A Semidefinite Bound for Mixing Rates of Markov Chains

Nabil Kahale*

AT&T Bell Laboratories, Murray Hill, NJ 07974. Email: kahale@research.att.com.

Abstract. We study the method of bounding the spectral gap of a reversible Markov chain by establishing canonical paths between the states. We provide natural examples where improved bounds can be obtained by allowing variable length functions on the edges. We give a simple heuristic for computing good length functions. Further generalization using multicommodity flow yields a bound which is an invariant of the Markov chain, and which can be computed at an arbitrary precision in polynomial time via semidefinite programming. We show that, for any reversible Markov chain on n states, this bound is off by a factor of order at most $\log^2 n$, and that this can be tight.

1 Introduction

Let $(X_m), m \geq 0$, be an irreducible Markov chain on a finite state space V with transition matrix P and stationary distribution π. We assume that P is reversible, that is

$$\pi(x)P(x,y) = \pi(y)P(y,x) = Q(x,y), \text{ for all } x, y \in V.$$

Under these conditions, all the eigenvalues of P are real, and will be denoted by $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1}$, where $n = |V|$.

It is known that the rate of convergence of a Markov chain can be bounded using the eigenvalues of P (see, e.g., [5]). Informally, the chain is rapidly mixing if $\max(\lambda_1, |\lambda_{n-1}|)$ is small compared to 1. Since we can assume the eigenvalues to be non-negative by replacing P with $\frac{1}{2}(I + P)$, it suffices to upper-bound $\tau = 1/(1 - \lambda_1)$ in order to bound the mixing rate of the (tranformed) Markov chain. Random walks have been used to establish approximation algorithms for hard combinatorial problems (see [10, 15, 12], and references therein). General results on Markov chains can be found in [1].

Let $G = (V, E)$ be the graph on the vertex set V, with $(x, y) \in E$ if and only if $Q(x, y) > 0$ and $x \neq y$. We assume that the elements of E are *directed*

* Based on DIMACS Tech Report 95-41, September 95. This work was partly done while the author was at the Massachusetts Institute of Technology, at the Institute for Mathematics and its Applications, at DIMACS and at XEROX Palo Alto Research Center.

edges. Following [10], Diaconis and Stroock [5] proposed a method for bounding τ by establishing a simple path γ_{xy} in G from x to y, for each $(x, y) \in V \times V$, $x \neq y$. Given a length function l that assigns a non-negative real number to each element of E, let

$$|\gamma_{xy}|_l = \sum_{e \in \gamma_{xy}} l(e).$$

Diaconis and Stroock showed the following Poincaré inequality:

$$\tau \leq \max_e \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_{Q^{-1}} \pi(x)\pi(y), \qquad (1)$$

where $|\gamma_{xy}|_{Q^{-1}}$ is the length of the path γ_{xy} according to the length function $l(e) = Q^{-1}(e)$. They also provided several examples where Eq. 1 gives a good bound on τ. Along the same lines, Sinclair [14] showed that

$$\tau \leq \max_e \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}| \pi(x)\pi(y), \qquad (2)$$

where $|\gamma_{xy}|$ is the number of edges in γ_{xy}. The two bounds in Eqs. 1 and 2 coincide in the case of the nearest-neighbor random walk on a graph, but are in general incomparable.

As observed in [6, 16], the bounds 1 and 2 can be generalized as follows. Let Γ be the set of paths γ_{xy}.

Proposition 1. *For any positive length function l, $\tau \leq \kappa(\Gamma, l)$, where*

$$\kappa(\Gamma, l) = \max_e \frac{1}{l(e)Q(e)} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_l \pi(x)\pi(y).$$

Proposition 1 can be shown by a straightforward generalization of the proofs of the aformentioned bounds. We give a different proof of Proposition 1 in Section 2, by using the theory of non-negative matrices. Under a general condition, this allows us to characterize the length functions l that minimize $\kappa(\Gamma, l)$. Most of the material in Section 2 has been obtained independently in [6, 16].

In Section 3, we give examples where the bound given by a simple length function improves upon the bounds 1 and 2. We provide a heuristic for finding such a length function. In Section 4, we extend our results to the case where Γ is a multicommodity flow. Multicommodity flow has be used in [14] in the case of the unit length function. We show that the best bound μ on τ that can be obtained via this method, over all flows and length functions, can be computed at an arbitrary precision in polynomial time by a reduction to a semidefinite program. The bound μ is an invariant of the Markov chain. It coincides with τ in the case of a birth-death chain. A simple expression for μ is derived for edge-transitive graphs. We also show that the gap between τ and μ is at most of order $\log^2 n$, and that for expander graphs the gap is actually of order $\log^2 n$. We give an application to the exclusion process. Our results can be easily adapted to bound the second smallest eigenvalue of the Laplacian of a graph, essentially by replacing $\pi(x)$ by 1 and $Q(e)$ by n, as discussed in Section 5.

