and corresponding stationary distribution $\widehat{\pi}$, it satisfies the pseudo-lifting definition with parameter $\delta = \varepsilon\left(1 - \frac{1}{2D}\right)$ if we choose $T = V(G)$ (i.e. T is the set of old nodes). Clearly, the number of edges in $\widehat{G}$ is $|E| + 2D|V|$. We claim the following bound on Mixing time.

Theorem 6: The mixing time of the Markov chain defined by $\widehat{Q}$ is $O\left(\frac{D}{\delta}\right)$.

Proof: Consider the following stopping rule.

(1) Walk until visiting v', and stop there with probability $\frac{\varepsilon}{2D}$.
(2) Otherwise, continue walking a directed path P'_{vw}, and choose an interior node of P'_{vw} uniformly at random, and stop there with probability $\frac{\varepsilon(D-1)}{2D}/(1 - \frac{\varepsilon}{2D})$.
(3) Otherwise, we will visit w, and stop there with probability $(1 - \varepsilon + \frac{\varepsilon}{2D})/\left(1 - \frac{\varepsilon}{2D} - \frac{\varepsilon(D-1)}{2D}\right)$.
(4) Otherwise, continue walking until getting a directed path P'_{wv}, and an interior node of P'_{wv} uniformly at random, and stop there.

With little calculation, it follows that when the stopping rule stops as per (1), (2) or (3), it induces the stopping distribution as per $\widehat{\pi}$ for nodes in $\{v'\} \cup V(G) \cup V(\mathcal{P}'_{vw})$. Now, when the step (4) is taken, till the walk gets on a directed path, it is evolving like original Markov chain with transition matrix P on nodes in $V(G)$. Since the distribution of the walk at the end of (3), conditioned to $V(G)$ is exactly π (note that $\widehat{\pi} = (1 - \gamma)\pi$, for an appropriate γ), it follows that the distribution π over the nodes of $V(G)$ is preserved under this walk till getting on a directed path. Therefore, it follows that when walk stops under (4), the distribution of $i \in \mathcal{P}'_{wv}$ is exactly $\widehat{\pi}_i$. Thus, we have established existence of a stopping rule that takes arbitrary starting distribution to stationary distribution $\widehat{\pi}$. Now, this stopping rule has average length $O(D/\varepsilon)$: since the probability of getting on a directed path P'_{wv} at w is $\Theta(\varepsilon/D)$, the expected number of walks in (1) and (4) is $O(D/\varepsilon)$. This completes the proof of Theorem 6. ■

B. Remark

To see the importance of Theorem 6, consider a *Barbell graph* of $2n$ nodes: *two complete graphs of n nodes connected by a single edge*. The size of this graph, $|E| = \Theta(n^2)$. Now, consider a random walk where the next transition is uniform among all the neighbor for each node. For such random walk, it is easy to check that $C = \Omega(n^2)$. Therefore, the mixing time of any lifted chain is at least $\Omega(n^2)$. However, this random walk is ill-designed because C can be decrease up to $O(n \log n)$ by defining its random walk in another way (i.e. increasing the probability of its linkage edge, and adding self-loops to non-linkage nodes not to change its stationary distribution). However, for this graph the diameter $D = O(1)$. Therefore, our construction provides pseudo-lifting with mixing time $O(1)$ with size of the lifted graph being $O(n^2)$ (i.e. only constant factor expansion in size).

V. Hierarchical construction for the Pseudo-Lifting: Construction of *L3*

Here, we describe a more sophisticated version of the pseudo-lifting described in the previous section. A caricature of this construction is presented in Figure 3. For this, we need an R-net $Y \subset V$ of graph $G = (V, E)$ defined as follows:

(i) For every $v \in V$, there exists $u \in Y$ such that the shortest path distance between u, v is at most R.
(ii) The distance between any two $y, z \in Y$ is more than R.

