Combinatorics, Probability and Computing (2014) 23, 585–606.
doi:10.1017/S096354831400025X

Intersection Conductance and Canonical Alternating Paths: Methods for General Finite Markov Chains

RAVI MONTENEGRO

Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA
(e-mail: ravi_montenegro@uml.edu)

Received 23 January 2008; revised 20 March 2014

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 the mixing time of a lazy walk on a Cayley graph with a symmetric generating set also applies to the non-lazy non-symmetric case, often even when there is no holding probability.

2010 *Mathematics subject classification*: 60J10

1. Introduction

Beginning with the work of Jerrum and Sinclair, geometric concepts such as conductance [8, 10] and canonical paths [14, 4] have played an important role in studying the mixing time of finite ergodic Markov chains. These methods were originally applied only to reversible lazy walks, and while little is lost in dropping reversibility or laziness [4, 6, 5, 12, 7], extensions which allow for dropping both conditions tend to be weak or difficult to use [6, 11, 5, 12]. A similar difficulty has been encountered with the method of blocking conductance [9], 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, and sharpen it with a simpler notion of vertex congestion than those proposed before. 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 ($\exists N \in \mathbb{N}, \forall x, y \in V : \mathsf{P}^N(x, y) > 0$), then it has a unique stationary distribution π satisfying $\pi\mathsf{P} = \pi$ and is ergodic ($\forall x \in V : \mathsf{P}^n(x, \cdot) \xrightarrow{n\to\infty} \pi$). The walk is lazy if the holding probability $\alpha = \min_{v\in V} \mathsf{P}(v, v)$ is at least $1/2$ and reversible if the time-reversal (adjoint)

$$\mathsf{P}^*(x, y) = \frac{\pi(y)\mathsf{P}(y, x)}{\pi(x)}$$

satisfies $\mathsf{P}^* = \mathsf{P}$. The L^2 mixing time $\tau(\epsilon) = \max_{x\in V} \min\{n : \|k_n^x - 1\|_{2,\pi} \leqslant \epsilon\}$ is the time it takes standard deviation of the density

$$k_n^x(y) = \frac{\mathsf{P}^n(x,y)}{\pi(y)}$$

to drop to ϵ.

Jerrum and Sinclair [8] and Lawler and Sokal [10] showed that mixing time of a lazy reversible walk can be bounded in terms of the conductance Φ, also known as the Cheeger constant.

Definition 1.1. Let $\mathsf{Q}(A,B) = \sum_{x\in A,\, y\in B} \pi(x)\mathsf{P}(x,y)$ denote the ergodic flow from $A \subset V$ to $B \subset V$. The *conductance* is given by

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

Theorem 1.2. *The mixing time of a lazy reversible finite ergodic Markov chain satisfies*

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

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

Our extension to the non-lazy non-reversible case uses a modified form of conductance, denoted by $\hat{\Phi}(A)$, which agrees with the usual conductance, $\Phi(A)$, in the lazy case. If a walk is lazy and $v \in A$, then $\mathsf{Q}(A,v) \geqslant \mathsf{Q}(A^c,v)$, whereas if $v \in A^c$ then $\mathsf{Q}(A^c,v) \geqslant \mathsf{Q}(A,v)$, and so

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

In other words, the overlap, or intersection, of the flow from A to V and the flow from A^c to V is the same as the flow $A \leftrightarrow A^c$ between A and A^c. The following is then equivalent to the usual conductance when a walk is lazy.

Definition 1.3. The *intersection conductance* $\hat{\Phi}(A)$ of $A \subset V$ (and A^c) is given by

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

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

Intuitively, when flows 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.

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

Theorem 1.4. *The mixing time of a finite ergodic Markov chain satisfies*

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

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.*, we work with *threshold limited ergodic flow* $\mathsf{Q}_t(A,v) = \min\{\mathsf{Q}(A,v),\, t\pi(v)\}$ instead of $\mathsf{Q}(A,v)$. When the 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 profiles, we define the following.

Definition 1.5. Given $t > 0$, the *intersection threshold conductance* of $A \subset V$ is

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

The *intersection threshold conductance profile* $\hat{\Phi}_t(r)$ is a function of set size, and given by

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

The *intersection threshold conductance* is $\hat{\Phi}_t = \min\limits_{\emptyset \subsetneq A \subsetneq V} \hat{\Phi}_t(A)$.

Note that, trivially, $\hat{\Phi}_t(A) = \hat{\Phi}_t(A^c)$, and so $\hat{\Phi}_t(1/2) = \hat{\Phi}_t$. Also, if $A \notin \{\emptyset, V\}$ then $\hat{\Phi}_t(A) = \hat{\Phi}(A) \leqslant 1$ for $t \geqslant 1/2$, and so for a lazy walk we have $\hat{\Phi}_{1/2}(A) = \Phi(A)$.

When $\hat{\Phi}_t(A) \approx \hat{\Phi}(A)$ then Theorem 1.4 can be sharpened by showing a mixing time bound that is typically t^{-1} times smaller.

Theorem 1.6. *Given threshold $t > 0$, the mixing time of a finite ergodic Markov chain satisfies*

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

More generally,

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

Many methods have been developed to lower-bound conductance. The most prominent of these is the method of canonical paths, introduced by Jerrum and Sinclair in their seminal paper on conductance [8].

Definition 1.7. A *canonical path* γ_{xy} is a path from x to y using only transitions $a \to b$ of P with non-zero probability $\mathsf{P}(a,b) > 0$:

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

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

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

Theorem 1.8 (Jerrum and Sinclair). *The mixing time of a lazy reversible finite ergodic Markov chain satisfies*

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

Sinclair later showed a Poincaré-type bound with 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 [14]. Diaconis and Stroock obtained an analogous bound for non-lazy reversible walks by including in Γ odd-length paths γ_{xx} from each vertex to itself [4]. Various authors have observed that these bounds apply to lazy non-reversible walks as well, and to non-lazy walks at the cost of a factor of $\alpha^{-1} = 1/\min_{v\in V} \mathsf{P}(v,v)$ [6, 11, 12].