2 Bounding τ via Paths and Length Functions

Let $\mathcal{Q}$ be the diagonal matrix indexed by E, with $\mathcal{Q}_{e,e} = Q(e)$, and Π the diagonal matrix indexed by $(V \times V) - \{(x,x) : x \in V\}$, with $\Pi_{(x,y),(x,y)} = \pi(x)\pi(y)$. We represent Γ by the matrix, also denote by Γ, whose rows are indexed by $(V \times V) - \{(x,x) : x \in V\}$ and columns indexed by E, with

$$\Gamma_{(x,y),e} = \begin{cases} 1 & \text{if } e \in \gamma_{xy}, \\ 0 & \text{otherwise.} \end{cases}$$

We say that a non-negative matrix is irreducible if its associated graph is strongly connected.

Theorem 2. *All the eigenvalues of the matrix $M = \mathcal{Q}^{-1}\Gamma^t \Pi \Gamma$ are real. Let $\mu_0(\Gamma)$ be the largest one. Then*

$$\tau \leq \mu_0(\Gamma). \tag{3}$$

Moreover,

$$\mu_0(\Gamma) = \min_l \kappa(\Gamma, l), \tag{4}$$

where l ranges over the set of positive length functions. If M is irreducible then $\mu_0(\Gamma) = \kappa(\Gamma, l)$ if and only if l is a positive eigenvector of M.

Proof. M is similar to the symmetric matrix $\sqrt{\mathcal{Q}}^{-1}\Gamma^t \Pi \Gamma \sqrt{\mathcal{Q}}^{-1}$, and thus all its eigenvalues are real. Let δ be the linear operator that maps every $\phi \in L^2(V)$ to $\phi' \in L^2(E)$, with $\phi'((x,y)) = \phi(x) - \phi(y)$. For $\phi \in L^2(V)$, we have $(\Gamma\phi')(x,y) = \phi(x) - \phi(y)$. Therefore, by the variational characterization of λ_1,

$$\begin{aligned}
\tau &= \max_\phi \frac{\sum_{(x,y)\in V\times V} \pi(x)\pi(y)(\phi(x)-\phi(y))^2}{\sum_{(x,y)\in V\times V} Q(x,y)(\phi(x)-\phi(y))^2} \\
&= \max_\phi \frac{||\sqrt{\Pi}\Gamma\phi'||^2}{||\sqrt{\mathcal{Q}}\phi'||^2} \\
&\leq \max_{l\in L^2(E)-\{0\}} \frac{||\sqrt{\Pi}\Gamma l||^2}{||\sqrt{\mathcal{Q}} l||^2} \\
&= \max_{l\in L^2(E)-\{0\}} \frac{||\sqrt{\Pi}\Gamma\sqrt{\mathcal{Q}}^{-1} l||^2}{||l||^2}
\end{aligned}$$

where, in the first two equations, ϕ ranges over the non-constant functions in $L^2(V)$. The last term is equal to the largest eigenvalue of $\sqrt{\mathcal{Q}}^{-1}\Gamma^t\Pi\Gamma\sqrt{\mathcal{Q}}^{-1}$, which is $\mu_0(\Gamma)$, since M and $\sqrt{\mathcal{Q}}^{-1}\Gamma^t\Pi\Gamma\sqrt{\mathcal{Q}}^{-1}$ have the same eigenvalues. Eq. 3 follows.

To show Eq. 4, we first note that

$$(Ml)(e) = \frac{1}{Q(e)} \sum_{\gamma_{xy}\ni e} |\gamma_{xy}|_l \pi(x)\pi(y). \tag{5}$$

Thus,

$$\kappa(\Gamma, l) = \max_e \frac{(Ml)(e)}{l(e)}. \tag{6}$$

for all $e \in E$. From the theory of non-negative matrices [13], any non-negative matrix B has a largest eigenvalue (in module) which is non-negative real, and equal to $\min_l \max_e (Bl)(e)/l(e)$, where l ranges over the set of positive vectors. (For the rest of this paper, we will refer to this eigenvalue simply as the largest eigenvalue of B.) Together with Eq. 6, this implies that the largest eigenvalue of M is equal to $\min_l \kappa(\Gamma, l)$, where l ranges over the set of positive length functions. Hence Eq. 4.

Finally, if l is a positive eigenvector of M, it corresponds to $\mu_0(\Gamma)$, by the theory of non-negative matrices. Thus, $(Ml)(e) = \mu_0(\Gamma)l(e)$. By Eq. 6, this implies that $\kappa(\Gamma, l) = \mu_0(\Gamma)$. Conversely, if M is irreducible and $\kappa(\Gamma, l) = \mu_0(\Gamma)$, then $Ml \leq \mu_0(\Gamma)l$, where inequality holds for every coordinate. By the theory of non-negative matrices, since M is irreducible and l positive, it follows that l is an eigenvector of M corresponding to $\mu_0(\Gamma)$.

Remark. Eq. 3 can be generalized as follows: For every i, $1 \leq i \leq n-1$,

$$\mu_{|E|-n+i} \leq \frac{1}{1-\lambda_i} \leq \mu_{i-1},$$

where μ_{i-1} is the i-th largest eigenvalue of M. The proof is similar to the case $i = 1$, and uses the usual minimax characterization of eigenvalues

$$\frac{1}{1-\lambda_i} = \min_H \max_{\phi \in H} \frac{\sum_{(x,y)\in V\times V} \pi(x)\pi(y)(\phi(x) - \phi(y))^2}{\sum_{e\in E} Q(e)\phi'(e)^2},$$

and a similar relation involving μ_i, where H ranges over the set of vector subspaces of $L_0^2(V)$ of dimension $n-i$, and $L_0^2(V)$ denotes the set of $\phi \in L^2(V)$ such that $\sum_{x\in V} \pi(x)\phi(x) = 0$. Note that H and $\delta(H)$ have the same dimension since the restriction of δ to $L_0^2(V)$ is injective. The lower bound follows similarly.

