

Intersection conductance and canonical alternating paths, methods for general finite Markov chains

Ravi Montenegro *

December 5, 2011

Abstract

We extend the conductance and canonical paths methods to the setting of general finite Markov chains, including non-reversible non-lazy walks. The new path method is used to show that a known bound for mixing time of a lazy walk on a Cayley graph with symmetric generating set also applies to the non-lazy non-symmetric case, often even when there is no holding probability.

Keywords : Mixing time, conductance, canonical paths, evolving sets, Cayley graph.

1 Introduction

Beginning with the work of Jerrum and Sinclair, geometric concepts such as conductance [6, 8] and canonical paths [12, 4] have played an important role in studying the mixing time of finite ergodic Markov chains. These methods originally applied only to reversible lazy walks, and while little is lost in dropping reversibility and some amount of laziness [4, 5, 10], extensions which allow for dropping both conditions tend to be weak or difficult to use [5, 9, 10]. Similar difficulty has been encountered with the method of blocking conductance [7], a geometric approach to sharpening conductance bounds by including a notion of vertex congestion but which has only been successfully applied to a few problems. In this paper we develop an extension of conductance to the general (non-lazy non-reversible) setting, along with a further extension which sharpens bounds by using a simple notion of vertex congestion. This makes it easier to use and allows for the proof of a new canonical path theorem which applies to general finite ergodic Markov chains.

Recall that if a finite Markov kernel P with sample space V is irreducible and aperiodic ($\exists N \in \mathbb{N}, \forall x, y \in V : P^N(x, y) > 0$) then it has a unique stationary distribution π satisfying $\pi P = \pi$ and is ergodic ($\forall x \in V : P^n(x, \cdot) \xrightarrow{n \rightarrow \infty} \pi$). The walk is lazy if the holding probability $\alpha = \min_{v \in V} P(v, v)$ is at least $1/2$, and is reversible if the time-reversal (adjoint) $P^*(x, y) = \frac{\pi(y)P(y, x)}{\pi(x)}$ satisfies $P^* = P$. The L^2 mixing time $\tau(\epsilon) = \max_{x \in V} \min\{n : \|k_n^x - 1\|_{2, \pi} \leq \epsilon\}$ is the worst-case time for standard deviation of the density $k_n^x(y) = \frac{P^n(x, y)}{\pi(y)}$ to drop to ϵ .

Jerrum and Sinclair [6] and Lawler and Sokal [8] showed that mixing time of a lazy reversible walk can be bounded in terms of the conductance, also known as the Cheeger constant,

$$\Phi = \min_{\emptyset \subsetneq A \subsetneq V} \Phi(A) \quad \text{where} \quad \Phi(A) = \frac{Q(A, A^c)}{\pi(A)\pi(A^c)}$$

*Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854, ravi_montenegro@uml.edu.

and the ergodic flow from $A \subset V$ to $B \subset V$ is $Q(A, B) = \sum_{x \in A, y \in B} \pi(x)P(x, y)$.

Theorem 1.1. *The mixing time of a lazy reversible finite ergodic Markov chain is*

$$\tau(\epsilon) \leq \frac{8}{\Phi^2} \log \frac{1}{\epsilon \sqrt{\pi_0}}$$

where $\pi_0 = \min_{v \in V} \pi(v)$.

Since π is stationary then $\pi(v) = Q(A, v) + Q(A^c, v)$ and $Q(A, A^c) = Q(A^c, A)$, and so in the case of a lazy walk

$$\frac{1}{2} \sum_{v \in V} \min\{Q(A, v), Q(A^c, v)\} = \frac{Q(A, A^c) + Q(A^c, A)}{2} = Q(A, A^c)$$

i.e. the flow $Q(A, \cdot)$ from A and $Q(A^c, \cdot)$ from A^c intersect at exactly the flow $Q(A, A^c) + Q(A^c, A)$ crossing between A and A^c . The intersection of the flows from A and A^c can then be used to define a modified form of conductance which agrees with normal conductance in the lazy case.

Definition 1.2. *The intersection conductance $\hat{\Phi}(A)$ of $A \subset V$ (and A^c) is given by*

$$\hat{\Phi}(A) = \hat{\Phi}(A, A^c) = \frac{\sum_{v \in V} \min\{Q(A, v), Q(A^c, v)\}}{2\pi(A)\pi(A^c)}$$

The intersection conductance is $\hat{\Phi} = \min_{\emptyset \subsetneq A \subsetneq V} \hat{\Phi}(A)$.

Intuitively, when flow from A and A^c overlap significantly then a walk starting in A will quickly mix with the ‘unoccupied’ space starting in A^c , and so good mixing may be expected. As our first result, in Section 3 we make this rigorous by showing that mixing time for general (non-lazy non-reversible) finite Markov chains can be bounded using the intersection conductance.

Theorem 1.3. *The mixing time of a finite ergodic Markov chain is*

$$\tau(\epsilon) \leq \frac{12}{\hat{\Phi}^2} \log \frac{1}{\epsilon \sqrt{\pi_0}}$$

As a further extension, suppose that a threshold t is fixed and ergodic flow is counted up to at most a t fraction of vertex capacity, i.e. work with *threshold limited ergodic flow* $Q_t(A, v) = \min\{Q(A, v), t\pi(v)\}$ instead of $Q(A, v)$. When ergodic flow is well distributed among vertices then it may be possible to make t quite small without decreasing the ergodic flow significantly, and so the optimal choice of t will measure some form of vertex congestion. More formally, and in the greater generality of conductance profile:

Definition 1.4. *Given $t > 0$ the intersection threshold conductance of $A \subset V$ is*

$$\hat{\Phi}_t(A) = \frac{\sum_{v \in V} \min\{Q_t(A, v), Q_t(A^c, v)\}}{2\pi(A)\pi(A^c)}$$

The intersection threshold conductance profile $\hat{\Phi}_t(r)$ is

$$\hat{\Phi}_t(r) = \begin{cases} \min_{0 < \pi(A) \leq r} \hat{\Phi}_t(A) & \text{if } r \leq 1/2 \\ \hat{\Phi}_t(1/2) & \text{if } r > 1/2 \end{cases}$$

The intersection threshold conductance is $\hat{\Phi}_t = \min_{\emptyset \subsetneq A \subsetneq V} \hat{\Phi}_t(A) = \hat{\Phi}_t(1/2)$.

For a lazy walk this is equal to normal conductance when $t \geq 1/2$, i.e. $\hat{\Phi}_{1/2}(A) = \Phi(A)$, while more generally $\hat{\Phi}_{1/2}(A) = \hat{\Phi}(A) \leq 1$. Our earlier result can be sharpened to show that the mixing time bound is t times smaller when a threshold is used. In particular:

Theorem 1.5. *Given threshold $t > 0$ the mixing time of a finite ergodic Markov chain is*

$$\tau(\epsilon) \leq \frac{12 \max\{t, \hat{\Phi}_t\}}{\hat{\Phi}_t^2} \log \frac{1}{\epsilon \sqrt{\pi_0}}$$

and more generally

$$\tau(\epsilon) \leq \begin{cases} \int_{4\pi_0}^{4/\epsilon^2} \frac{12 \max\{t, \hat{\Phi}_t(r)\}}{r \hat{\Phi}_t(r)^2} dr & \text{in general} \\ \int_{\pi_0}^{1/\epsilon^2} \frac{6 \max\{t, \hat{\Phi}_t(r)\}}{r \hat{\Phi}_t(r)^2} dr & \text{if } r \frac{\hat{\Phi}_t^2\left(\frac{1}{1+r^2}\right)}{\max\left\{t, \hat{\Phi}_t\left(\frac{1}{1+r^2}\right)\right\}} \text{ is convex in } r \end{cases}$$

In their seminal paper Jerrum and Sinclair introduced the method of canonical paths as a means of lower bounding conductance [6].