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

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

This is equivalent to a path from x to y alternating between edges of P and its time-reversal 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 $x_1, x_2, \ldots, x_n$ 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

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

A mixing time result then follows from Theorem 1.6. With some extra work a bound is also possible if standard canonical paths are used.

Theorem 1.9. *Consider a finite ergodic Markov chain. Given a set of canonical alternating paths Γ, the mixing time is*

$$\tau(\epsilon) \leqslant 48\rho_e \max\left\{\rho_v, \frac{1}{2}\right\} \log \frac{1}{\epsilon\sqrt{\pi_0}}.$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

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

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

The congestions satisfy $\rho_v \leqslant \rho_e$, so this generalizes Jerrum and Sinclair's result, up to a constant factor. 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, we show that known complexity results for the lazy max-degree random walk on an undirected graph apply just as well to the max-degree walk on an Eulerian directed graph, *i.e.*, a strongly connected graph with in-degree equal to out-degree at each vertex. This is true if each vertex has a self-loop, and often true even when there are no 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 [12, 13] will be required for the proof of Theorem 1.6, 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 (*e.g.*, conductance) to a property of the original walk (*e.g.*, mixing time) is to construct a *dual process*: a walk P_D on $\Omega = \{S \subset V\}$ and a *link*, or transition matrix, Λ from Ω to V such that

$$\mathsf{P}\Lambda = \Lambda \mathsf{P}_D.$$

In particular, $\mathsf{P}^n\Lambda = \Lambda\mathsf{P}_D^n$ 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 [13] in a slightly improved form by Montenegro and Tetali [12].

To understand the method we require some new terminology.

Definition 2.1. Given a 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

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

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

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

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

Note that

$$\text{for all } r \geqslant 1/2, \quad \psi(r) = \psi(1/2) = \min_{A\notin\{\emptyset,V\}} \psi(A).$$

This follows from the relation $\psi(A) = \psi(A^c)$, a consequence of $(A_u)^c = (A^c)_{1-u}$ for every u with $\nexists v : \mathsf{Q}(A,v) = u\pi(v)$, *i.e.*, u-a.e. since V is finite.

Since $\mathsf{Q}(V,v) = \pi(v)$, 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

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

and so by Jensen's inequality

$$\mathbb{E}\sqrt{\pi(A_u)(1-\pi(A_u))} \leqslant \sqrt{\pi(A)(1-\pi(A))},$$

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 [13] first made this intuition rigorous, although we use a sharper result of Montenegro and Tetali [12].

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

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

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 $u \leqslant 1/2$ and $A_u \subset A$ when $u > 1/2$, and so

$$\begin{aligned} \mathsf{Q}(A,A^c) &= \sum_{v\in A^c} \frac{\mathsf{Q}(A,v)}{\pi(v)}\, \pi(v) \\ &= \sum_{v\in A^c} \int_0^{1/2} 1_{\{u:\, \mathsf{Q}(A,v)/\pi(v)\geqslant u\}}\, \pi(v)\, du \\ &= \int_0^{1/2} \pi(A_u \setminus A)\, du. \end{aligned}$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

The identities $\int_0^1 \pi(A_u)\,du = \pi(A)$ and $\mathsf{Q}(A,A^c) = \mathsf{Q}(A^c,A)$ can be used to write this in terms of area between A_u and A as

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

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 $\psi(A) \geqslant \Phi(A)^2/2$ [12]. Theorem 2.2 then shows a stronger version of Theorem 1.2.

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

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

To prove the result of this paper we will use Theorem 2.2 and a similar, but much more elaborate, approach to lower-bounding the root profile.

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 1.5. By Theorem 2.2 this induces upper bounds on mixing time, including the main result of this paper, Theorem 1.6, 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 the usual conductance.

Definition 3.1. If A, $B \subset V$ and $t \in [0,1]$, then define the *threshold flow* $\mathsf{Q}_t(A,B)$ by

$$\mathsf{Q}_t(A,B) = \sum_{v \in B} \mathsf{Q}_t(A,v)$$

and the *threshold conductance* $\Phi_t(A)$ by

$$\Phi_t(A) = \frac{\min\{\mathsf{Q}_t(A,A^c),\, \mathsf{Q}_t(A^c,A)\}}{\pi(A)\pi(A^c)}.$$

The *threshold conductance profile* is given by $\Phi_t(r) = \min_{\pi(A)\leqslant r} \Phi_t(A)$ while the *threshold conductance* is $\Phi_t = \min_{\emptyset \neq A \neq V} \Phi_t(A)$.

When a walk is lazy then $\mathsf{Q}_{1/2}(A,A^c) = \mathsf{Q}(A,A^c)$ and $\Phi_{1/2}(A) = \Phi(A)$ 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 overly concentrated on a few vertices then $\Phi_t(A) \approx \Phi(A)$, even for fairly small t. The extra information provided by t will be used here to substantially improve on the conductance lower bound for $\psi(A)$.

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

Theorem 3.2. *Given $A \subset V$, $t \in [0,1]$, and $\alpha = \min_{v\in V} \mathsf{P}(v,v)$ a lower bound on holding probability, then*

$$\psi(A) \geqslant \begin{cases} \dfrac{1}{12t}\, \min\{\hat{\Phi}_t(A)^2,\, t\hat{\Phi}_t(A)\}, \\ \dfrac{\min\{\alpha, t\}}{4t^2}\, \Phi_t(A)^2. \end{cases}$$

When the flow from $A \leftrightarrow A^c$ is not concentrated at any vertices, then $\Phi_t(r) = \Theta(\Phi(r))$, and this typically leads to a mixing bound t^{-1} times that of Theorem 1.2, a significant improvement when the threshold is small. In the extreme case when ergodic flow is spread uniformly over the complement, a threshold of

$$t = \min \frac{\mathsf{Q}(A, A^c)}{\min\{\pi(A), \pi(A^c)\}} \sim \Phi$$

can be used and still have $\Phi_t(r) = \Phi(r)$, leading to a mixing time upper bound matching the best-case lower bound of $\tau(\epsilon) = \Omega\big(1/\Phi\big)$.