Such an R-net can be found in graph greedily. Given R-net Y, match each node w to the nearest $y \in Y$ (breaking ties arbitrarily). Let $C_y = \{w | \ w \text{ matched to } y\}$ for $y \in Y$. Clearly, $V = \cup_{y \in Y} C_y$. Finally, for each $y \in Y$ and for any $w \in C_y$ we have path $\mathcal{P}_{wy}, \mathcal{P}_{yw}$ between w and y of length exactly R; for each $y \in Y$, $\mathcal{P}_{yv}$ and $\mathcal{P}_{vy}$ between y, v of length exactly D.

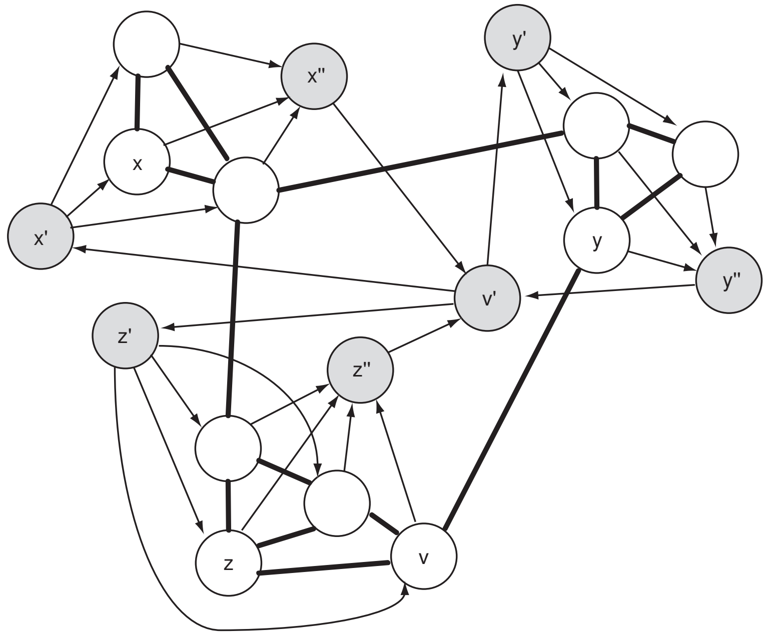

Fig. 3. A caricature of hierarchical pseudo lifting. The original graph is given by "white" nodes and "solid" lines. In the lifted graph, "grey" nodes and "arrow" lines (which represent copies of directional paths) are additional ones. For example, the "arrow" line from v' to x' represents a copy of the directed path from v to x. Nodes x, y and z form a net, and the vertex v is chosen arbitrarily.

A. Construction

Now, we construct the lifted graph $\widehat{G}$. As before, select an arbitrary node $v \in V$ and create its copy v' again. Further, for each $y \in Y$ create two copies y'_1 and y'_2. Now, add directed paths $\mathcal{P}'_{wy}$, a copy of $\mathcal{P}_{wy}$, from w to y'_1 and add $\mathcal{P}'_{yv}$, a copy of $\mathcal{P}_{yv}$, from y'_1 to v'. Similarly, add $\mathcal{P}'_{vy}$ and $\mathcal{P}'_{yw}$ between v', y'_2 and y'_2, w. This construction adds $2D|Y| + nR$ edges in G to give $\widehat{G}$. Now, the ergodic flow $\widehat{Q}$ on $\widehat{G}$ is defined as follows: for any (i,j) of $\widehat{G}$,

$$\widehat{Q}_{ij} = \begin{cases} \frac{\varepsilon}{2(R+D)}\pi_w & \text{if } (i,j) \in E(\mathcal{P}'_{wy}) \text{ or } E(\mathcal{P}'_{yw}) \\ \frac{\varepsilon}{2(R+D)}\pi(C_y) & \text{if } (i,j) \in E(\mathcal{P}'_{yv}) \text{ or } E(\mathcal{P}'_{vy}) \\ (1-\varepsilon)Q_{ij} & \text{if } (i,j) \in E(G) \end{cases},$$