Definition 1.6. *A canonical path γ_{xy} is a path from x to y using only valid transitions of P :*

$$x = x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow x_n = y$$

Let $\Gamma = \{\gamma_{xy} : x, y \in V\}$ include a canonical path for each pair of distinct vertices $x, y \in V$. The edge-congestion is given by

$$\rho_e = \rho_e(\Gamma) = \max_{u, v \in V: P(u, v) > 0} \frac{1}{\pi(u)P(u, v)} \sum_{(u, v) \in \gamma_{xy} \in \Gamma} \pi(x)\pi(y)$$

Theorem 1.7 (Jerrum and Sinclair). *The mixing time of a lazy reversible finite ergodic Markov chain is*

$$\tau(\epsilon) \leq 8\rho_e^2 \log \frac{1}{\epsilon \sqrt{\pi_0}}$$

Sinclair later showed a Poincaré type bound with order ρ_e^2 replaced by $\rho_e \ell$ where $\ell = \max |\gamma_{xy}|$ is the length of the longest path, usually an improvement on the ρ_e^2 result [12]. Diaconis and Strook extend this to non-lazy reversible walks if Γ also includes odd length paths γ_{xx} from each vertex to itself [4]. Various authors have observed that the Poincaré approach applies to lazy non-reversible walks as well, and to non-lazy walks at the cost of a factor of α^{-1} [5, 9, 10].

As our second main result, in Section 4 we find that a generalization of the canonical path method can be used to lower bound intersection threshold conductance, giving a canonical path method for general finite Markov chains. In particular we consider *canonical alternating paths*, even length paths which alternate between forward and reversed edges of P , i.e.

$$x = x_0 \rightarrow x_1 \leftarrow x_2 \rightarrow x_3 \leftarrow \cdots \rightarrow x_{2n-1} \leftarrow x_{2n} = y$$

This is equivalent to a path from x to y alternating between edges of P and P^* . The canonical paths used in each of the methods discussed in the previous paragraph can be used to construct

canonical alternating paths – for instance by adding self-loops at x_1, x_2, \dots, x_n in the lazy case – so this new type of path generalizes each of those settings. For an appropriately defined notion of vertex congestion ρ_v we show that

$$\hat{\Phi}_{\rho_v/\rho_e} \geq 1/2\rho_e.$$

A mixing time result then follows from Theorem 1.5. With some extra work a bound is also possible if standard canonical paths are used:

Theorem 1.8. *A finite ergodic Markov chain with canonical alternating paths Γ has mixing time*

$$\tau(\epsilon) \leq 48\rho_e \max\{\rho_v, \frac{1}{2}\} \log \frac{1}{\epsilon\sqrt{\pi_0}}$$

If ordinary canonical paths are used and $\alpha = \min_{v \in V} P(v, v)$ is the minimal holding probability then

$$\tau(\epsilon) \leq 4\rho_v \max\left\{\frac{\rho_v}{\alpha}, \rho_e\right\} \log \frac{1}{\epsilon\sqrt{\pi_0}}$$

The congestions satisfy $\rho_v \leq \rho_e$, so this generalizes Jerrum and Sinclair’s result. Our results also hold for multicommodity flows.

We finish in Section 5 with a few examples of how our new tools can be used. First, it is shown that known complexity results for the lazy simple random walk on an undirected graph apply just as well to the simple random walk on an Eulerian directed graph, i.e. a strongly connected graph with in-degree=out-degree at each vertex, both when there is a self-loop at each vertex and often even without self-loops. A more interesting problem is to study random walks on Cayley graphs, i.e. walks on groups, for which we show that known bounds [1, 3] for a lazy walk or a walk with a symmetric set of generators can be extended to the non-lazy non-symmetric case.

2 Evolving Sets

The Evolving Set methodology [11, 10] will be required for the proof of Theorem 1.5, from which most results of the paper follow. We give here a very brief introduction to those elements required in our proof. The reader interested only in the canonical path results may skip to Section 4.

One approach to relating a property of sets (conductance) to a property of the original walk (mixing time) is to construct a *dual process*: a walk P_D on some state space Ω and a *link*, or transition matrix, Λ from Ω to V such that

$$P\Lambda = \Lambda P_D.$$

In particular, $P^n\Lambda = \Lambda P_D^n$ and so the evolution of P^n and P_D^n will be closely related. A natural candidate to link a walk on sets to a walk on states is the projection $\Lambda(S, y) = \frac{\pi(y)}{\pi(S)} 1_S(y)$. Diaconis and Fill [2] have shown that for certain classes of Markov chains a walk based on the Evolving Set process discussed below is the unique dual process with link Λ , so this is the most natural walk on sets to consider. Our discussion of Evolving Sets will be based on work of Morris and Peres [11] in a slightly improved form of Montenegro and Tetali [10].

To understand the method requires some new terminology.

Definition 2.1. Given set $A \subset V$ a step of the evolving set process is given by choosing $u \in [0, 1]$ uniformly at random, and transitioning to the set

$$A_u = \{y \in V \mid \mathbf{Q}(A, y) \geq u\pi(y)\} = \{y \in V \mid \mathbf{P}^*(y, A) \geq u\}.$$

The root profile $\psi : (0, \infty) \rightarrow [0, 1]$ is given by

$$\psi(r) = \min_{0 < \pi(A) \leq r} \psi(A) \quad \text{where} \quad \psi(A) = 1 - \frac{\int_0^1 \sqrt{\pi(A_u)(1 - \pi(A_u))} du}{\sqrt{\pi(A)(1 - \pi(A))}}$$

when $r \in (0, 1/2]$, and $\psi(r) = \psi(1/2) = \min_{A \notin \{\emptyset, V\}} \psi(A)$ when $r > 1/2$.

The claim $\psi(1/2) = \min_{A \notin \{\emptyset, V\}} \psi(A)$ follows from the relation $\psi(A) = \psi(A^c)$, a consequence of $(A_u)^c = (A^c)_{1-u}$ for every u with $\#v : \mathbf{Q}(A, v) = u\pi(v)$, i.e. u -a.e. since V is finite.

Since $\mathbf{Q}(V, v) = \pi(v)$ then A_u consists of those vertices receiving at least a u -fraction of their steady state probability from A . If u is chosen uniformly from $[0, 1]$ then

$$\mathbb{E}\pi(A_u) = \int_0^1 \pi(A_u) du = \sum_{y \in V} \pi(y) \frac{\mathbf{Q}(A, y)}{\pi(y)} = \pi(A)$$

and so by Jensen's Inequality $\mathbb{E}\sqrt{\pi(A_u)(1 - \pi(A_u))} \leq \sqrt{\pi(A)(1 - \pi(A))}$, with equality if and only if $\pi(A_u) = \pi(A)$ u -a.e. It follows that a large root profile $\psi(A)$ indicates that $\pi(A_u)$ differs significantly from $\pi(A)$, and in particular the flow from A is spread over a large space and so the walk expands quickly from A . Morris and Peres [11] make this intuition rigorous, although we use a sharper result of Montenegro and Tetali [10]:

Theorem 2.2. A finite ergodic Markov chain with root profile lower bounded by $\psi(r)$ has

$$\tau(\epsilon) \leq \begin{cases} \int_{4\pi_0}^{4/\epsilon^2} \frac{dr}{r\psi(r)} & \text{in general} \\ \int_{\pi_0}^{1/\epsilon^2} \frac{dr}{2r\psi(r)} & \text{if } r\psi\left(\frac{1}{1+r^2}\right) \text{ is convex} \end{cases}$$

The root profile is typically lower bounded by writing constraints of interest in terms of the evolving set process. For instance, when conductance is being considered then a lazy walk has $A \subset A_u$ when $u \leq 1/2$ and $A_u \subset A$ when $u > 1/2$ and so

$$\mathbf{Q}(A, A^c) = \sum_{v \in A^c} \frac{\mathbf{Q}(A, v)}{\pi(v)} \pi(v) = \sum_{v \in A^c} \int_0^{1/2} \mathbf{1}_{\mathbf{Q}(A, v)/\pi(v) \geq u} \pi(v) du = \int_0^{1/2} \pi(A_u \setminus A) du$$

The identities $\int_0^1 \pi(A_u) du = \pi(A)$ and $\mathbf{Q}(A, A^c) = \mathbf{Q}(A^c, A)$ can be used to write this as an area problem:

$$\mathbf{Q}(A, A^c) = \int_0^{1/2} (\pi(A_u) - \pi(A)) du = \int_{1/2}^1 (\pi(A) - \pi(A_u)) du = \mathbf{Q}(A^c, A) \quad (2.1)$$