3 Good Length Functions

We assume in this section that we are given a set of paths in G between every pair of vertices. Theorem 2 shows that, under the general condition that M is irreducible, a length function that gives the best bound on τ is an eigenvector of M corresponding to its largest eigenvalue. Such an eigenvector is generally not easy to compute, however. We give in this section a method for finding a length function that tends to give a good bound on τ.

Proposition 3. *For any positive length function l, $\kappa(\Gamma, Ml) \leq \kappa(\Gamma, l)$. Moreover,*

$$\mu_0(\Gamma) \geq \min_e \frac{(Ml)(e)}{l(e)}. \tag{7}$$

Proof. Let $l' = Ml$. By Eq. 6, $\kappa(\Gamma, l)$ is the smallest number such that $l' \leq \kappa(\Gamma, l)l$. Since M is a non-negative matrix, it follows that $Ml' \leq \kappa(\Gamma, l)Ml = \kappa(\Gamma, l)l'$. Using Eq. 6 again, we conclude that $\kappa(\Gamma, l') \leq \kappa(\Gamma, l)$, as desired. Eq. 7 follows from the fact that, for any non-negative matrix B and any positive vector l, the largest eigenvalue of B is at least $\min_e (Bl)(e)/l(e)$.

Proposition 3 provides a method for computing better and better bounds on $\mu_0(\Gamma)$ by starting with any positive vector l, computing Ml via Eq. 5, and bounding $\mu_0(\Gamma)$ by $\max_e (Ml)(e)/l(e)$. We can obtain a better bound by repeating the same process, starting with the length function Ml. In the limit, the bound that we get is exactly $\mu_0(\Gamma)$ since, for any non-negative matrix B and any positive vector l, $\max_e (B^{t+1}l)(e)/(B^t l)(e)$ converges to the largest eigenvalue of B as t goes to infinity.

In practice, it suffices to calculate Ml within a constant multiplicative factor. Indeed, if l' approximates Ml within a constant multiplicative factor, then $\kappa(\Gamma, Ml)$ and $\kappa(\Gamma, l')$ differ only by a constant multiplicative factor.

Observe that if γ_{xy} and γ_{yx} consist of the same sequence of edges, but in reverse order, for any $(x, y) \in (V \times V) - \{(x, x) : x \in V\}$, then there exists a symmetric length function l that minimizes $\kappa(\Gamma, l)$. This can be seen by replacing a function l_1 that minimizes $\kappa(\Gamma, l_1)$ with $l_1(x, y) + l_1(y, x)$. Note also that if l is symmetric, then so is the improved length function Ml.

3.1 Examples

An important reversible Markov chain is the nearest-neighbor random walk on a graph $G = (V, E)$. For this Markov chain, $P(x, y) = 1/d_x$, for $(x, y) \in E$, where d_x is the degree of x in G. Also, $\pi(x) = d_x/|E|$ and $Q(e) = 1/|E|$, where $|E| = \sum_{x \in V} d_x$ is the number of *directed edges* in G.

d-ary tree. In a d-ary tree T of depth h, each node which is not a leave has exactly d descendants. The case $d = 2$ was treated in [5], where it was shown that Eq. 1 gives a bound on τ off by a multiplicative factor of h. Indeed, the number of nodes in T is $n = (d^{h+1} - 1)/(d - 1)$. There is a unique simple path between every pair of points in the tree. For a symmetric length function l, let $\mathrm{diam}_l(T)$ be the maximum l-length of any simple path in the tree. For any edge $e = (u, v)$ such that the depth of u is i and the depth of v is $i-1$, with $1 \leq i \leq h$, we have

$$\begin{aligned} (Ml)(e) &= \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_l \pi(x)\pi(y) \qquad (8) \\ &\leq |E| \mathrm{diam}_l(T) \sum_{x : \exists y, \gamma_{xy} \ni e} \pi(x) \\ &\leq 2\mathrm{diam}_l(T) \frac{d^{h+1-i}}{d-1}. \end{aligned}$$

If l is the unit length function $l(e) = 1$, Eq. 8 shows that $(Ml)(e) \leq 4hd^{h+1-i}/(d-1)$. This upper bound on $(Ml)(e)$ is tight, up to a constant factor. It implies that

$\kappa(\Gamma, l) \leq 4hd^h/(d-1)$. To obtain a better bound on μ_0, we replace l by Ml, which is (up to a constant factor) equivalent to replacing l by l_1, with $l_1(e) = d^{-i}$. By Eq. 8, $(Ml_1)(e) \leq 4d^{h+1-i}/(d-1)^2$, and so $\kappa(\Gamma, l_1) \leq 4d^{h+1}/(d-1)^2$. On the other hand, if ϕ is the vector that takes value 1 on a subtree rooted at one of the d children of the root of T, and value -1 at another subtree, and value 0 elsewhere, the numerator in the variational expression of τ is at least $1/(2d)$, the denominator is $4/|E|$ and so $\tau \geq (d^h - 1)/(4(d-1))$. In conclusion,

$$\frac{d^h - 1}{4(d-1)} \leq \tau \leq \mu_0 \leq \frac{4d^{h+1}}{(d-1)^2}.$$

This determines τ and μ_0 up to a constant factor. The constants in the above inequality can be easily improved, and we have made no attempt to optimize them.