where $\pi(C_y) = \sum_{w \in C_y} \pi_w$. It can be checked that $\sum_{ij} \widehat{Q}_{ij} = 1, \sum_j \widehat{Q}_{ij} = \sum_j \widehat{Q}_{ji}$. Hence it defines a Markov chain on $\widehat{G}$. The stationary distribution of this lifted chain is

$$\widehat{\pi}_i = \begin{cases} \frac{\varepsilon}{2(R+D)}\pi_w & \text{if } i \in (V(\mathcal{P}'_{wy}) \cup V(\mathcal{P}'_{yw})) \backslash \{w, y'_1, y'_2\} \\ \frac{\varepsilon}{2(R+D)}\pi(C_y) & \text{if } i \in (V(\mathcal{P}'_{yv}) \cup V(\mathcal{P}'_{vy})) \backslash \{v'\} \\ (1-\varepsilon)\pi_i + \frac{\varepsilon}{2(R+D)}\pi_i & \text{if } i \in V(G) \\ \frac{\varepsilon}{2(R+D)} & \text{if } i = v' \end{cases}$$

To establish this lifting as pseudo-lifting of original Markov chain P, consider $T = V(G)$ and parameter $\delta = \varepsilon\left(1 - \frac{1}{2(R+D)}\right)$. The $\widehat{G}$ has exactly $|E| + 2R|V| + 2D|Y|$ edges as noted earlier. We state and prove the following result about the mixing time.

Theorem 7: The mixing time of the Markov chain defined by $\widehat{Q}$ is $O\left(\frac{D}{\delta}\right)$.

Proof: Consider the following stopping rule. Walk until visiting v', and toss a coin X with the following probability.

$$X = \begin{cases} 0 & \text{with probability } \frac{\varepsilon}{2(R+D)} \\ 1 & \text{with probability } \frac{\varepsilon D}{2(R+D)} \\ 2 & \text{with probability } \frac{\varepsilon(R-1)}{2(R+D)} \\ 3 & \text{with probability } 1-\delta \\ 4 & \text{with probability } \frac{\varepsilon(R-1)}{2(R+D)} \\ 5 & \text{with probability } \frac{\varepsilon D}{2(R+D)} \end{cases}$$

If $X = 0$, stop at v'. Else if, $X = 1$ and 5, walk until getting on to $\mathcal{P}'_{vy}$ and $\mathcal{P}'_{yv}$ respectively, select its node on the path except for v' uniformly at random to stop there. If $X = 2$ and 4, walk until getting $\mathcal{P}'_{yw}$ and $\mathcal{P}'_{wy}$ respectively, select its interior node on the path uniformly at random to stop there. If $X = 3$, walk until getting an old node, and stop there. It can be checked, using arguments similar to that in proof of Theorem 6, that the distribution of the stopped node is precisely $\widehat{\pi}$.

Similar to the proof of Theorem 6, we can show that the expected length of stopping rule is $O(\frac{R+D}{\varepsilon}) = O(\frac{D}{\varepsilon}) = O(\frac{D}{\delta})$. This is primarily true because the probability of getting on a directed path $\mathcal{P}'_{wy}$ at w is $\Theta(\varepsilon/(R+D))$. ■

B. Application to Constant Doubling Dimension Graph

Here, we apply the hierarchical construction to the case of the graph with constant doubling dimension. To this end, let G have constant doubling dimension ρ with respect to the shortest path metric.

Corollary 8: Graph $G = (V, E)$ with constant doubling dimension ρ and diameter D. Then, the hierarchical construction gives δ pseudo-lifting $\widehat{G}$ with its size $O(D|V|^{1-\frac{1}{\rho+1}})$ and mixing time is $O(D/\delta)$.