Breaking the definition of $\psi(A)$ into an integral of $u \in [0, 1/2]$ and one of $u \in [1/2, 1]$, applying Jensen's Inequality, and then making a few simplifications leads to the relation $\psi(A) \geq \Phi(A)^2/2$. Theorem 2.2 then implies a stronger version of Jerrum and Sinclair's result given in Theorem 1.1:

Theorem 2.3. *A lazy finite ergodic Markov chain has $\psi(A) \geq \Phi^2(A)/2$ and mixing time*

$$\tau(\epsilon) \leq \frac{2}{\Phi^2} \log \frac{1}{\epsilon \sqrt{\pi_0}}$$

3 Mixing bounds with Threshold Conductances

In this section we prove a lower bound on the root profile $\psi(A)$ in terms of the (intersection) threshold conductance of Definition 1.4. By Theorem 2.2 this induces upper bounds on mixing time, including the main result of this paper, Theorem 1.5, a bound on mixing time in terms of the intersection threshold conductance. To further indicate the improvement provided by use of thresholds we also give a bound in terms of a quantity which more strongly resembles ordinary conductance:

Definition 3.1. *If $A \subset V$ and $t \in [0, 1]$ then the threshold conductance $\hat{\Phi}_t(A)$ is*

$$\hat{\Phi}_t(A) = \frac{\min\{Q_t(A, A^c), Q_t(A^c, A)\}}{\pi(A)\pi(A^c)}$$

The threshold conductance profile $\hat{\Phi}_t(r)$ and the threshold conductance $\hat{\Phi}_t$ are defined as were the related quantities in Definition 1.4.

For a lazy walk $Q_{1/2}(A, A^c) = Q(A, A^c)$ and $\hat{\Phi}_{1/2}(A) = \Phi(A)$ agree with standard notions of ergodic flow and conductance. Intuitively, if ergodic flow from A to A^c and from A^c to A are not unduly concentrated on a few vertices then $\hat{\Phi}_t(A) \approx \Phi(A)$ even for fairly small t . The extra information provided by t will be used here to show a substantially improved result.

And now to the main result:

Theorem 3.2. *Given $A \subset V$ then*

$$\psi(A) \geq \begin{cases} \frac{1}{12t} \min\{\hat{\Phi}_t(A)^2, t\hat{\Phi}_t(A)\} \\ \frac{\min\{\alpha, t\}}{4t^2} \hat{\Phi}_t(A)^2 \end{cases}$$

The second bound simply sharpens a special case of the first, as $\hat{\Phi}_{\min\{\alpha, t\}}(A) \geq \frac{\min\{\alpha, t\}}{t} \hat{\Phi}_t(A) \leq 2 \min\{\alpha, t\}$. When $\hat{\Phi}_t(r) = \Theta(\Phi(r))$ then this leads to a mixing bound $\frac{t^2}{\min\{\alpha, t\}}$ times that of Theorem 1.1, a significant improvement when the threshold is small. In the extreme case when ergodic flow is spread uniformly over the complement then $\hat{\Phi}_t(r) = \Phi(r)$ when $t = \min \frac{Q(A, A^c)}{\min\{\pi(A), \pi(A^c)\}} \leq \Phi$, improving the mixing time bound from $\tau(\epsilon) \leq \frac{2}{\Phi^2} \log \frac{1}{\epsilon \sqrt{\pi_0}}$ to $\tau(\epsilon) = O\left(\frac{1}{\Phi} \log \frac{1}{\epsilon \sqrt{\pi_0}}\right)$ if $\alpha \geq \Phi$. With conductance profile this can even match the best case lower bound of $\tau(\epsilon) = \Omega\left(\frac{1}{\Phi}\right)$.

Proof. Start with the first bound of the theorem, that $\psi(A) \geq \frac{1}{12t} \min\{\check{\phi}_t(A)^2, t\check{\phi}_t(A)\}$. The proof will involve relating evolving sets to the numerator of $\hat{\Phi}_t(A)$, with (3.2) replacing the use of (2.1) for conductance. Breaking the definition of $\psi(A)$ into an integral over $u \in [0, t]$ and one over $u \in [t, 1]$, applying Jensen's Inequality, and then making a few simplifications completes the proof. The argument for the second case is similar.

Suppose $t \geq 1/2$. Since $\min\{Q(A, v), Q(A^c, v)\} \leq \pi(v)/2$ then $\min\{Q_t(A, v), Q_t(A^c, v)\} = \min\{Q_{1/2}(A, v), Q_{1/2}(A^c, v)\}$ and consequently $\hat{\Phi}_t(A) = \hat{\Phi}_{1/2}(A)$. The result at $t = 1/2$ is then stronger than that at $t > 1/2$, so without loss assume that $t \leq 1/2$.

When $u \leq 1/2$ then

$$\begin{aligned} \pi(A_u \setminus A_{1-u}) &= \pi(\{v \in V : Q(A, v) \geq u\pi(v), Q(A^c, v) > u\pi(v)\}) \\ &= \pi(\{v \in V : \min\{Q(A, v), Q(A^c, v)\} \geq u\pi(v)\}) \\ &\quad - \pi(\{v \in V : Q(A, v) \geq u\pi(v), Q(A^c, v) = u\pi(v)\}) \end{aligned}$$

The set $\{u \in [0, 1] : \exists v \in V, Q(A^c, v) = u\pi(v)\}$ is finite and so it has Lebesgue measure zero. Since $u \leq 1/2$ then $A_{1-u} \subset A_u$ and so

$$\begin{aligned} \sum_{v \in V} \min\{Q_t(A, v), Q_t(A^c, v)\} &= \sum_{v \in V} \int_0^t \pi(v) \delta_{\min\{Q_t(A, v), Q_t(A^c, v)\} \geq u\pi(v)} du \\ &= \int_0^t [\pi(A_u \setminus A_{1-u}) + \pi(\{v \in V : Q(A, v) \geq u\pi(v), Q(A^c, v) = u\pi(v)\})] du \\ &= \int_0^t (\pi(A_u) - \pi(A_{1-u})) du + 0 \\ &= \int_0^t (\pi(A_u) - \pi(A)) du + \int_{1-t}^1 (\pi(A) - \pi(A_u)) du \tag{3.2} \\ &= \int_0^t (\pi(A_u) - \pi(A)) du + \int_0^t (\pi((A^c)_u) - \pi(A^c)) du \end{aligned}$$

The final equality uses the relation $(A_u)^c = (A^c)_{1-u}$ for a.e. $u \in [0, 1]$. It follows that

$$\max \left\{ \int_0^t (\pi(A_u) - \pi(A)) du, \int_0^t (\pi((A^c)_u) - \pi(A^c)) du \right\} \geq \hat{\Phi}_t(A) \pi(A) \pi(A^c)$$

Since $\psi(A) = \psi(A^c)$ and $\hat{\Phi}_t(A) = \hat{\Phi}_t(A^c)$ then the roles of A and A^c can be swapped and this inequality will still hold. So without loss assume $\int_0^t (\pi(A_u) - \pi(A)) du \geq \int_0^t (\pi((A^c)_u) - \pi(A^c)) du$. By Jensen's Inequality, the Martingale identity $\int_0^1 \pi(A_u) du = \pi(A)$, and concavity of $f(x) = \sqrt{x(1-x)}$ then

$$\begin{aligned} \int_0^1 f(\pi(A_u)) du &\leq t f \left(\int_0^t \pi(A_u) \frac{du}{t} \right) + (1-t) f \left(\int_t^1 \pi(A_u) \frac{du}{1-t} \right) \\ &= t f \left(\pi(A) + \frac{\int_0^t (\pi(A_u) - \pi(A)) du}{t} \right) + (1-t) f \left(\pi(A) - \frac{\int_0^t (\pi(A_u) - \pi(A)) du}{1-t} \right) \end{aligned}$$