Proof. The proof will require us to break the definition of $\psi(A)$ into an integral over $u \in [0,t]$ and one over $u \in [t,1]$, apply Jensen's inequality, and then make a few simplifications to complete the proof. The simplifications relate evolving sets to the numerator of $\hat{\Phi}_t(A)$, with (2.1) replaced by a similar term that also depends on the threshold t.

We start by showing the first bound of the theorem, that is,

$$\psi(A) \geqslant \frac{1}{12t}\, \min\{\hat{\Phi}_t(A)^2,\, t\hat{\Phi}_t(A)\}.$$

As discussed after Definition 1.5, when $t \geqslant 1/2$ then $\hat{\Phi}_t(A) = \hat{\Phi}_{1/2}(A)$. The theorem is then stronger at $t = 1/2$ than at $t > 1/2$, so without loss assume that $t \leqslant 1/2$.

If $u \leqslant 1/2$,

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

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

$$\begin{aligned} \sum_{v\in V} \min\{\mathsf{Q}_t(A,v),\, \mathsf{Q}_t(A^c,v)\} &= \sum_{v\in V} \int_0^t \pi(v)\, 1_{\{u:\min\{\mathsf{Q}_t(A,v),\,\mathsf{Q}_t(A^c,v)\}\geqslant u\pi(v)\}}\, du \\ &= \int_0^t \big[\pi(A_u \setminus A_{1-u}) + \pi\big(\{v \in V \,:\, \mathsf{Q}(A,v) \geqslant u\pi(v),\, \mathsf{Q}(A^c,v) = u\pi(v)\}\big)\big]\, du \\ &= \int_0^t \big(\pi(A_u) - \pi(A_{1-u})\big)\, du + 0 \end{aligned}$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

$$= \int_0^t (\pi(A_u) - \pi(A))\,du + \int_{1-t}^1 (\pi(A) - \pi(A_u))\,du \tag{3.1}$$
$$= \int_0^t (\pi(A_u) - \pi(A))\,du + \int_0^t (\pi((A^c)_u) - \pi(A^c))\,du.$$

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

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

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

$$\int_0^t (\pi(A_u) - \pi(A))\,du \geqslant \int_0^t (\pi((A^c)_u) - \pi(A^c))\,du.$$

By Jensen's inequality, the martingale identity $\int_0^1 \pi(A_u)\,du = \pi(A)$, and concavity of $f(x) = \sqrt{x(1-x)}$, we have

$$\int_0^1 f(\pi(A_u))\,du \leqslant t\,f\left(\int_0^t \pi(A_u)\,\frac{du}{t}\right) + (1-t)\,f\left(\int_t^1 \pi(A_u)\,\frac{du}{1-t}\right)$$
$$= t\,f\left(\pi(A) + \frac{\int_0^t (\pi(A_u) - \pi(A))\,du}{t}\right) + (1-t)\,f\left(\pi(A) - \frac{\int_0^t (\pi(A_u) - \pi(A))\,du}{1-t}\right).$$

By concavity of f this is decreasing in

$$\int_0^t (\pi(A_u) - \pi(A))\,du \geqslant \hat{\Phi}_t(A)\pi(A)\pi(A^c),$$

and so it is maximized at the lower bound. The root profile is then bounded by

$$1 - \psi(A) \leqslant t\sqrt{\left(1 + \frac{\hat{\Phi}_t(A)(1-\pi(A))}{t}\right)\left(1 - \frac{\hat{\Phi}_t(A)\pi(A)}{t}\right)} \tag{3.2}$$
$$+ (1-t)\sqrt{\left(1 - \frac{\hat{\Phi}_t(A)(1-\pi(A))}{1-t}\right)\left(1 + \frac{\hat{\Phi}_t(A)\pi(A)}{1-t}\right)}.$$

Suppose $\hat{\Phi}_t(A) \geqslant 1 - 2t$. Equation (3.2) can be simplified by using the relation

$$\sqrt{XY} + \sqrt{(1-X)(1-Y)} \leqslant \sqrt{1 - (X-Y)^2}$$

with $X = t + \hat{\Phi}_t(A)(1-\pi(A))$ and $Y = t - \hat{\Phi}_t(A)\pi(A)$; see Lemma A.1 in the Appendix for a proof. Then

$$\psi(A) \geqslant 1 - \sqrt{1 - \hat{\Phi}_t(A)^2} \geqslant \frac{1}{2}\,\hat{\Phi}_t(A)^2.$$

Since $\hat{\Phi}_t(A) \geqslant 1 - 2t$, then $\hat{\Phi}_t(A) \geqslant 1/3t\,\min\{\hat{\Phi}_t(A), t\}$, which gives the first inequality of the theorem.

Now suppose $\hat{\Phi}_t(A) < 1 - 2t$. To simplify (3.2) observe that

$$\frac{d^2}{dx^2}\sqrt{(1 + c(1-x))(1-cx)} = \frac{-c^4}{4\big[(1 + c(1-x))(1-cx)\big]^{3/2}} \leqslant 0,$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

and so the upper bound on $1-\psi(A)$ is concave in $\pi(A)$. Also,

$$\frac{d}{dx}\bigg|_{x=0}\sqrt{(1+c(1-x))(1-cx)} = \frac{-c(2+c)}{2\sqrt{1+c}},$$

and so the derivative at $\pi(A)=0$ of the upper bound in (3.2) is

$$\frac{d}{d\pi(A)}\bigg|_{\pi(A)=0} = t\,\frac{-\frac{\hat{\Phi}_t(A)}{t}\left(2+\frac{\hat{\Phi}_t(A)}{t}\right)}{2\sqrt{1+\hat{\Phi}_t(A)/t}} + (1-t)\,\frac{\frac{\hat{\Phi}_t(A)}{1-t}\left(2-\frac{\hat{\Phi}_t(A)}{1-t}\right)}{2\sqrt{1-\hat{\Phi}_t(A)/(1-t)}}$$
$$= \frac{\hat{\Phi}_t(A)}{2}\left(\sqrt{1+\frac{\hat{\Phi}_t(A)}{t}} - \sqrt{1-\frac{\hat{\Phi}_t(A)}{1-t}}\right)\left(\left[\left(1+\frac{\hat{\Phi}_t(A)}{t}\right)\left(1-\frac{\hat{\Phi}_t(A)}{1-t}\right)\right]^{-1/2} - 1\right).$$