Proof: The property of doubling dimension graph implies that there exists an R-net Y such that $|Y| \leq (2D/R)^\rho$ (cf. [2]). Consider $R = D2^{\frac{\rho}{\rho+1}}|V|^{-\frac{1}{\rho+1}}$. This is appropriate choice as by definition of doubling dimension, $|V| \leq D^\rho$ and hence $R = D2^{\frac{\rho}{\rho+1}}|V|^{-\frac{1}{\rho+1}} > D|V|^{-\frac{1}{\rho+1}} > |V|^{\frac{1}{\rho}-\frac{1}{\rho+1}} > 1$. Given this, the size of the lifted graph $\widehat{G}$ is

$$|\widehat{E}| = |E| + 2R|V| + 2D|Y| \leq |E| + 2D\left(\frac{2^{\frac{\rho}{\rho+1}}}{|V|^{\frac{1}{\rho+1}}}\right)|V| + 2D\left(2\frac{|V|^{\frac{1}{\rho+1}}}{2^{\frac{\rho}{\rho+1}}}\right)^\rho = |E| + O(D|V|^{1-\frac{1}{\rho+1}}).$$

Since $|E| = O(|V|)$ and $D = \Omega(|V|^{1/\rho})$, we have that $|\widehat{E}| = O(D|V|^{1-\frac{1}{\rho+1}})$. ■

VI. Conclusion

Motivated by application of Markov chain in designing fast distributed algorithms, we considered the problem of constructing lifting Markov chains for arbitrary graphs with small mixing time and size. We first obtained a slimmer lifting by means of expander graphs compared to the result of Chen et. al. [5]. To overcome the limitation of this notion of lifting in terms of the mixing time, we introduced new notion of lifting, the pseudo-lifting. This allows us to construct a simple lifting that have the fastest mixing time for a given graph irrespective of the starting Markov chain and the stationary distribution. Using our hierarchical construction, we showed how the size of lifting can be optimized by utilizing information about the geometry of the underlying graph. We believe that our technique of hierarchical construction can be applied recursively to reduce size of lifting with little loss in the mixing time. Our pseudo lifting can allow for designing fast distributed iterative algorithms.

References

[1] D. J. Aldous. Some inequalities for reversible markov chains. In *J. London Math. Soc. 25*, pages 564–576, 1982.

[2] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[3] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing markov chain on a graph. *SIAM Review*, 46(4):667–689, 2004.

[4] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Gossip algorithms: Design, analysis and applications. In *Proceedings of IEEE INFOCOM 2005*, pages 1653–1664, 2005.

[5] F. Chen, L. Lov'asz, and I. Pak. Lifting markov chains to speed up mixing. In *STOC '99: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 275–281, New York, NY, USA, 1999. ACM Press.

[6] P. Diaconis, S. Holmes, and R. Neal. Analysis of a non-reversible markov chain sampler. *Ann. Appl. Probab.*, 10:726–752, 2000.

[7] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, page 534, Washington, DC, USA, 2003. IEEE Computer Society.

[8] Juha Heinonen, editor. *Lectures on Analysis on Metric Spaces*. Springer, 2001.

[9] R. Horn and C. Johnson, editors. *Matrix Analysis*. Cambridge Univ. Press., 1985.

[10] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 561–568, New York, NY, USA, 2004. ACM Press.

[11] Claire Kenyon, Yuval Rabani, and Alistair Sinclair. Biased random walks, lyapunov functions, and stochastic analysis of best fit bin packing. In *SODA '96: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 351–358, Philadelphia, PA, USA, 1996. Society for Industrial and Applied Mathematics.

[12] P. Kolman and C. Scheideler. Improved bounds for the unsplittable flow problem. In *SODA '02: Proceedings of the 13th annual ACM-SIAM symposium on Discrete algorithms*, pages 184–193, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics.

[13] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. In *FOCS '88: Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 422–431, 1988.

[14] J. Sun S. Boyd, P. Diaconis and L. Xiao. Fastest mixing markov chain on a path. In *The American Mathematical Monthly*, pages 113(1):70–74, 2006.

[15] John N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

[16] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. In *Systems and Control Letters*, pages 53:65–78, 2004.