By concavity of f this is decreasing in $\int_0^t (\pi(A_u) - \pi(A)) du \geq \hat{\Phi}_t(A) \pi(A) \pi(A^c)$ and so it is maximized at the lower bound. The root profile is then bounded by

$$\begin{aligned} 1 - \psi(A) &\leq t \sqrt{\left(1 + \frac{\hat{\Phi}_t(A)(1 - \pi(A))}{t}\right) \left(1 - \frac{\hat{\Phi}_t(A)\pi(A)}{t}\right)} \\ &\quad + (1-t) \sqrt{\left(1 - \frac{\hat{\Phi}_t(A)(1 - \pi(A))}{1-t}\right) \left(1 + \frac{\hat{\Phi}_t(A)\pi(A)}{1-t}\right)} \end{aligned} \tag{3.3}$$

Treating t and $\hat{\Phi}_t(A)$ as constants, this is concave in $\pi(A)$ with negative derivative at $\pi(A) = 0$ when $\hat{\Phi}_t(A) < 1 - 2t$. The bound in (3.3) is then maximized at $\pi(A) = 0$, with

$$\begin{aligned} 1 - \psi(A) &\leq t\sqrt{1 + \frac{\hat{\Phi}_t(A)}{t}} + (1-t)\sqrt{1 - \frac{\hat{\Phi}_t(A)}{1-t}} \\ &\leq t\left(1 + \frac{\hat{\Phi}_t(A)}{2t} - \frac{1}{12}\min\left\{\frac{\hat{\Phi}_t(A)^2}{t^2}, \frac{\hat{\Phi}_t(A)}{t}\right\}\right) + (1-t)\left(1 - \frac{\hat{\Phi}_t(A)}{2(1-t)}\right) \end{aligned}$$

The second inequality applies the relations $\sqrt{1+x} \leq 1 + \frac{x}{2} - \delta_{x \geq 0} \frac{\min\{x^2, x\}}{12}$ and $\sqrt{1-x} \leq 1 - \frac{x}{2}$. This gives the theorem when $\hat{\Phi}_t(A) < 1 - 2t$. When $\hat{\Phi}_t(A) \geq 1 - 2t$ then (3.3) is maximized at $\pi(A) = \frac{1-(1-2t)\hat{\Phi}_t(A)}{2}$, leading to the bound $\psi(A) \geq \frac{\hat{\Phi}_t(A)^2}{2} \geq \frac{1}{6t} \min\{\hat{\Phi}_t(A)^2, t\hat{\Phi}_t(A)\}$.

Now consider the second bound of the theorem, $\psi(A) \geq \frac{\min\{\alpha, t\}}{4t^2} \hat{\Phi}_t(A)^2$.

Suppose $t \geq 1/2$. If $\alpha \geq 1/2$ then $\hat{\Phi}_t(A) = \hat{\Phi}_{1/2}(A)$ and so the bound follows from the case of $t = 1/2$. If $\alpha < 1/2$ then $\hat{\Phi}_{1/2}(A) \geq \frac{1}{2t}\hat{\Phi}_t(A)$ and again the bound follows from the case of $t = 1/2$. The result at $t = 1/2$ is then stronger than that at $t > 1/2$, so without loss assume that $t \leq 1/2$.

Let $T = \min\{\alpha, t\} \leq 1/2$. Observe that $\mathbf{Q}_T(A, v) = \int_0^T \pi(A_u \cap \{v\}) du$ and so

$$\begin{aligned} \mathbf{Q}_T(A, V) &= \sum_{v \in V} \int_0^T \pi(A_u \cap \{v\}) du = \int_0^T \pi(A_u) du \\ \mathbf{Q}_T(A, V) &= \mathbf{Q}_T(A, A) + \mathbf{Q}_T(A, A^c) = \pi(A)T + \mathbf{Q}_T(A, A^c) \end{aligned}$$

Then $\int_0^T \pi(A_u) du = \mathbf{Q}_T(A, V) = \pi(A)T + \mathbf{Q}_T(A, A^c)$. Swapping the roles of A and A^c in this identity, and recalling that u -a.e. $(A_u)^c = (A^c)_{1-u}$, it follows that

$$\int_{1-T}^1 \pi(A_u) du = \int_0^T (1 - \pi((A^c)_u)) du = \pi(A)T - \mathbf{Q}_T(A^c, A)$$

By Jensen's Inequality and the identity $\int_0^1 \pi(A_u) du = \pi(A)$:

$$\begin{aligned} \int_0^1 \sqrt{\pi(A_u)(1 - \pi(A_u))} du &= \int_0^T + \int_T^{1-T} + \int_{1-T}^1 f(\pi(A_u)) du \\ &\leq T f\left(\int_0^T \pi(A_u) \frac{du}{T}\right) + (1-2T) f\left(\int_T^{1-T} \pi(A_u) \frac{du}{1-2T}\right) + T f\left(\int_{1-T}^1 \pi(A_u) \frac{du}{T}\right) \\ &= T f\left(\pi(A) + \frac{\mathbf{Q}_T(A, A^c)}{T}\right) + (1-2T) f\left(\pi(A) - \frac{\mathbf{Q}_T(A, A^c) - \mathbf{Q}_T(A^c, A)}{1-2T}\right) \\ &\quad + T f\left(\pi(A) - \frac{\mathbf{Q}_T(A^c, A)}{T}\right) \end{aligned}$$

This is decreasing in $\mathbf{Q}_T(A, A^c)$ when $\mathbf{Q}_T(A, A^c) \geq \mathbf{Q}_T(A^c, A)$, and decreasing in $\mathbf{Q}_T(A^c, A)$ when $\mathbf{Q}_T(A^c, A) \geq \mathbf{Q}_T(A, A^c)$, and so is maximized when $\mathbf{Q}_T(A, A^c) = \mathbf{Q}_T(A^c, A) = \hat{\Phi}_t(A)\pi(A)\pi(A^c)$. It

follows that

$$\begin{aligned}
1 - \psi(A) &\leq \frac{T f\left(\pi(A) + \frac{\hat{\Phi}_t(A)\pi(A)\pi(A^c)}{T}\right) + (1 - 2T) f(\pi(A)) + T f\left(\pi(A) - \frac{\hat{\Phi}_t(A)\pi(A)\pi(A^c)}{T}\right)}{f(\pi(A))} \\
&= 2T \sqrt{\left(\frac{1}{2} + \frac{\hat{\Phi}_T(A)}{2T}\pi(A^c)\right)\left(\frac{1}{2} - \frac{\hat{\Phi}_T(A)}{2T}\pi(A)\right)} + (1 - 2T) \\
&\quad + 2T \sqrt{\left(\frac{1}{2} - \frac{\hat{\Phi}_T(A)}{2T}\pi(A^c)\right)\left(\frac{1}{2} + \frac{\hat{\Phi}_T(A)}{2T}\pi(A)\right)} \\
&\leq 2T \sqrt{1 - (\hat{\Phi}_T(A)/2T)^2} + 1 - 2T \leq 1 - \hat{\Phi}_T(A)^2/4T
\end{aligned}$$

The final line was by the relation $\sqrt{XY} + \sqrt{(1-X)(1-Y)} \leq \sqrt{1-(X-Y)^2}$; see Lemma A.1 in the Appendix for a proof. If $T = \min\{\alpha, t\} = t$ then the theorem is immediate. If $T = \min\{\alpha, t\} = \alpha$ then use the relation $\hat{\Phi}_T(A) \geq \frac{\alpha}{t} \hat{\Phi}_t(A)$. \square

4 New Canonical path bounds

When ergodic flow leaving a set A is not heavily concentrated at a few vertices then use of threshold conductance methods can improve substantially over use of conductance. In this section we explore such a situation by considering the case when canonical paths are well distributed among the vertices, and thus the ergodic flow appearing on edges of canonical paths is not heavily concentrated at any vertex.

What is most novel about our approach is that it applies to general finite Markov chains, i.e. non-reversible and non-lazy, a result made possible by using the paths to bound intersection threshold conductance $\hat{\Phi}_t(A)$. Ordinary canonical paths cannot be used in this setting as there are walks which do not converge but for which it is easy to construct canonical paths, such as the simple random walk on a cycle with an even number of vertices (i.e. $P(i, i-1 \bmod n) = P(i, i+1 \bmod n) = 1/2$) and the clockwise walk on a cycle of odd length (i.e. $P(i, i+1 \bmod n) = 1$). However, if path length is required to have some *parity* (all even or all odd) then the simple walk on an even cycle no longer has natural canonical paths, while a requirement that the path *alternate* between edges of P and P^* likewise prevents the clockwise walk from having canonical paths. In fact, these two conditions will be sufficient.