Observe that since $\hat{\Phi}_t(A) \leqslant 1-2t$, then

$$\frac{\hat{\Phi}_t(A)}{2} \geqslant 0 \quad \text{and} \quad \sqrt{1+\frac{\hat{\Phi}_t(A)}{t}} - \sqrt{1-\frac{\hat{\Phi}_t(A)}{1-t}} \geqslant 0,$$

while

$$\left(1+\frac{\hat{\Phi}_t(A)}{t}\right)\left(1-\frac{\hat{\Phi}_t(A)}{1-t}\right) = 1 + \frac{\hat{\Phi}_t(A)(1-2t)-\hat{\Phi}_t(A)^2}{t(1-t)} \geqslant 1\,.$$

It follows that

$$\frac{d}{d\pi(A)}|_{\pi(A)=0} < 0,$$

and so (3.2) is maximized at $\pi(A)=0$ with

$$1-\psi(A) \leqslant t\sqrt{1+\frac{\hat{\Phi}_t(A)}{t}} + (1-t)\sqrt{1-\frac{\hat{\Phi}_t(A)}{1-t}}$$
$$\leqslant t\left(1+\frac{\hat{\Phi}_t(A)}{2t} - \frac{1}{12}\min\left\{\frac{\hat{\Phi}_t(A)^2}{t^2}, \frac{\hat{\Phi}_t(A)}{t}\right\}\right) + (1-t)\left(1-\frac{\hat{\Phi}_t(A)}{2(1-t)}\right)$$
$$= 1 - \frac{t}{12}\min\left\{\frac{\hat{\Phi}_t(A)^2}{t^2}, \frac{\hat{\Phi}_t(A)}{t}\right\},$$

using the inequalities

$$\sqrt{1+x} \leqslant 1 + \frac{x}{2} - 1_{\{x\geqslant 0\}}\frac{\min\{x^2,x\}}{12} \quad \text{and} \quad \sqrt{1-x} \leqslant 1-\frac{x}{2}.$$

This completes the proof of the first bound of the theorem. Now consider the second bound of the theorem,

$$\psi(A) \geqslant \frac{\min\{\alpha,t\}}{4t^2}\,\Phi_t(A)^2.$$

Suppose $t \geqslant 1/2$. If $\alpha \geqslant 1/2$ then $\Phi_t(A) = \Phi_{1/2}(A)$, and so the bound follows from the case of $t = 1/2$. If $\alpha < 1/2$ then

$$\Phi_{1/2}(A) \geqslant \frac{1}{2t}\Phi_t(A),$$

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 $t \leqslant 1/2$.

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

$$\mathsf{Q}_T(A, V) = \sum_{v \in V} \int_0^T \pi(A_u \cap \{v\})\,du = \int_0^T \pi(A_u)\,du,$$
$$\mathsf{Q}_T(A, V) = \mathsf{Q}_T(A, A) + \mathsf{Q}_T(A, A^c) = \pi(A)T + \mathsf{Q}_T(A, A^c).$$

Then

$$\int_0^T \pi(A_u)\,du = \mathsf{Q}_T(A, V) = \pi(A)T + \mathsf{Q}_T(A, A^c).$$

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

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

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

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

This is decreasing in $\mathsf{Q}_T(A, A^c)$ when $\mathsf{Q}_T(A, A^c) \geqslant \mathsf{Q}_T(A^c, A)$, and decreasing in $\mathsf{Q}_T(A^c, A)$ when $\mathsf{Q}_T(A^c, A) \geqslant \mathsf{Q}_T(A, A^c)$, and so it is maximized when

$$\mathsf{Q}_T(A, A^c) = \mathsf{Q}_T(A^c, A) = \Phi_t(A)\pi(A)\pi(A^c).$$

It follows that

$$\begin{aligned}
&1 - \psi(A) \\
&\quad \leqslant \frac{T f\left(\pi(A) + \frac{\Phi_t(A)\pi(A)\pi(A^c)}{T}\right) + (1 - 2T) f(\pi(A)) + T f\left(\pi(A) - \frac{\Phi_t(A)\pi(A)\pi(A^c)}{T}\right)}{f(\pi(A))} \\
&\quad = 2T \sqrt{\left(\frac{1}{2} + \frac{\Phi_T(A)}{2T}\pi(A^c)\right)\left(\frac{1}{2} - \frac{\Phi_T(A)}{2T}\pi(A)\right)} + (1 - 2T) \\
&\qquad + 2T \sqrt{\left(\frac{1}{2} - \frac{\Phi_T(A)}{2T}\pi(A^c)\right)\left(\frac{1}{2} + \frac{\Phi_T(A)}{2T}\pi(A)\right)} \\
&\quad \leqslant 2T\sqrt{1 - (\Phi_T(A)/2T)^2} + 1 - 2T \leqslant 1 - \Phi_T(A)^2/4T.
\end{aligned}$$

The final line was by the relation

$$\sqrt{XY} + \sqrt{(1-X)(1-Y)} \leqslant \sqrt{1-(X-Y)^2}.$$

For a proof see Lemma A.1 in the Appendix. If $T = \min\{\alpha, t\} = t$ then the theorem is immediate. If $T = \min\{\alpha, t\} = \alpha$ then use the relation

$$\Phi_T(A) \geqslant \frac{\alpha}{t}\,\Phi_t(A).$$

□

4. New canonical path bounds

The threshold conductance bounds can improve substantially over conductance bounds when ergodic flow between A and A^c is not heavily concentrated at a few vertices of A or A^c. In this section we explore such a situation: when canonical paths are well distributed among the vertices then the ergodic flow appearing on edges of canonical paths is not heavily concentrated at any vertex.