Birth-death chain. In a birth-death chain, G consists of a single line, whose vertices are labeled from 0 to $n-1$. This chain is always reversible. In this case too, the paths are uniquely determined. Moreover, $\tau = \mu_0$. This is because

$$M = \begin{pmatrix} M_1 & 0 \\ 0 & M_1 \end{pmatrix},$$

where the first $n(n-1)/2$ rows are indexed by the pairs $(i, i+1)$, $0 \leq i \leq n-2$, and the last $n(n-1)/2$ rows are indexed by the pairs $(i+1, i)$, $0 \leq i \leq n-2$. In the proof of Theorem 2, we can restrict the summations to $(x, y) \in V \times V$, with $x < y$, and replace all matrices, vectors and operators involved with their restriction to the indices $(x, y) \in V \times V$, $x < y$, and $(i, i+1)$, $0 \leq i \leq n-2$. We conclude that τ is at most the largest eigenvalue of M_1. But, since any vector l indexed by $(i, i+1)$, $0 \leq i \leq n-2$, is of the form $(\phi(i) - \phi(i+1))$, for $\phi \in L^2(V)$, the third equation is in fact an equality, and so τ is equal to the largest eigenvalue of M_1, which is equal to μ_0. (Note that the equality $\tau = \mu_0$ does not hold for all trees; it is shown in Subsection 4.1 that the two quantities differ for stars with more than 2 leaves.)

Consider now the particular case where $P(i-1, i) = \beta = 1 - P(i, i-1)$ for $1 \leq i \leq n-1$, and $P(n-1, n-1) = \beta = 1 - P(0, 0)$, where $0 < \beta < 1/2$ is a constant. The chain is ergodic and reversible, with stationary distribution $\pi(i) = r^i/Z$, where $r = \beta/(1-\beta)$, and $Z = (1-r^n)/(1-r)$. This example was considered in [14] where it is shown that, when r is fixed, the bound on τ given by Eq. 2 is linear in n, whereas τ is of constant order. Indeed, $Q(i-1, i) = \beta r^{i-1}/Z$, for $1 \leq i \leq n-1$. For the unit length function l that assigns 1 to every edge, we have $(Ml)(i-1, i) = \Theta(i)$, where the constants behind Θ may depend on β. Thus the maximum of $(Ml)(e)/l(e)$ differs from its mimimum by a factor of order n. A similar situation happens for the length function Ml, since $(M^2 l)(i-1, i) = \Theta(i^2)$. The reverse situation happens for the length function $l_{DS}(e) = 1/Q(e)$, where $(Ml_{DS})(i-1, i)/l_{DS}(i-1, i) = \Theta(n-i)$. This suggests that an intermediate length function of the form $l_\alpha(i-1, i) = \alpha^{i-1}$, with $1 < \alpha < r^{-1}$, would give a good bound on τ. By bounding $|\gamma_{xy}|_{l_\alpha}$ by $\alpha^y/(\alpha - 1)$, a simple calculation shows that

$$\frac{(Ml_\alpha)(e)}{l_\alpha(e)} \le \frac{(1+r)\alpha}{(\alpha-1)(1-r\alpha)}.$$

Thus, any such α gives a constant bound on τ. This bound is minimum for $\alpha = r^{-1/2}$, in which case it yields $\lambda_1 \le 2r^{1/2}/(1+r)$. This is very close to the exact value of λ_1, which is equal to $2r^{1/2}\cos(\pi/n)/(1+r)$. Note that the *Cheeger inequality* (see [14]) gives a constant but weaker bound on τ. Examples similar to this one can be found in [4].

3.2 Bounded-degree Graphs

Further simplification can be acheived for the nearest neighbor walk on a bounded-degree graph. Let d be the maximum degree of G, l a positive length function, Γ a set of simple paths between every pair of distinct points, cong(e) the number of pairs (x, y) such that $e \in \gamma_{xy}$, and $\mathrm{diam}_l(G,\Gamma)$ the maximum of $|\gamma_{xy}|_l$ over all pairs x and y. By Eq. 5,

$$(Ml)(e) \le \frac{\mathrm{diam}_l(G,\Gamma)d^2}{|E|}\mathrm{cong}(e). \tag{9}$$

In many cases, the right-hand side of Eq. 9 provides a rather accurate estimate on Ml. For example, for d-ary trees, it is off by a factor of order d^2. Thus, by using the length function *cong* that assigns cong(e) to e, we expect to get a good bound on $\mu_0(\Gamma)$. By Eq. 9,

$$\kappa(\Gamma, cong) \le \frac{\mathrm{diam}_{cong}(G,\Gamma)d^2}{|E|}.$$

Thus,

Proposition 4. *For the nearest-neighbor walk on a graph with maximum degree d,*

$$\tau \le \frac{diam_{cong}(G,\Gamma)d^2}{|E|}.$$

3.3 Examples

Proposition 4 often gives a good bound on τ, as shown in the examples below.

d-ary tree. The congestion of an edge at distance i from the root is at most $(n-1)d^{h-i}/(d-1)$. Thus, $\mathrm{diam}_{cong}(T) \le 2(n-1)d^{h+1}/(d-1)^2$. By Proposition 4, it follows that $\tau \le d^{h+3}/(d-1)^2$, which is off from the exact value by a factor of order d^2.