Definition 4.1. A canonical alternating path γ_{xy} is an even length path from x to y alternating between valid transitions of P and P^* :

$$x = x_0 \xrightarrow{P} x_1 \xrightarrow{P^*} x_2 \xrightarrow{P} x_3 \xrightarrow{P^*} \cdots \xrightarrow{P^*} x_{2n} = y$$

Equivalently, the path alternates between forward and reversed edges of P :

$$x = x_0 \rightarrow x_1 \leftarrow x_2 \rightarrow x_3 \leftarrow \cdots \rightarrow x_{2n-1} \leftarrow x_{2n} = y$$

Let $\Gamma = \{\gamma_{xy} : x, y \in V\}$ be a set including a canonical alternating path for each ordered pair of distinct vertices $x, y \in V$.

Define $v \in \gamma_{xy}$ if $v = x_{2i+1}$ for some i , i.e. v is the terminal point of an edge in γ_{xy} .

Define $(u, v) \in \gamma_{xy}$ if $u \rightarrow v$ or $v \leftarrow u$ appears in path γ_{xy} , i.e. (u, v) is an edge of \mathbf{P} or (v, u) an edge of \mathbf{P}^* in the path.

The definitions of $v \in \gamma_{xy}$ and $(u, v) \in \gamma_{xy}$ can be relaxed to allow both initial and terminal endpoints of edges, but this leads to weaker results.

As discussed in the Introduction, different canonical path methods require different types of paths. However, each of these types of paths induce natural canonical alternating paths, as shown below, so our new definition provides a unifying framework:

- For a finite reversible walk with holding probability $\alpha > 0$ Diaconis and Strook [4] and Sinclair [12] use simple paths:

$$x = x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_n = y$$

Such paths can be used to study non-reversible walks as well [10]. In either case, to construct an alternating path insert a self-loop at each vertex after the initial one and consider the loop as a transition of \mathbf{P}^* since $\mathbf{P}^*(v, v) = \mathbf{P}(v, v)$:

$$x = x_0 \rightarrow x_1^{\circlearrowleft} \rightarrow x_2^{\circlearrowleft} \rightarrow \cdots \rightarrow x_n^{\circlearrowleft} = y$$

- For a general finite reversible walk Diaconis and Strook [4] use paths in which γ_{xy} is constructed for $x \neq y$ with no parity requirement, but with looping paths γ_{xx} which must be odd length. If the length $|\gamma_{xy}|$ is even then it is also an alternating path since $\mathbf{P}^* = \mathbf{P}$ for reversible walks, while if $|\gamma_{xy}|$ is odd then γ_{xy} followed by γ_{yy} will be an alternating path.
- For a general finite walk Mihail [9] and Fill [5] show that it suffices to study conductance of \mathbf{PP}^* or canonical paths with edges in \mathbf{PP}^* . A path γ_{xy} with edges from \mathbf{PP}^* can be made into an alternating path by replacing each edge $\mathbf{PP}^*(x, y) > 0$ with a pair of edges $x \xrightarrow{\mathbf{P}} z \xrightarrow{\mathbf{P}^*} y$ such that $\mathbf{P}(x, z) > 0$ and $\mathbf{P}^*(z, y) > 0$.
- In an earlier version of this paper we required odd length paths alternating between \mathbf{P} and \mathbf{P}^* , including looping paths γ_{xx} . Given $x \neq y$ the path γ_{xy} followed by γ_{yy} in reverse will be an alternating path.

Notions of both edge and vertex congestion will be required in our results:

Definition 4.2. Given (ordinary or alternating) canonical paths Γ between every pair of distinct vertices $x, y \in V$ the vertex congestion is

$$\rho_v = \rho_v(\Gamma) = \max_{v \in V} \frac{1}{2\pi(v)} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni v} \pi(x)\pi(y)$$

and the edge congestion is

$$\rho_e = \rho_e(\Gamma) = \max_{e=(u,v) \in E} \frac{1}{2\pi(u)\mathbf{P}(u,v)} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni (u,v)} \pi(x)\pi(y)$$

The redefined edge congestion is no larger than the earlier definition and is equivalent to it when, as is typically the case, the path γ_{xy} does not use any of the same (directed) edges as γ_{yx} , e.g. in the reversible case γ_{xy} can be assumed to traverse the edges of γ_{yx} in reverse. The conductance and mixing time bounds to be shown in this section are weaker for larger ρ_e , so using the earlier definition will give weaker but still valid results.

This brings us to the main result of the section, a lower bound on the intersection threshold conductance $\hat{\Phi}_t$ in terms of canonical alternating paths.

Lemma 4.3. *Given canonical alternating paths Γ then*

$$\hat{\Phi}_{\rho_v/\rho_e} \geq 1/2\rho_e$$

If, instead, Γ consists of ordinary canonical paths then

$$\hat{\Phi}_{\rho_v/\rho_e} \geq 1/\rho_e$$

The proof is given later. Combining this with our results on threshold conductances leads to new upper bounds on mixing time:

Theorem 4.4. *Consider a finite ergodic Markov kernel \mathbf{P} with canonical ordinary or alternating paths. Assume $\epsilon \leq 1$. Let*

$$\mathbf{P}_0^*(\Gamma) = \min \{ \mathbf{P}^*(b, a) : \exists \gamma_{xy} \ni (a, b) \}$$

be the smallest transition in the reversal $\mathbf{P}^(b, a) = \frac{\pi(a)\mathbf{P}(a,b)}{\pi(b)}$ of the edges in the paths.*

If Γ consists of canonical alternating paths then \mathbf{P} has mixing time

$$\tau(\epsilon) \leq 48\rho_e \max \left\{ \rho_v, \frac{1}{2} \right\} \left(\log \frac{1}{\epsilon\sqrt{\pi_0}} - \frac{1}{4} (\log(4\rho_v\rho_e\mathbf{P}_0^*(\Gamma)) - 1)^+ \right)$$

where $x^+ = \max\{x, 0\}$.

If, instead, Γ consists of ordinary canonical paths then

$$\tau(\epsilon) \leq 4\rho_v \max \left\{ \frac{\rho_v}{\alpha}, \rho_e \right\} \left(\log \frac{1}{\epsilon\sqrt{\pi_0}} - \frac{1}{4} (\log(\rho_v\rho_e\mathbf{P}_0^*(\Gamma)) - 1)^+ \right)$$

These can be simplified by using the relations:

$$\rho_v \leq \rho_e \leq \frac{\rho_v}{\mathbf{P}_0^*(\Gamma)} < \frac{1}{2\pi_0\mathbf{P}_0^*(\Gamma)} \quad (4.4)$$

With ordinary paths a lower bound of $\rho_v \geq 1 - \min \pi(v)$ also holds, while alternating canonical paths have the weaker $\rho_v \geq \frac{1}{2}(1 - \max \pi(v))$. Although the mixing bounds of the theorem are not monotone, they do hold for any upper bounds on ρ_v and ρ_e which satisfy the constraints just mentioned.

The $\max \left\{ \frac{\rho_v}{\alpha}, \rho_e \right\}$ term is just an upper bound on the edge congestion when alternating paths are constructed by adding self-loops at all but the initial vertices of the ordinary paths. This maximum and the correction term with $\log(\rho_v\rho_e\mathbf{P}_0^*(\Gamma))$ are not simply artifacts of the proof. A simple illustration of this is the walk on a cycle \mathbb{Z}_n with $\mathbf{P}(i, i) = \alpha \in (0, 1)$ and $\mathbf{P}(i, i+1) = 1 - \alpha$. With the obvious choice of paths this has $\rho_v = \frac{n-1}{2}$, $\rho_e = \frac{n-1}{2(1-\alpha)}$, $\mathbf{P}_0^*(\Gamma) = 1 - \alpha$, $\pi_0 = 1/n$ and mixing time $\tau(\epsilon) = \Theta \left(\frac{n^2}{\min\{\alpha, 1-\alpha\}} \log \frac{1}{\epsilon} \right)$. In contrast, the lazy walk on the complete graph K_n has mixing time $\tau(1/4) = \Omega(\log n)$ and so the π_0 term is sometimes necessary.