The most novel aspect of our new result is that it applies to general finite Markov chains, *i.e.*, non-reversible and non-lazy. This is made possible by using the paths to bound intersection threshold conductance $\hat{\Phi}_t(A)$ rather than the usual conductance $\Phi(A)$ or $\Phi_t(A)$. This will require a new type of canonical path.

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

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

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 $\Gamma = \{\gamma_{xy} : x, y \in V,\, x \neq y\}$ 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 P or (v, u) an edge of P^* in the path.

Some simplification is possible in the definitions of $v \in \gamma_{xy}$ and $(u, v) \in \gamma_{xy}$ by ignoring orientation of edges, although this increases congestion and leads to weaker results. However, it is not possible to replace our use of alternating canonical paths by either ordinary paths or by ordinary paths supplemented by odd-length cycles, as is possible for reversible non-lazy walks and lazy non-reversible walks. For instance, for an odd-length cycle $\mathbb{Z}_n$ the clockwise walk $\mathsf{P}(i, i+1 \mod n) = 1$ does not converge. It has an obvious set of ordinary canonical paths and odd-length cycles for each vertex, but not if paths are required to alternate between P and P^*. The parity requirement is needed because on an even-length cycle $\mathbb{Z}_n$ the simple random walk $\mathsf{P}(i, i-1 \mod n) = \mathsf{P}(i, i+1 \mod n) = 1/2$ does not converge, but there is an obvious set of (ordinary or alternating) canonical paths.

As discussed in the Introduction, there are several types of canonical path methods. However, each of these types of paths induce natural canonical alternating paths, so our new definition provides a unifying framework.

- Consider a finite reversible walk with holding probability $\alpha > 0$. Diaconis and Stroock [4] and Sinclair [14] study mixing times using ordinary canonical paths:

$$x = x_0 \to x_1 \to x_2 \to x_3 \to \cdots \to x_n = y.$$

  An alternating path can be constructed by inserting a self-loop at each vertex, except the initial one, and treating the loop as a transition of P^* (since $\mathsf{P}^*(v,v) = \mathsf{P}(v,v)$):

$$x = x_0 \to \overset{\circlearrowleft}{x_1} \to \overset{\circlearrowleft}{x_2} \to \cdots \to \overset{\circlearrowleft}{x_n} = y.$$

- Consider a finite reversible walk with no condition on α. Diaconis and Stroock [4] study mixing times using ordinary canonical paths supplemented by by cycles γ_{xx} of odd length. If $|\gamma_{xy}|$ is even then γ_{xy} is an alternating path since $\mathsf{P}^* = \mathsf{P}$ for reversible walks, while if $|\gamma_{xy}|$ is odd then γ_{xy} followed by γ_{yy} will be an even-length alternating path.
- Consider a general finite walk. Mihail [11] and Fill [6] show that it suffices to study conductance of PP^*, and so canonical paths with edges in PP^* can be used. To construct an alternating path note that an edge (x,y) with $\mathsf{PP}^*(x,y) > 0$ can be replaced by a pair of edges $x \xrightarrow{\mathsf{P}} z \xrightarrow{\mathsf{P}^*} y$ such that $\mathsf{P}(x,z) > 0$ and $\mathsf{P}^*(z,y) > 0$.
- In an earlier version of this paper we required odd-length paths alternating between P and P^*, both when $y \neq x$ and when $y = x$. To construct an even-length alternating path between x and $y \neq x$ first follow the path γ_{xy}, and then do γ_{yy} in reverse.

Our generalization of canonical path bounds will require notions of both edge and vertex congestion.

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

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

Our redefinition of edge congestion is equivalent to the standard definition (see Definition 1.7) 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. More generally, it is no larger than the standard definition, and since the conductance and mixing time bounds to be shown in this section are weaker for larger ρ_e, then using the earlier definition will give a weaker but still valid result.

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

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. *When $\Gamma = \cup_{x \neq y} \gamma_{xy}$ consists of canonical alternating paths, then*

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

If, instead, Γ consists of ordinary canonical paths then

$$\Phi_{\rho_v/\rho_e} \geqslant 1/\rho_e.$$

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 P with canonical ordinary or alternating paths. Assume $\epsilon \leqslant 1$. Let*

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

be the smallest transition in the reversal

$$\mathsf{P}^*(b,a) = \frac{\pi(a)\mathsf{P}(a,b)}{\pi(b)}$$

of the edges in the paths.

If Γ consists of canonical alternating paths, then P has mixing time

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

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

If, instead, Γ consists of ordinary canonical paths, then

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

The key equations of the proof, (4.2) and (4.3) below, are both monotone in ρ_v and ρ_e, and so any upper bounds on ρ_v and ρ_e can be substituted. Some simplifications can be made by using the relations

$$\rho_v \leqslant \rho_e \leqslant \frac{\rho_v}{\mathsf{P}_0^*(\Gamma)} < \frac{1}{2\pi_0\mathsf{P}_0^*(\Gamma)}. \tag{4.1}$$

The lower bound $\rho_v \geqslant 1 - \min \pi(v)$ for ordinary paths and $\rho_v \geqslant \frac{1}{2}(1 - \max \pi(v))$ for alternating canonical paths can be useful for simplifying the maxima. For instance, with canonical alternating paths, if $\max \pi(v) \leqslant 1/2$ then $\max\{\rho_v,\ 1/2\} \leqslant 2\rho_v$.

The terms $\max\{\rho_v/\alpha,\ \rho_e\}$ and $\log(\rho_v\rho_e\mathsf{P}_0^*(\Gamma))$ are not simply artifacts of the proof. A simple illustration of this is the walk on a cycle $\mathbb{Z}_n$ with $\mathsf{P}(i,i) = \alpha \in (0,1)$ and $\mathsf{P}(i,i+1) = 1-\alpha$. With the obvious choice of paths this has

$$\rho_v = \frac{n-1}{2}, \quad \rho_e = \frac{n-1}{2(1-\alpha)}, \quad \mathsf{P}_0^*(\Gamma) = 1-\alpha, \quad \pi_0 = 1/n$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

and mixing time

$$\tau(\epsilon) = \Theta\left(\frac{n^2}{\min\{\alpha, 1-\alpha\}} \log \frac{1}{\epsilon}\right),$$

matching the prediction of our theorem. The π_0 term cannot always be dropped, however, as the lazy walk on the complete graph K_n has mixing time $\tau(1/4) = \Omega(\log n)$.