Elongated grid. Consider the subgraph of an $(au+1)\times(bv+1)$–grid induced on the vertices (i, j), with $0 \le i \le au$ and $0 \le j \le bv$, such that i is multiple of a or j is multiple of b. We establish a path from (i, j) to (i', j') by first going to $(i, \lfloor j/b \rfloor b)$, then to $(\lfloor i'/a \rfloor a, \lfloor j/b \rfloor b)$, then to $(\lfloor i'/a \rfloor a, j')$, then to (i', j'). Each time, we use a geodesic (or empty) path. A simple calculation shows that the congestion of a horizontal edge is at most $u(a+b)n$, where n is the number of vertices, and the congestion of a vertical edge is at most $v(a+b)n$. Thus, $\mathrm{diam}_{cong}(G,\Gamma) \leq (au^2+bv^2)(a+b)n$. Since $|E| \geq 2n$, Proposition 4 shows that $\tau \leq 8(au^2+bv^2)(a+b)$.

On the other hand, let $\phi \in L_0^2(V)$ be the function that assigns $2i - au$ to (i,j). Then the numerator in the variational expression of τ is at least $a^2u^2/2$, and the denominator is $8auv/|E|$, and so $\tau \geq au|E|/(16v) \geq au^2(a+b)/8$. Similarly, $\tau \geq bv^2(a+b)/8$, and so $\tau \geq (au^2+bv^2)(a+b)/16$. Thus, τ is of order $(au^2+bv^2)(a+b)$. Here again, we have made no attempt to optimize the constants in the calculation.

Note that the bound on τ given by Eqs. 1 and 2 (using the same paths) may be of different order than τ. For example, when $u=b=1$ and $a=v^2$, τ is of order v^4, but the bound given by Eqs. 1 and 2 is of order v^5.

Double-grid. Consider two $a \times a$ grids having one common origin O. We use Proposition 4 to show that $\tau = \Theta(a^2 \log a)$. To see this, we start by showing that there exist geodesic paths that connect every point in the double-grid to O such that the congestion of an edge at distance h from O is $O(a^2/(h+1))$. This can be done by backward induction on the distance from O. Indeed, label the nodes in one grid by (i,j), $0 \leq i,j \leq a-1$, the origin O being labeled $(0,0)$. Starting with the node $(a-1,a-1)$, recursively route an $i/(i+j)$ fraction of the paths traversing (i,j) to $(i-1,j)$, and the remaining fraction to $(i,j-1)$. Rounding is done in favor of the first set of paths. If c_h is the maximum congestion of a node at distance h from O then

$$c_{h-1} \leq \frac{h+1}{h} c_h + 2,$$

which implies that $c_h \leq 4a^2/(h+1)$, as desired. By connecting every pair of points by first connecting them to O, the congestion of an edge at distance h from O becomes of order $a^4/(h+1)$. Proposition 4 then implies that $\tau = O(a^2 \log a)$. A lower bound of order $a^2 \log a$ follows using the variational expression of τ and the function that assigns $\log(h+1)$ to the vertices at distance h from O in one specified grid, and $-\log(h+1)$ to their symmetric counterparts. Diaconis and Saloff-Coste [2] have obtained independently the same estimate on τ using a different approach.

4 A Semidefinite Relaxation of $\mu_0(\Gamma)$

The results and definitions in the preceding sections can be generalized in an obvious manner by replacing paths with flows of unit value, along [14]. The use of flow will allow us to apply general results from optimization of convex functions, as shown below. Let F be the set of non-negative real matrices Γ whose rows are indexed by $(V \times V) - \{(x,x) : x \in V\}$ and columns indexed by E, and such that each row indexed by (x,y) represents a flow in G from x to y with unit value. In other words, $\Gamma_{(x,y)}$ is a convex combination of the indicator functions of simple paths in G from x to y. As before, let $\mu_0(\Gamma)$ be the largest eigenvalue of $M = \mathcal{Q}^{-1}\Gamma^t \Pi \Gamma$. For a non-negative length function l on E, denote by $\mathrm{dist}_l(x, y)$ the length according to l of a shortest path from x to y.

Lemma 5. *If C is a closed convex set of non-negative matrices in $R^{p\times q}$, then*

$$\min_{B\in C} ||B|| = \max_{l\in(\mathcal{R}^+)^q:||l||\leq 1} \min_{B\in C} ||Bl||, \tag{10}$$

where $||B|| = \max_{l\in\mathcal{R}^q:||l||\leq 1} ||Bl||$.

Proof. The right-hand side of Eq. 10 is upper-bounded by the left-hand side by weak duality. We show that equality holds.