Proof. First consider ordinary canonical paths. The theorem is trivial if $|V| = 1$, so assume that $|V| \geq 2$ and $0 < \min \pi(v) \leq 1/2$. Lemma 4.3 and Theorem 3.2 bound the root profile as

$$\psi(A) \geq \frac{\min\{\alpha, \rho_v/\rho_e\}}{4(\rho_v/\rho_e)^2} \frac{1}{\rho_e^2} = \frac{1}{4\rho_v \max\{\rho_v/\alpha, \rho_e\}}$$

Evaluating the integral in the convex case of Theorem 2.2 gives the result when $\rho_v \rho_e P_0^*(\Gamma) \leq e$, so assume that $\rho_v \rho_e P_0^*(\Gamma) > e$. Since the walk is ergodic then the state space is strongly connected, and so $Q(A, A^c) > 0$ for every set $A \notin \{\emptyset, V\}$. In particular, $Q_{P_0^*(\Gamma)}(A, A^c) \geq \pi_0 P_0^*(\Gamma)$, and similarly $Q_{P_0^*(\Gamma)}(A^c, A) \geq \pi_0 P_0^*(\Gamma)$, and so $\hat{\Phi}_{P_0^*(\Gamma)}(A) \geq \frac{\pi_0 P_0^*(\Gamma)}{\pi(A)\pi(A^c)}$. By Theorem 3.2

$$\psi(A) \geq \max \left\{ \frac{1}{4\rho_v \max\{\rho_v/\alpha, \rho_e\}}, \frac{\pi_0^2 \min\{\alpha, P_0^*(\Gamma)\}}{4\pi(A)^2} \right\}$$

The convexity condition of Theorem 2.2 is easily verified for this lower bound and so

$$\tau(\epsilon) \leq \int_0^c \frac{2r \, dr}{\pi_0^2 \min\{\alpha, P_0^*(\Gamma)\}} + 2\rho_v \max \left\{ \frac{\rho_v}{\alpha}, \rho_e \right\} \int_c^{1/\epsilon^2} \frac{dr}{r}$$

if $c = \pi_0 \sqrt{\min\{\alpha, P_0^*(\Gamma)\} \rho_v \max\{\rho_v/\alpha, \rho_e\}} \leq 1/\epsilon^2$. The theorem follows by integrating and using equations (4.4), $\rho_v \rho_e P_0^*(\Gamma) \geq 1$ and $1/2 \leq \rho_v < 1/2\pi_0$.

For alternating canonical paths $\psi(A) \geq (48 \max\{\rho_v, 1/2\} \rho_e)^{-1}$ by Lemma 4.3 and Theorem 3.2. Theorem 2.2 shows the mixing bound when $4\rho_v \rho_e P_0^*(\Gamma) \leq e$. To improve $\psi(A)$ use $\hat{\Phi}_{P_0^*(\Gamma)} \geq \frac{\pi_0 P_0^*(\Gamma)}{\pi(A)}$ because there is at least one alternating path between some $x \in A$ and $y \in A^c$, and this path will contain some vertex $v \in \gamma_{xy}$ with incoming edges from both A and A^c . Again use Theorem 3.2 to lower bound $\psi(A)$, and this time split the mixing time integral of Theorem 2.2 at $c = 2\pi_0 \sqrt{\rho_v \rho_e P_0^*(\Gamma)}$. \square

We return now to the proof of our main result:

Proof of Lemma 4.3. Consider either the ordinary canonical path or alternating path case. The stationary distribution and edge capacity can be lower bounded using paths:

$$\begin{aligned} \forall v \in V : \quad \pi(v) &\geq \frac{1}{2\rho_v} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni v} \pi(x)\pi(y) \\ &\geq \frac{1}{2\rho_v} \sum_{\substack{(x,y): \exists u \in A, \\ (u,v) \in \gamma_{xy} \cup \gamma_{yx}}} \pi(x)\pi(y) \\ \forall u \in A, v \in V : \quad \pi(u)P(u,v) &\geq \frac{1}{2\rho_e} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni (u,v)} \pi(x)\pi(y). \end{aligned}$$

Hence, for any $v \in V$:

$$\begin{aligned}
Q_{\rho_v/\rho_e}(A, v) &= \min \left\{ \sum_{u \in A} \pi(u) P(u, v), \frac{\rho_v}{\rho_e} \pi(v) \right\} \\
&\geq \frac{1}{2\rho_e} \min \left\{ \sum_{u \in A} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni (u, v)} \pi(x)\pi(y), \sum_{\substack{(x, y): \exists u \in A, \\ (u, v) \in \gamma_{xy} \cup \gamma_{yx}}} \pi(x)\pi(y) \right\} \\
&= \frac{1}{2\rho_e} \sum_{\substack{(x, y): \exists u \in A, \\ (u, v) \in \gamma_{xy} \cup \gamma_{yx}}} \pi(x)\pi(y) \tag{4.5}
\end{aligned}$$

In the ordinary canonical path case it follows from (4.5) that

$$\begin{aligned}
Q_{\rho_v/\rho_e}(A, A^c) &= \sum_{v \in A^c} Q_{\rho_v/\rho_e}(A, v) \\
&\geq \sum_{v \in A^c} \frac{1}{2\rho_e} \sum_{\substack{(x, y): \exists u \in A, \\ (u, v) \in \gamma_{xy} \cup \gamma_{yx}}} \pi(x)\pi(y) \\
&\geq \frac{1}{2\rho_e} \sum_{(x, y) \in A \times A^c} 2\pi(x)\pi(y) = \frac{\pi(A)\pi(A^c)}{\rho_e}
\end{aligned}$$

The second inequality is because a path from some $x_0 \in A$ to $x_n \in A^c$ must have some $x_i \in A$ and $x_{i+1} \in A^c$. Taking $A \leftarrow A^c$ shows that $Q_{\rho_v/\rho_e}(A^c, A) \geq \frac{\pi(A)\pi(A^c)}{\rho_e}$ as well. It follows that $\hat{\Phi}_{\rho_v/\rho_e}(A) \geq 1/\rho_e$.

If the paths are alternating then (4.5) shows that

$$\begin{aligned}
&\sum_{v \in V} \min\{Q_{\rho_v/\rho_e}(A, v), Q_{\rho_v/\rho_e}(A^c, v)\} \\
&\geq \sum_{v \in V} \frac{1}{2\rho_e} \sum_{\substack{(x, y): \exists u \in A, w \in A^c: \\ (u, v) \in \gamma_{xy} \cup \gamma_{yx}, \\ (w, v) \in \gamma_{xy} \cup \gamma_{yx}}} \pi(x)\pi(y) \\
&\geq \frac{1}{2\rho_e} \sum_{(x, y) \in A \times A^c} 2\pi(x)\pi(y) \geq \frac{\pi(A)\pi(A^c)}{\rho_e}
\end{aligned}$$

The second inequality is because an (even length) alternating path from $x_0 \in A$ to $x_{2n} \in A^c$ must have some $x_{2i} \in A$ and $x_{2(i+1)} \in A^c$, and so is counted when $v = x_{2i+1}$. It follows that $\hat{\Phi}_{\rho_v/\rho_e}(A) \geq 1/2\rho_e$. \square

Remark 4.5. Ordinary canonical paths can be applied to three types of mixing bounds for lazy walks, with lead term ρ_e^2 for Jerrum and Sinclair's result, $\rho_e \ell$ for the Poincaré bound, and $\rho_e \rho_v$ for ours. Since $\rho_v \leq \rho_e$ then our bound improves on Jerrum and Sinclair's. To compare to the

Poincaré bound define average vertex congestion and average path length by

$$\bar{\rho}_v = \sum_{v \in V} \pi(v) \left[\frac{1}{2\pi(v)} \sum_{\gamma_{xy} \cup \gamma_{yx} \ni v} \pi(x) \pi(y) \right] \quad \text{and} \quad \bar{\ell} = \frac{\sum_{x \neq y} \pi(x) \pi(y) |\gamma_{xy}|}{\sum_{x \neq y} \pi(x) \pi(y)}$$

respectively. Then

$$\begin{aligned} \bar{\rho}_v &= \frac{1}{2} \sum_{x \neq y} \sum_{v \in \gamma_{xy} \cup \gamma_{yx}} \pi(x) \pi(y) \leq \sum_{x \neq y} \pi(x) \pi(y) |\gamma_{xy}| \\ &= \bar{\ell} \sum_{x \neq y} \pi(x) \pi(y) = \bar{\ell} (1 - \|\pi\|_2^2). \end{aligned}$$

Likewise $\bar{\rho}_v \geq \frac{\bar{\ell}+1}{2}(1 - \|\pi\|_2^2)$, so $\bar{\rho}_v = \Theta(\bar{\ell}(1 - \|\pi\|_2^2))$. When a few paths are very long or the distribution is concentrated near a single vertex then it is likely that $\rho_v \ll \ell$ and our bound is best. However, it is more often the case that path length varies little and a bottleneck causes a few states to have high vertex congestion, and so $\ell \ll \rho_v$. In contrast to the lazy case, when the holding probability is small then our result can be significantly better, even when $\ell \ll \rho_v$; see the next section for examples.