Proof of Theorem 4.4. First consider ordinary canonical paths. The theorem is trivial if $|V| = 1$, so assume that $|V| \geqslant 2$ and $0 < \min \pi(v) \leqslant 1/2$. Lemma 4.3 and Theorem 3.2 bound the root profile as

$$\psi(A) \geqslant \frac{\min\{\alpha, \rho_v/\rho_e\}}{4(\rho_v/\rho_e)^2} \frac{1}{\rho_e^2} = \frac{1}{4\rho_v \max\{\rho_v/\alpha, \rho_e\}}. \tag{4.2}$$

Evaluating the integral in the convex case of Theorem 2.2 gives the result when $\rho_v\rho_e\mathsf{P}_0^*(\Gamma) \leqslant e$, so assume that $\rho_v\rho_e\mathsf{P}_0^*(\Gamma) > e$. Since the walk is ergodic then the state space is strongly connected, and so $\mathsf{Q}(A, A^c) > 0$ for every set $A \notin \{\emptyset, V\}$. In particular, $\mathsf{Q}_{\mathsf{P}_0^*(\Gamma)}(A, A^c) \geqslant \pi_0\mathsf{P}_0^*(\Gamma)$, and similarly $\mathsf{Q}_{\mathsf{P}_0^*(\Gamma)}(A^c, A) \geqslant \pi_0\mathsf{P}_0^*(\Gamma)$, and so

$$\Phi_{\mathsf{P}_0^*(\Gamma)}(A) \geqslant \frac{\pi_0\mathsf{P}_0^*(\Gamma)}{\pi(A)\pi(A^c)}.$$

By Theorem 3.2,

$$\psi(A) \geqslant \max\left\{\frac{1}{4\rho_v \max\{\rho_v/\alpha, \rho_e\}}, \frac{\pi_0^2 \min\{\alpha, \mathsf{P}_0^*(\Gamma)\}}{4\pi(A)^2}\right\}. \tag{4.3}$$

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

$$\tau(\epsilon) \leqslant \int_0^c \frac{2r\,dr}{\pi_0^2 \min\{\alpha, \mathsf{P}_0^*(\Gamma)\}} + 2\rho_v \max\left\{\frac{\rho_v}{\alpha}, \rho_e\right\} \int_c^{1/\epsilon^2} \frac{dr}{r}$$

if

$$c = \pi_0\sqrt{\min\{\alpha, \mathsf{P}_0^*(\Gamma)\}\, \rho_v \max\{\rho_v/\alpha, \rho_e\}} \leqslant 1/\epsilon^2.$$

The theorem follows by integrating and using equations (4.1), $\rho_v\rho_e\mathsf{P}_0^*(\Gamma) \geqslant 1$ and $1/2 \leqslant \rho_v < 1/2\pi_0$.

For alternating canonical paths, $\psi(A) \geqslant \left(48\max\{\rho_v, 1/2\}\rho_e\right)^{-1}$ by Lemma 4.3 and Theorem 3.2. Theorem 2.2 shows the mixing bound when $4\rho_v\rho_e\mathsf{P}_0^*(\Gamma) \leqslant e$. To improve $\psi(A)$, use

$$\hat{\Phi}_{\mathsf{P}_0^*(\Gamma)} \geqslant \frac{\pi_0\mathsf{P}_0^*(\Gamma)}{\pi(A)},$$

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 $c = 2\pi_0\sqrt{\rho_v\rho_e\mathsf{P}_0^*(\Gamma)}$. □

We now return 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:

$$\text{for all } v \in V, \qquad \pi(v) \geqslant \frac{1}{2\rho_v} \sum_{\gamma_{xy}\cup\gamma_{yx}\ni v} \pi(x)\pi(y)$$
$$\geqslant \frac{1}{2\rho_v} \sum_{\substack{(x,y):\,\exists u\in A,\\ (u,v)\in\gamma_{xy}\cup\gamma_{yx}}} \pi(x)\pi(y),$$
$$\text{for all } u \in A,\ v \in V, \qquad \pi(u)\mathsf{P}(u,v) \geqslant \frac{1}{2\rho_e} \sum_{\gamma_{xy}\cup\gamma_{yx}\ni(u,v)} \pi(x)\pi(y)$$

Hence, for any $v \in V$,

$$\mathsf{Q}_{\rho_v/\rho_e}(A,v) = \min\left\{\sum_{u\in A} \pi(u)\mathsf{P}(u,v),\ \frac{\rho_v}{\rho_e}\,\pi(v)\right\}$$
$$\geqslant \frac{1}{2\rho_e}\,\min\left\{\sum_{u\in A}\ \sum_{\gamma_{xy}\cup\gamma_{yx}\ni(u,v)} \pi(x)\pi(y),\ \sum_{\substack{(x,y):\,\exists u\in A,\\ (u,v)\in\gamma_{xy}\cup\gamma_{yx}}} \pi(x)\pi(y)\right\}$$
$$= \frac{1}{2\rho_e} \sum_{\substack{(x,y):\,\exists u\in A,\\ (u,v)\in\gamma_{xy}\cup\gamma_{yx}}} \pi(x)\pi(y). \tag{4.4}$$

In the ordinary canonical path case it follows from (4.4) that

$$\mathsf{Q}_{\rho_v/\rho_e}(A,A^c) = \sum_{v\in A^c} \mathsf{Q}_{\rho_v/\rho_e}(A,v)$$
$$\geqslant \sum_{v\in A^c} \frac{1}{2\rho_e} \sum_{\substack{(x,y):\,\exists u\in A,\\ (u,v)\in\gamma_{xy}\cup\gamma_{yx}}} \pi(x)\pi(y)$$
$$\geqslant \frac{1}{2\rho_e} \sum_{(x,y)\in A\times A^c} 2\pi(x)\pi(y) = \frac{\pi(A)\pi(A^c)}{\rho_e}.$$

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 $x_{i+1} \in A^c$. Replacing A with A^c shows that

$$\mathsf{Q}_{\rho_v/\rho_e}(A^c,A) \geqslant \frac{\pi(A)\pi(A^c)}{\rho_e}$$

as well. It follows that $\Phi_{\rho_v/\rho_e}(A) \geqslant 1/\rho_e$.