First, assume that any matrix in C is positive, i.e, all entries are strictly positive. Since C is closed convex, the norm function attains its minimum on C at some matrix B_0. Since the matrix $B_0^t B_0$ is positive, it follows from the theory of non-negative matrices [13] that it has a unique (positive) largest eigenvalue. Therefore, the norm function is differentiable at B_0, since $||B||$ is equal to the square root of the largest eigenvalue of $B^t B$. Moreover, its gradient at B_0 is equal to uv^t, where $||u|| = ||v|| = 1$ and $B_0 v = ||B_0|| u$ (see, e.g., [18].) Since the norm function and C are convex, it follows that $\mathrm{tr}((B - B_0)vu^t) \geq 0$ for all $B \in C$. Indeed, if a convex function f (the norm function, in this case) attains its minimum on a convex set C at x_0, and if f is differentiable at x_0, then $(x - x_0)\cdot \nabla_{x_0} f \geq 0$, for any x in C. Here the dot product of two matrices B_1 and B_2 is $\mathrm{tr}(B_1 B_2^t)$. But $\mathrm{tr}(B_0 vu^t) = ||B_0||\mathrm{tr}(uu^t) = ||B_0||$, and $\mathrm{tr}(Bvu^t) \leq ||Bv||\,||u|| = ||Bv||$. Thus, $||Bv|| \geq ||B_0||$. Since $||B_0 v|| = ||B_0||$, it follows that $\min_{B\in C} ||Bv|| = ||B_0||$. We conclude that Eq. 10 holds under the assumption that any matrix in C is positive (clearly, l can be assumed to be non-negative.) The general case can be reduced to the previous case by compactness and by considering the sets $C_i = C + (1/i)\mathbf{1}$, where i is a positive integer. Indeed, if l_i is a vector that maximizes the right-hand side of Eq. 10 for the set C_i, and l is the limit of a subsequence of l_i, it is not hard to show that $\min_{B\in C} ||B|| = \min_{B\in C} ||Bl||$.

We note that Eq. 10 may not hold if C is an arbitrary closed convex set of matrices in $R^{p\times q}$. For example, if C is the set of $p \times p$ matrices whose trace is 1, then the left-hand side is equal to $1/p$, but the right-hand side is equal to 0.

Theorem 6. *Let $\mu = \min_{\Gamma\in F} \mu_0(\Gamma)$. Then $\tau \leq \mu$. For any $\epsilon > 0$, μ can be computed within a factor of $1 + \epsilon$ in polynomial time in n, $\log(1/\epsilon)$ and the bitsize of the input (the transition probabilities). Moreover,*

$$\mu = \max_l \frac{\sum_{(x,y)\in V\times V, x\neq y} \pi(x)\pi(y)\mathrm{dist}_l(x, y)^2}{\sum_{e\in E} Q(e) l(e)^2}, \tag{11}$$

where l ranges over the set of non-negative symmetric length functions on E.

Proof. A proof similar to that of Theorem 2 shows that $\tau \le \mu_0(\Gamma)$, for any $\Gamma \in F$. This is because $(\Gamma(\delta(\phi)))(x,y) = \phi(x) - \phi(y)$. Hence $\tau \le \mu$. Given $\mathcal{Q}$ and Π, the problem of minimizing $\mu_0(\Gamma) = ||\sqrt{\Pi}\Gamma\sqrt{\mathcal{Q}}^{-1}||^2$ can be reduced to a semidefinite program since, as noted in e.g. [17], $||B|| \le t$ if and only if the matrix

$$\begin{pmatrix} tI & B \\ B^t & tI \end{pmatrix}$$

is positive semidefinite. Under general conditions that can be easily checked here, semidefinite programs can be solved at an arbitrary precision in polynomial time [9]. Thus, μ can be computed at an arbitrary precision in polynomial time using, e.g., the ellipsoid method.

On the other hand,

$$\begin{aligned}
\mu &= \min_{\Gamma \in F} ||\sqrt{\Pi}\Gamma\sqrt{\mathcal{Q}}^{-1}||^2 \\
&= \max_{l:||l|| \le 1} \min_{\Gamma \in F} ||\sqrt{\Pi}\Gamma\sqrt{\mathcal{Q}}^{-1} l||^2 \\
&= \max_{l:||\sqrt{\mathcal{Q}}l|| \le 1} \min_{\Gamma \in F} ||\sqrt{\Pi}\Gamma l||^2 \\
&= \max_{l:||\sqrt{\mathcal{Q}}l|| \le 1} \sum_{(x,y) \in V \times V, x \ne y} \pi(x)\pi(y)\mathrm{dist}_l(x,y)^2 \\
&= \max_l \frac{\sum_{(x,y) \in V \times V, x \ne y} \pi(x)\pi(y)\mathrm{dist}_l(x,y)^2}{\sum_{e \in E} Q(e)l(e)^2}.
\end{aligned}$$

The second equation follows from Lemma 5. The fourth equation follows from the fact that a flow from x to y is a convex combination of simple paths from x to y. By convexity of the norm function, there exists $\Gamma \in F$ that minimizes $||\sqrt{\Pi}\Gamma\sqrt{\mathcal{Q}}^{-1}||$ and such that $\Gamma_{(x,y),(u,v)} = \Gamma_{(y,x),(v,u)}$. By the proof of Lemma 5 and the remark following Proposition 3, we can assume that l is symmetric.