5 Examples

To demonstrate our method we give two examples where the new canonical path theorems extend previously known bounds into the general non-reversible non-lazy case. First, the classical problem of the max-degree walk on a graph, then the more interesting case of walks on Cayley graphs, i.e. random walks on groups.

Example 5.1. An Eulerian multigraph is a strongly connected graph with in-degree=out-degree at each vertex, a natural generalization of the undirected multigraph into the directed graph setting. Suppose an Eulerian multigraph has n vertices and maximum out-degree d . Let $d(x, y)$ denote the number of directed edges from x to y , so that $d(x) = \sum_y d(x, y)$ is the out-degree of x . The max-degree walk has $\mathbf{P}(x, y) = \frac{d(x, y)}{d}$ if $y \neq x$ and $\mathbf{P}(x, x) = 1 - \frac{d(x) - d(x, x)}{d}$. The stationary distribution $\pi = 1/n$ is uniform and the walk is lazy if $d(x) - d(x, x) \leq d/2$ at each vertex, i.e. there are many self-loops.

Suppose that every vertex has a self-loop. Then holding probability $\alpha \geq 1/d$, which is sufficient to prevent significant periodicity effects. For every $x \neq y$ let γ_{xy} be a path from x to y . The congestions are at worst

$$\rho_v \leq \frac{\sum_{x \neq y} \pi(x) \pi(y)}{2\pi_0} \leq \frac{n}{2} \quad \text{and} \quad \rho_e \leq \frac{\sum_{x \neq y} \pi(x) \pi(y)}{2\pi_0 \mathbf{P}_0(\Gamma)} \leq \frac{dn}{2}$$

and so by Theorem 4.4

$$\tau(\epsilon) \leq dn^2 \log \frac{2}{\epsilon}$$

Without the self-loop requirement, if the graph is connected under canonical alternating paths then it is still true that $\rho_v \leq \frac{n}{2}$ and $\rho_e \leq \frac{nd}{2}$ and so by Theorem 4.4

$$\tau(\epsilon) \leq 12 dn^2 \log \frac{2}{\epsilon}.$$

Recall that strong connectivity is not enough for ergodicity, e.g. the cycle walk $P(i, i+1 \pmod n) = 1$ on \mathbb{Z}_n .

In contrast, the Poincaré method introduces an extra factor of α^{-1} , and so it can match the self-looping case only for a walk with constant (in n and d) holding probability. Diaconis and Strook's extension to non-lazy walks works only for the reversible case (i.e. $d(x, y) = d(y, x)$ for every x, y). Mihail and Fill's extension with PP^* replaces the $(\min_{P(x, y) > 0} P(x, y))^{-1}$ term in ρ_e with $(\min_{PP^*(x, y) > 0} PP^*(x, y))^{-1}$, which typically replaces the order d term with order d^2 .

Example 5.2. The Cayley graph of a group G with (non-symmetric) generating set $S \subset G - \{id\}$, i.e. $\bigcup_{n=0}^{\infty} S^n = G$, has edge set (g, gs) for all $g \in G, s \in S$. If $p : G \rightarrow [0, 1]$ is a probability distribution supported on $S \cup \{id\}$ then $P(g, gs) = p(s)$ defines a Markov chain with uniform stationary distribution $\pi = 1/|G|$. Represent each $g \in G$ as a product of generators $g = s_1 s_2 \cdots s_k$, define $\Delta = \max |g|$ to be the length of the longest such representation, and let $N(g, s) \leq \Delta$ denote the number of times generator s appears in the representation of g .

Babai [1] showed $\tau(\epsilon) = O(\frac{\Delta^2}{\min_{s \in S} p(s)} \log \frac{|G|}{\epsilon})$ for the lazy walk with symmetric generating set, i.e. $p(id) \geq 1/2$ with $S = S^{-1}$ and $\forall s \in S : p(s) = p(s^{-1})$. Diaconis and Saloff-Coste [3] use (ordinary) canonical paths to bound the spectral gap by the Poincaré approach. This can be plugged into spectral gap bounds on mixing time (e.g. Corollary 2.15 of [10]), leading to the following generalizations of Babai's result to the symmetric and non-symmetric cases respectively:

$$\begin{aligned} \tau(\epsilon) &\leq \max \left\{ \frac{1}{2p(id)}, \Delta \max_{g \in G, s \in S} \frac{N(g, s)}{p(s)} \right\} \left(\frac{1}{2} \log |G| + \log \frac{1}{\epsilon} \right) \\ \tau(\epsilon) &\leq \frac{\Delta}{p(id)} \max_{g \in G, s \in S} \frac{N(g, s)}{p(s)} \left(\frac{1}{2} \log |G| + \log \frac{1}{\epsilon} \right) \end{aligned} \quad (5.6)$$

Consider now our new method of canonical alternating paths. Let Δ_{alt} be the diameter measured using canonical alternating paths, i.e. $\Delta_{alt} = 2 \min \{N : G = \bigcup_{n=0}^N (SS^{-1})^n\}$, and let $N_{alt}(g, s)$ count frequency of generators when group elements are written in terms of alternating canonical paths, e.g. $g = s_1 s_2^{-1} s_3 s_4^{-1} \cdots s_{2n-1} s_{2n}^{-1}$. Congestions for canonical alternating paths lead to the bounds

$$\rho_v < \frac{\Delta_{alt}}{4}, \quad \rho_e < \frac{1}{2} \max_{g \in G, s \in S} \frac{N_{alt}(g, s)}{p(s)}$$

The proof is left to the Appendix as it uses essentially the same approach used for ordinary canonical paths in [3]. Theorem 4.4 leads to a mixing bound of

$$\tau(\epsilon) \leq 6 \Delta_{alt} \max_{g \in G, s \in S} \frac{N_{alt}(g, s)}{p(s)} \left(\frac{1}{2} \log \frac{4|G|}{\Delta_{alt}} + \log \frac{1}{\epsilon} \right) \quad (5.7)$$

To see that this generalizes Diaconis and Saloff-Coste's non-symmetric result recall the construction of alternating paths from ordinary canonical paths by adding a self-loop at each vertex except the starting point. Such a set of alternating paths will have $\Delta_{alt} = 2\Delta$, while the self-loops cause

$$\max_{g \in G, s \in S} \frac{N_{alt}(g, s)}{p(s)} = \max \left\{ \frac{\Delta}{p(id)}, \max_{g \in G, s \in S} \frac{N(g, s)}{p(s)} \right\} \leq \frac{1}{p(id)} \max_{g \in G, s \in S} \frac{N(g, s)}{p(s)}$$

The inequality was because $\Delta = \max_{g \in G} \sum_{s \in S} p(s) \frac{N(g, s)}{p(s)} \leq \max_{g \in G} \max_{s \in S} \frac{N(g, s)}{p(s)}$.

It follows that, up to a constant, if ordinary canonical paths are replaced by canonical alternating paths then past results for walks on Cayley graphs hold even when there is no holding probability.