If the paths are alternating then (4.4) shows that

$$\sum_{v\in V} \min\{\mathsf{Q}_{\rho_v/\rho_e}(A,v),\ \mathsf{Q}_{\rho_v/\rho_e}(A^c,v)\} \geqslant \sum_{v\in V} \frac{1}{2\rho_e} \sum_{\substack{(x,y):\exists u\in A,\, w\in A^c:\\ (u,v)\in\gamma_{xy}\cup\gamma_{yx},\\ (w,v)\in\gamma_{xy}\cup\gamma_{yx}}} \pi(x)\pi(y)$$
$$\geqslant \frac{1}{2\rho_e} \sum_{(x,y)\in A\times A^c} 2\pi(x)\pi(y) \geqslant \frac{\pi(A)\pi(A^c)}{\rho_e}.$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

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 it is counted when $v = x_{2i+1}$. It follows that $\hat{\Phi}_{\rho_v/\rho_e}(A) \geqslant 1/2\rho_e$. □

Remark 4.5. Ordinary canonical paths can be used to bound the mixing time of a lazy walk in three main ways: Jerrum and Sinclair's bound with lead term ρ_e^2, the Poincaré bound with $\rho_e \ell$, and our bound with $\rho_e \rho_v$. Since $\rho_v \leqslant \rho_e$, our bound improves on Jerrum and Sinclair's. To compare to the Poincaré bound, define the 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

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

Likewise

$$\bar{\rho}_v \geqslant \frac{\bar{\ell} + 1}{2}(1 - \|\pi\|_2^2),$$

and so

$$\bar{\rho}_v = \Theta\big(\bar{\ell}(1 - \|\pi\|_2^2)\big).$$

Our bound is best when there are a few very long paths or the distribution is concentrated near a single vertex, as it is then likely that $\rho_v \ll \ell$. However, it is more often the case that path length varies little and a bottleneck causes a few states to have high vertex congestion, in which case $\ell \ll \rho_v$. In contrast, 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 setting: first, the classical problem of the max-degree walk on a graph; then a more interesting example, walks on Cayley graphs, *i.e.*, random walks on groups.

Example 5.1. An Eulerian multigraph is a strongly connected graph with equal in-degree and out-degree at each vertex. This is 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 $\mathsf{P}(x, y) = (d(x, y))/d$ if

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

$y \neq x$ and $\mathsf{P}(x,x) = 1 - (d(x) - d(x,x))/d$. The stationary distribution $\pi = 1/n$ is uniform and the walk is lazy if $d(x) - d(x,x) \leqslant d/2$ at each vertex.

Suppose that $\mathsf{P}(x,x) > 0$ at every vertex, so that the holding probability is $\alpha \geqslant 1/d$. For instance, it suffices that each max-degree vertex has a self-loop. For every $x \neq y$, let γ_{xy} be an ordinary path from x to y. Then

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

By Theorem 4.4,

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

Suppose the self-loop requirement that every $\mathsf{P}(x,x) > 0$ is dropped. Ergodicity is not guaranteed, even if the Eulerian multigraph is strongly connected, *e.g.*, the cycle walk $\mathsf{P}(i, i+1 \mod n) = 1$ on $\mathbb{Z}_n$. However, if the graph is connected under canonical alternating paths then it is still true that $\rho_v \leqslant n/2$ and $\rho_e \leqslant nd/2$, and so by Theorem 4.4

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

This improves on previous general bounds. For instance, Sinclair's Poincaré bound 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 Stroock'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 $(\min_{\mathsf{P}(x,y)>0} \mathsf{P}(x,y))^{-1}$ term in ρ_e with $(\min_{\mathsf{PP}^*(x,y)>0} \mathsf{PP}^*(x,y))^{-1}$, which typically replaces the order d term with order d^2.

Example 5.2. Given a finite group G, a (non-symmetric) generating set $S \subset G - \{\text{id}\}$ is any subset with the property that $\bigcup_{n=0}^{\infty} S^n = G$. The Cayley graph of G has edge set (g, gs) for all $g \in G$, $s \in S$. If $p : G \to [0,1]$ is a probability distribution supported on $S \cup \{\text{id}\}$, then $\mathsf{P}(g, gs) = p(s)$ defines a Markov chain on the Cayley graph 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 $N(g,s) \leqslant \Delta$ denote the number of times generator s appears in the representation of g.

Babai [1] showed

$$\tau(\epsilon) = O\left(\frac{\Delta^2}{\min_{s\in S} p(s)} \log \frac{|G|}{\epsilon}\right)$$

for the lazy walk with symmetric generating set, *i.e.*, $p(\text{id}) \geqslant 1/2$ with $S = S^{-1}$ and $p(s) = p(s^{-1})$ for all $s \in S$. 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 [12]), leading to the following generalizations

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

of Babai's result to the symmetric and non-symmetric cases respectively:

$$\tau(\epsilon) \leqslant \max\left\{\frac{1}{2\,p(\mathrm{id})},\, \Delta \max_{g\in G,\, s\in S} \frac{N(g,s)}{p(s)}\right\} \left(\frac{1}{2}\log|G| + \log\frac{1}{\epsilon}\right),$$

$$\tau(\epsilon) \leqslant \frac{\Delta}{p(\mathrm{id})} \max_{g\in G,\, s\in S} \frac{N(g,s)}{p(s)} \left(\frac{1}{2}\log|G| + \log\frac{1}{\epsilon}\right). \tag{5.1}$$