Note that, in general, it is easier to compute eigenvalues than to solve semidefinite programs. However, whereas other related quantities associated with a Markov chain such as minimum bisection are NP-hard to compute [8], μ can be computed at an arbitrary precision in polynomial time.

It is easy to show directly that τ is upper-bounded by the right-hand side of Eq. 11. This can be seen by considering $l(x,y) = |\phi(x) - \phi(y)|$, where ϕ is an eigenvector of P corresponding to λ_1. We now show that the best bound on τ using the Poincaré inequalities is off by a factor of order at most $\log^2 n$. We apply a technique used in [11] in the context of multicommodity flow.

Theorem 7 [11]. *There exists a universal constant c such that, for any metric space (X,d) on n vertices, there exists an embedding ψ of X in R^p, for some integer p, such that*

$$d(x,y) \le ||\psi(x) - \psi(y)|| \le c d(x,y) \log n,$$

for any elements $x, y \in X$.

Theorem 8. *For any irreducible, reversible Markov chain on n states, $\mu \leq c^2 \tau \log^2 n$.*

Proof. By Theorem 6, it suffices to show that, for any non-negative symmetric length function l,

$$\frac{\sum_{(x,y)\in V\times V, x\neq y} \pi(x)\pi(y)\mathrm{dist}_l(x,y)^2}{\sum_{e\in E} Q(e)l(e)^2} \leq c^2 \tau \log^2 n. \tag{12}$$

By continuity, we can assume without loss of generality that l is positive, so that dist_l defines a metric on V. By Theorem 7, it suffices for our needs to show that, for any embedding ψ of V in R^p,

$$\frac{\sum_{(x,y)\in V\times V, x\neq y} \pi(x)\pi(y)||\psi(x)-\psi(y)||^2}{\sum_{(x,y)\in V\times V, x\neq y} Q(x,y)||\psi(x)-\psi(y)||^2} \leq \tau. \tag{13}$$

Let $\phi_i(x)$ be the ith coordinate of $\phi(x)$, $1 \leq i \leq p$. To show Eq. 13 it suffices to prove that, for $1 \leq i \leq p$,

$$\frac{\sum_{(x,y)\in V\times V, x\neq y} \pi(x)\pi(y)(\psi_i(x)-\psi_i(y))^2}{\sum_{(x,y)\in V\times V, x\neq y} Q(x,y)(\psi_i(x)-\psi_i(y))^2} \leq \tau.$$

But this follows immediately from the variational definition of τ. This concludes the proof.

A graph is edge-transitive if, for any $(x,y),(x',y') \in E$, there exists an automorphism of the graph that either takes x to x' and y to y', or that takes x to y' and y to x'.

Corollary 9. *For any reversible Markov chain, we have $\mu \geq D$, where D is the expected squared distance between two random points in G chosen independently according to the stationary distribution. Equality $\mu = D$ occurs for the nearest-neighbor walk on an edge-transitive graph.*

Proof. The inequality $\mu \geq D$ follows from Theorem 6 by considering the unit length function l.

Assume now that the graph is edge-transitive. By the definition of μ and Eq. 4, $\mu \leq \kappa(\Gamma, l)$ for any $\Gamma \in F$ and any l. Furthermore, Diaconis and Stroock [5] produce l (unit length function) and $\Gamma \in F$ (average combination of geodesic paths) so that $\kappa(\Gamma, l) = D$. Thus $\mu \leq D$.

4.1 Examples

We calculate μ for the complete bipartite graph $K_{a,b}$ and, up to a constant factor, for expander graphs, and give an application to the exclusion process.

Complete bipartite graph $K_{a,b}$. This graph is edge-transitive, and so $\mu = D = 5/2 - 1/a - 1/b$, whereas $\tau = 1$ for $a+b \geq 3$. Corollary 9 shows that μ

is equal to $\kappa(\Gamma, l)$, where l is the unit length function, and Γ is determined by assigining the same weight to all shortest paths between two given points. In contrast, in the case $a = 2$, deterministic paths will give an $\Omega(b)$ bound on τ, uniformly in all length functions. This example was treated in [14] for $a = 2$ and the unit length function (see also [3]).

Expander Graphs. A d-regular graph G is a (d, λ_1)-expander if all the eigenvalues of the adjacency matrix of G, besides d, are at most $\lambda_1 d$, where $\lambda_1 < 1$ is a fixed constant. It is known [14, Th. 8] that, for any (d, λ_1)-expander on n vertices, one can establish simultaneously a flow of unit value between every pair of distinct points such that the total amount of flow traversing any given edge is $O(n \log n)$, where the constant behind O depends on λ_1. Moreover, every such flow is a convex combination of paths of length $O(\log n)$. Using the unit length function, we conclude that $\mu_0(\Gamma) = O(\log^2 n)$. A lower bound on μ can be obtained using Corollary 9. Indeed, since most of the vertices in the graph are at distance $\Omega(\log_{d-1} n)$ from any given point, $D = \Omega(\log_{d-1}^2 n)$. Thus $\mu = \Theta(\log^2 n)$, where the constants behind Θ depend on d and λ_1, whereas $\tau = \Theta(1)$. This example shows the sharpness of Theorem 8. Note that the Cheeger inequality provides a constant bound on τ in this case.