References

- [1] L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing (STOC 1991)*, pages 164–174, 1991.
- [2] P. Diaconis and J. Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18(4):1483–1522, 1990.
- [3] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, 21(4):2131–2156, 1993.
- [4] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1:36–61, 1991.
- [5] J. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, 1(1):62–87, 1991.
- [6] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved. *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC 1988)*, pages 235–243, 1988.
- [7] R. Kannan, L. Lovász, and R. Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability and Computing*, 15(4):541–570, 2006.
- [8] G. Lawler and A. Sokal. Bounds on the l^2 spectrum for markov chains and markov processes: a generalization of cheeger’s inequality. *Transactions of the American Mathematical Society*, 309:557–580, 1988.
- [9] M. Mihail. Conductance and convergence of markov chains—a combinatorial treatment of expanders. *30th Annual Symposium on Foundations of Computer Science*, pages 526–531, 1989.
- [10] R. Montenegro and P. Tetali. *Mathematical Aspects of Mixing Times in Markov Chains*, volume 1:3 of *Foundations and Trends in Theoretical Computer Science*. NOW Publishers, Boston-Delft, June 2006.
- [11] B. Morris and Y. Peres. Evolving sets, mixing and heat kernel bounds. *Probability Theory and Related Fields*, 133(2):245–266, 2005.
- [12] A. Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

A Appendix

The following inequality was used in the proof of Theorem 3.2.

Lemma A.1. *If $X, Y \in [0, 1]$ then*

$$g(X, Y) = \sqrt{XY} + \sqrt{(1-X)(1-Y)} \leq \sqrt{1 - (X-Y)^2}.$$

Proof. Observe that

$$g(X, Y)^2 = 1 - (X+Y) + 2XY + \sqrt{[1 - (X+Y) + 2XY]^2 - [1 - 2(X+Y) + (X+Y)^2]}.$$

Now, $\sqrt{A^2 - B} \leq A - B$ if $A^2 \geq B$, $A \leq \frac{1+B}{2}$ and $A \geq B$ (square both sides to show this). These conditions are easily verified with $A = 1 - (X+Y) + 2XY$ and $B = 1 - 2(X+Y) + (X+Y)^2$, and so

$$\begin{aligned} g(X, Y)^2 &\leq 2[1 - (X+Y) + 2XY] - [1 - 2(X+Y) + (X+Y)^2] \\ &= 1 + 2XY - X^2 - Y^2 = 1 - (X-Y)^2 \end{aligned}$$

□

In order to study walks on Cayley graphs it was claimed that congestion bounds of Diaconis and Saloff-Coste [3] generalize easily. We show this here.

Lemma A.2. *Consider group G with (non-symmetric) generating set S . Use the notation of Example 5.2 to describe a walk on the Cayley graph of $G = \langle S \rangle$.*

There are ordinary canonical paths with

$$\rho_v < \Delta, \quad \rho_e < \max_{g \in G, s \in S} \frac{N(g, s)}{p(s)}.$$

If p is symmetric, i.e. $\forall s \in S : p(s) = p(s^{-1})$, then $\rho_v < \frac{\Delta+1}{2}$.

There are canonical alternating paths with

$$\rho_v < \frac{\Delta_{alt}}{4}, \quad \rho_e < \frac{1}{2} \max_{g \in G, s \in S} \frac{N_{alt}(g, s)}{p(s)}$$

and $N_{alt}(g, s) \leq \frac{1}{2} \Delta_{alt} < |G|$ for every $g \in G, s \in S$.

Proof. First consider ordinary canonical paths. Given $x, y \in G$ let $g = x^{-1}y = s_1 s_2 \cdots s_k$ and define path $\gamma_{x,y}$ by $x \rightarrow xs_1 \rightarrow \cdots \rightarrow xg = y$. Recall that $\pi = 1/|G|$ is uniform.

To bound vertex-congestion observe that the same number of paths pass through each vertex, because if $\gamma_{x,y}$ includes vertex v then $\gamma_{wv^{-1}x, wv^{-1}y}$ includes vertex w , and vice-versa. Then

$$\begin{aligned} \rho_v = \bar{\rho}_v &\leq \frac{1}{|G|} \sum_{g \in G} \frac{1}{2\pi(g)} \sum_{(x,y): g \in \gamma_{xy} \cup \gamma_{yx}} \pi(x)\pi(y) \\ &\leq \frac{1}{2} \sum_{x \neq y} \pi(x)\pi(y) (|\gamma_{xy}| + |\gamma_{yx}|) \\ &\leq \Delta \left(1 - \sum_{g \in G} \pi(g)^2 \right) = \Delta \left(1 - \frac{1}{|G|} \right) \end{aligned} \tag{A.8}$$

When p is symmetric then assume the representation for g^{-1} to be the inverse of that for g , i.e. if $g = s_1 s_2 \cdots s_k$ then $g^{-1} = s_k^{-1} s_{k-1}^{-1} \cdots s_1^{-1}$. This does not increase Δ so it can only improve the bound on ρ_v . If $g \in \gamma_{xy}$ then $g \in \gamma_{yx}$, and vice-versa, so (A.8) improves to $\frac{1}{2} \sum_{x \neq y} \pi(x)\pi(y) (|\gamma_{xy}| + 1)$ and $\rho_v \leq \frac{\Delta+1}{2} \left(1 - \frac{1}{|G|}\right)$.

Now consider edge-congestion. Without loss assume that id does not appear in any paths. If $\gamma_{x,y}$ includes edge (v, vs) then $\gamma_{wv^{-1}x, wv^{-1}y}$ includes edge (w, ws) , and vice-versa, and so for fixed $s \in S$ the number of paths through edge (g, gs) is independent of the choice of $g \in G$. Hence,

$$\begin{aligned} \rho_e &\leq \max_{s \in S} \frac{1}{|G|} \sum_{g \in G} \frac{1}{2\pi(g)\mathbb{P}(g, gs)} \sum_{(x,y): (g,gs) \in \gamma_{xy} \cup \gamma_{yx}} \pi(x)\pi(y) \\ &\leq \max_{s \in S} \frac{1}{2p(s)} 2 \sum_{x \neq y} N(x^{-1}y, s) \pi(x)\pi(y) \\ &\leq \max_{s \in S} \frac{1}{p(s)} \max_{g \in G} N(g, s) \left(1 - \frac{1}{|G|}\right) \end{aligned}$$

Finally, when alternating canonical paths are used then again assume the representation of each g^{-1} to be the inverse of that for g , so that $g \in \gamma_{xy} \Leftrightarrow g \in \gamma_{yx}$ and $(g, h) \in \gamma_{xy} \Leftrightarrow (g, h) \in \gamma_{yx}$. Then, arguing as before,

$$\begin{aligned} \rho_v = \bar{\rho}_v &= \frac{1}{|G|} \sum_{g \in G} \frac{1}{2\pi(g)} \sum_{(x,y): g \in \gamma_{xy} \cup \gamma_{yx}} \pi(x)\pi(y) \\ &= \frac{1}{2} \sum_{x \neq y} \pi(x)\pi(y) \frac{|\gamma_{xy}|}{2} \\ &\leq \frac{\Delta_{alt}}{4} \left(1 - \sum_{g \in G} \pi(g)^2\right) = \frac{\Delta_{alt}}{4} \left(1 - \frac{1}{|G|}\right) \end{aligned}$$

Similarly minor changes show that $\rho_e < \frac{1}{2} \max_{s \in S} \frac{1}{p(s)} \max_{g \in G} N_{alt}(g, s)$.

For the final statement, in the representation $g = s_1 s_2^{-1} \cdots s_{2k-1} s_{2k}^{-1}$ remove all even length subcycles, reducing the problem to the case where there are no even length subcycles. In particular, if $i, j \leq k$ then $s_1 s_2^{-1} \cdots s_{2i-1} s_{2i}^{-1} = s_1 s_2^{-1} \cdots s_{2j-1} s_{2j}^{-1} \Rightarrow i = j$. This guarantees that $\{id, s_1 s_2^{-1}, \dots, s_1 s_2^{-1} \cdots s_{2k-1} s_{2k}^{-1}\}$ is a set of $k+1$ distinct elements, and so $k+1 \leq |G|$, while it also guarantees that $s s^{-1}$ never appears and so also $N_{alt}(g, s) \leq |g|/2 \leq |G| - 1$. \square