Now consider our new method of canonical alternating paths. Let Δ_{alt} be the diameter measured using canonical alternating paths, *i.e.*,

$$\Delta_{\mathrm{alt}} = \min\left\{2N \,:\, G = \bigcup_{n=0}^{N} (SS^{-1})^n\right\},$$

and let $N_{\mathrm{alt}}(g,s)$ be the count frequency of generators when group elements are written in terms of canonical alternating paths, *e.g.*, $g = s_1\, s_2^{-1}\, s_3\, s_4^{-1} \cdots s_{2n-1}\, s_{2n}^{-1}$. Then

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

The proof is left to the Appendix because it uses essentially the same approach that Diaconis and Saloff-Coste used for ordinary canonical paths in [3]. Substituting these into Theorem 1.9 shows that

$$\tau(\epsilon) \leqslant 6\,\Delta_{\mathrm{alt}} \max_{g\in G,\, s\in S} \frac{N_{\mathrm{alt}}(g,s)}{p(s)} \left(\frac{1}{2}\log|G| + \log\frac{1}{\epsilon}\right) \tag{5.2}$$

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

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

The inequality holds because

$$\Delta = \max_{g\in G} \sum_{s\in S} p(s)\, \frac{N(g,s)}{p(s)} \leqslant \max_{g\in G} \max_{s\in S} \frac{N(g,s)}{p(s)}\ .$$

Substituting this into (5.2) shows that (5.2) generalizes (5.1) to the setting of walks on Cayley graphs with no holding probability.

Appendix

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

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

$$\sqrt{XY} + \sqrt{(1-X)(1-Y)} \leqslant \sqrt{1-(X-Y)^2}.$$

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

Proof. Observe that with $g(X,Y) = \sqrt{XY} + \sqrt{(1-X)(1-Y)}$ then

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

Now, $\sqrt{A^2-B} \leqslant A-B$ if $A^2 \geqslant B$, $A \leqslant (1+B)/2$ and $A \geqslant B \geqslant 0$ (to show this, square both sides). These conditions are easily verified with

$$A = 1-(X+Y)+2XY \quad \text{and} \quad B = 1-2(X+Y)+(X+Y)^2,$$

and so

$$g(X,Y)^2 \leqslant 2\big[1-(X+Y)+2XY\big] - \big[1-2(X+Y)+(X+Y)^2\big] \\ = 1+2XY - X^2 - Y^2 = 1-(X-Y)^2.$$ □

In order to study walks on Cayley graphs it was claimed that the 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., $p(s) = p(s^{-1})$ for all $s \in S$, then

$$\rho_v < \frac{\Delta+1}{2}.$$

There are canonical alternating paths with

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

and

$$N_{\mathrm{alt}}(g,s) \leqslant \frac{1}{2}\Delta_{\mathrm{alt}} < |G| \quad \text{for every } g \in G,\ s \in S.$$

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 $x \to xs_1 \to \cdots \to xg = y$. 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.*

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

Then

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

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.1) improves to

$$\frac{1}{2} \sum_{x\neq y} \pi(x)\pi(y)\left(|\gamma_{xy}| + 1\right) \quad \text{and} \quad \rho_v \leqslant \frac{\Delta+1}{2}\left(1 - \frac{1}{|G|}\right).$$

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,

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

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,

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

Similarly, minor changes show that

$$\rho_e < \frac{1}{2} \max_{s\in S} \frac{1}{p(s)} \max_{g\in G} N_{\text{alt}}(g, s).$$

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

https://doi.org/10.1017/S096354831400025X Published online by Cambridge University Press

subcycles. In particular, if $i, j \leqslant k$ 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 $\{\text{id}, s_1 s_2^{-1}, \ldots, s_1 s_2^{-1} \cdots s_{2k-1} s_{2k}\}$ is a set of $k+1$ distinct elements, and so $k+1 \leqslant |G|$, while it also guarantees that $s s^{-1}$ never appears and so also $N_{\text{alt}}(g, s) \leqslant |g|/2 \leqslant |G| - 1$. □

References

[1] Babai, L. (1991) Local expansion of vertex-transitive graphs and random generation in finite groups. *Proc. 23rd Annual ACM Symposium on Theory of Computing: STOC 1991*, pp. 164–174.

[2] Diaconis, P. and Fill, J. (1990) Strong stationary times via a new form of duality. *Ann. Probab.* **18** 1483–1522.

[3] Diaconis, P. and Saloff-Coste, L. (1993) Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.

[4] Diaconis, P. and Stroock, D. (1991) Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

[5] Dyer, M., Goldberg, L., Jerrum, M. and Martin, R. (2006) Markov chain comparison. *Probab. Surv.* **3** 89–111.

[6] Fill, J. (1991) Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

[7] Greenhill, C. (2013) Making Markov chains less lazy. arXiv:1203.6668

[8] Jerrum, M. and Sinclair, A. (1988) Conductance and the rapid mixing property for Markov chains: The approximation of the permanent resolved. In *Proc. 20th Annual ACM Symposium on Theory of Computing: STOC 1988*, pp. 235–243.

[9] Kannan, R., Lovász, L. and Montenegro, R. (2006) Blocking conductance and mixing in random walks. *Combin. Probab. Comput.* **15** 541–570.

[10] Lawler, G. and Sokal, A. (1988) Bounds on the l^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580.

[11] Mihail, M. (1989) Conductance and convergence of Markov chains: A combinatorial treatment of expanders. In *30th Annual Symposium on Foundations of Computer Science*, pp. 526–531.

[12] Montenegro, R. and Tetali, P. (2006) *Mathematical Aspects of Mixing Times in Markov Chains*, Vol. 1:3 of *Foundations and Trends in Theoretical Computer Science*, NOW Publishers.

[13] Morris, B. and Peres, Y. (2005) Evolving sets, mixing and heat kernel bounds. *Probab. Theory Rel. Fields* **133** 245–266.

[14] Sinclair, A. (1992) Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1** 351–370.