The exclusion process. Given a connected graph $G = (V, E)$ and $1 \leq r < n$, the exclusion process is the Markov chain on all r-sets of V defined as follows. If the current state is at A, pick a random element x of A with probability proportional to its degree, pick a neighbor y of x uniformly at random, and move to $A' = A - \{x\} \cup \{y\}$. Note that A' may be equal to A. It is known (see [3] and references therein) that this Markov chain is reversible. Bounds on the mixing rate of the exclusion process were established in [7, 3]. In particular, it was shown in [3] that when G is regular, for any set Γ of canonical paths in G,

$$\tau(r) \leq r\kappa(\Gamma, \mathbf{1}), \tag{14}$$

where $\tau(r)$ is the inverse of the eigenvalue gap for the exclusion process. Here the Markov chain on G is the usual nearest-neighbor walk on G. The proof-technique in [3] shows that in fact, when G is regular, $\tau(r) \leq r\mu$. Together with Theorem 8, this implies that $\tau(r) \leq c^2 r\tau \log^2 n$. Up to a constant factor, this inequality also holds for nearly-regular graphs. Diaconis and Saloff-Coste [3] have raised the question whether $\tau(r) \leq \gamma r\tau$, for some universal constant γ and all graphs. They showed [3] a lower bound $\tau(r) = \Omega(r\tau)$ for nearly-regular graphs.

5 Concluding Remarks

1. Our results can be easily adapted to bound the second smallest eigenvalue of the Laplacian of a graph. The Laplacian is the matrix indexed by the vertices which for $i \neq j$, has a 0 in entry (i, j) if vertices i and j are not connected, and a -1 if they are connected; while on the diagonal, the degrees of the vertices appear in the corresponding order. Essentially, this can be done by replacing $\pi(x)$ by 1 and $Q(e)$ by n in the proofs and results. In particular,

if τ_{lap} is the inverse of the second smallest eigenvalue of the Laplacian (the smallest one is 0), then

$$n\tau_{lap} \leq \max_e \frac{1}{l(e)} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_l,$$

for any positive length function l, and so

$$n\tau_{lap} \leq \mathrm{diam}_{cong}(G, \Gamma).$$

The best bound on τ_{lap} that can be obtained using multicommodity flow and a variable length function is

$$\mu_{lap} = \max_l \frac{\sum_{(x,y)\in V\times V} \mathrm{dist}_l(x,y)^2}{n \sum_{e\in E} l(e)^2},$$

where l ranges over the set of non-negative symmetric length functions. Moreover, $\mu_{lap} \leq O(\log^2 n)\tau_{lap}$.

2. It is easy to show that, for any set Γ of canonical paths and any positive length function l, $\kappa(\Gamma, \mathbf{1}) \leq (\max_{x\neq y} |\gamma_{xy}|)\kappa(\Gamma, l)$. A similar remark holds for any flow matrix Γ. Thus, it is not possible to improve the bound on τ by more than a factor of $\max_{x\neq y} |\gamma_{xy}|$ by using an optimal length function rather than the unit length function. Our birth-death chain and tree examples show that such an improvement can be achieved for some chains.

6 Acknowledgments

The author is grateful to David Aldous, Noga Alon, Stephen Boyd, Persi Diaconis, Jim Fill, Joel Friedman, Stephen Guattery, Gary Miller, Laurent Saloff-Coste, and Alistair Sinclair for helpful discussions.

References

1. D. Aldous. *Reversible Markov Chains and random walks on graphs.* Book in preparation.
2. P. Diaconis and L. Saloff-Coste. Personal Communication, 1995.
3. P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov Chains. *The Annals of Applied Probability*, 3:696–730, 1993.
4. P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? In *27th Annual ACM Symposium on Theory of Computing*, pages 112–129. ACM Press, 1995.
5. P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov Chains. *The Annals of Applied Probability*, 1:36–61, 1991.
6. J. Fill. Unpublished manuscript, July 1990.
7. J. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains with an application to the exclusion process. *The Annals of Applied Probability*, 1:62–87, 1991.

8. M. R. Garey and D. S. Johnson. *Computers and intractability: a guide to the theory of NP–completeness.* Freeman and Company, San Fransisco, 1979.
9. M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.
10. M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. on Comput.*, 18:1149–1178, 1989.
11. N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.
12. L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Strutures & Algorithms*, 4(4):359–412, 1993.
13. E. Seneta. *Non-negative matrices and Markov Chains.* Springer-Verlag, 1981.
14. A. Sinclair. Improved bounds for mixing rates of Markov Chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.
15. A. Sinclair and M. Jerrum. Approximate counting, uniform generation, and rapidly mixing Markov Chains. *Information and Computation*, 82:93–113, 1989.
16. A. D. Sokal. Optimal Poincaré inequalities for the spectra of Markov Chains. Unpublished manuscript, September 1992.
17. L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 1995. To appear.
18. G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and Appl.*, 170:33–45, 1992.

This article was processed using the LaTeX macro package with LLNCS style