Improved bounds for mixing rates of Markov chains and multicommodity flow

Alistair Sinclair
Department of Computer Science, University of Edinburgh
The King's Buildings, Edinburgh EH9 3JZ, Scotland

Appeared in
Combinatorics, Probability & Computing **1** (1992), pp. 351–370.

Part of this work was done while the author was visiting the International Computer Science Institute at Berkeley and the Center for Discrete Mathematics and Theoretical Computer Science at Rutgers University. A preliminary version of this paper appeared in the *Proceedings of the 1st Latin American Symposium on Theoretical Informatics*, São Paulo, Brazil, April 1992 (Springer Lecture Notes in Computer Science Vol. 583).

Abstract

The paper is concerned with tools for the quantitative analysis of finite Markov chains whose states are combinatorial structures. Chains of this kind have algorithmic applications in many areas, including random sampling, approximate counting, statistical physics and combinatorial optimisation. The efficiency of the resulting algorithms depends crucially on the mixing rate of the chain, i.e., the time taken for it to reach its stationary or equilibrium distribution.

The paper presents a new upper bound on the mixing rate, based on the solution to a multicommodity flow problem in the Markov chain viewed as a graph. The bound gives sharper estimates for the mixing rate of several important complex Markov chains. As a result, improved bounds are obtained for the runtimes of randomised approximation algorithms for various problems, including computing the permanent of a 0-1 matrix, counting matchings in graphs, and computing the partition function of a ferromagnetic Ising system. Moreover, solutions to the multicommodity flow problem are shown to capture the mixing rate quite closely: thus, under fairly general conditions, a Markov chain is rapidly mixing if and only if it supports a flow of low cost.

1 Summary

In recent years, Markov chain simulation has emerged as a powerful algorithmic paradigm. Its chief application is to the random sampling of combinatorial structures from a specified probability distribution. Such a sampling procedure lies at the heart of efficient probabilistic algorithms for a wide variety of problems, such as approximating the size of combinatorially defined sets, estimating the expectation of certain operators in statistical physics, and combinatorial optimisation by stochastic search.

The algorithmic idea is simple. Suppose we wish to sample the elements of a large but finite set X of structures from a distribution π. First, construct a Markov chain whose states are the elements of X and which converges asymptotically to the stationary or equilibrium distribution π over X; it is usually possible to do this using as transitions simple random perturbations of the structures in X. Then, starting from an arbitrary state, simulate the chain until it is close to equilibrium; the distribution of the final state will be close to the desired distribution π.

To take a typical example, let H be a connected graph and X the set of spanning trees of H, and suppose we wish to sample elements of X from a uniform distribution. Consider the Markov chain $\mathcal{MC}(H)$ with state space X which, given a spanning tree $T \in X$, makes transitions as follows: select uniformly at random an edge e of H which does not belong to T, add e to T, thereby creating a single cycle C, and finally remove an edge of C uniformly at random to create a new spanning tree T'. It is not hard to check that this Markov chain converges to the uniform distribution over X.

Analysing the efficiency of the above technique in a given application presents a considerable challenge. The key issue is to determine the *mixing rate* of the chain, i.e., the number of simulation steps needed to ensure that it is sufficiently close to its equilibrium distribution π. An efficient algorithm can result only if this number is reasonably small, which usually means dramatically less than the size of the state space X itself. For example, in the spanning tree problem above we would want $\mathcal{MC}(H)$ to reach equilibrium in time bounded by some polynomial in n, the size of the problem instance H; however, the number of states $|X|$ will typically be exponential in n. Informally, we will call chains having this property *rapidly mixing*. (More correctly, this is a property of *families* of chains, such as $\mathcal{MC}(H)$, parameterised on problem instances.)

The classical theory of Markov chains has not been greatly concerned with a quantitative study of the approach to equilibrium. This has led to the development recently of new analytic tools, based on coupling, stopping times and group representation theory, which have been successfully applied to chains with a regular structure such as random walks on certain special graphs or groups. The book by Diaconis [7] gives an excellent survey.

Markov chains arising in the combinatorial applications mentioned above are typically much more complex, however. The first analyses of such chains were

made possible using a quantity called the *conductance* [23, 24]. Suppose we view a (reversible) Markov chain as a weighted graph G, whose vertices are states and whose edges are transitions. Then the conductance $\Phi \equiv \Phi(G)$ is essentially the *edge expansion* of G. Equivalently, Φ may be viewed as the probability that the chain in equilibrium escapes from a subset S of the state space in one step, minimised over small subsets S. (Precise definitions of this and other quantities are given in later sections.) It is intuitively reasonable that Φ should be related to the mixing rate: if the above escape probability is small for some S then the cut edges separating S from $X - S$ constitute a "constriction" or bottleneck which prevents rapid convergence to equilibrium. Conversely, a large value of Φ means that the chain cannot get trapped by any small region of the space, and hence should be rapidly mixing.

A useful piece of technology for obtaining lower bounds on Φ in complex examples was developed in [11, 23]. The idea is to construct a *canonical path* γ_{xy} in the graph G between each ordered pair of distinct states x and y. If the paths can be chosen in such a way that no edge is overloaded by paths, then the chain cannot contain a constriction, so Φ is not too small. (The existence of a constriction between S and $X - S$ would imply that *any* choice of paths must overload the edges in the constriction.)

More precisely, suppose ρ is the maximum loading of an edge by paths; then it is not hard to show (see Theorem 3 of Section 2) that $\Phi \geq (2\rho)^{-1}$, so ρ does indeed provide a bound on the mixing rate of the chain. The power of this observation lies in the fact that a good collection $\Gamma = \{\gamma_{xy}\}$ of canonical paths can sometimes be constructed for which ρ can be bounded rather tightly; indeed, the quantity ρ arises very naturally from a combinatorial encoding technique, as explained in Section 3.

In a recent paper [8], Diaconis and Stroock observed that path arguments similar to that described above can lead *directly* to bounds on the mixing rate, independently of the conductance Φ. In this paper, we present a new bound which is a modification of that of Diaconis and Stroock. The new bound also involves the maximum loading of an edge by paths, but takes into account the lengths of the paths. A simplified form of the bound (Corollary 6 of Section 2) relates the mixing rate to the product $\rho\ell$ for a collection of paths Γ, where ℓ is the length of a longest path in Γ. This bound turns out to be sharper than the conductance-based bound above when the maximum path length ℓ is small compared to ρ.

In Section 3 of the paper, we illustrate the effectiveness of the new bound by obtaining significantly improved estimates for the mixing rate of several important complex Markov chains, which have been used in the design of algorithms for problems involving monomer-dimer systems, matchings in graphs, the Ising model, and almost uniform generation of combinatorial structures. The factors saved in the mixing rate translate directly to the runtime of the algorithms that use the chains. These improvements apparently do not follow from the similar bound given by Diaconis and Stroock because the Markov chains in question have widely differing weights on their edges. (The two bounds are equivalent if the edge weights are uniform, e.g., in the case of random walk on a graph.)

Finally, in Section 4, we address the problem of characterising the rapid mixing property for reversible Markov chains. It is already known that the conductance Φ characterises rapid mixing, in the sense that Φ^{-1} essentially measures the mixing rate up to a polynomial factor (in fact, a square). In view of the foregoing results, it is natural to ask whether a similar characterisation in terms of the path measure ρ also holds. This would mean that whenever a Markov chain is rapidly mixing a proof using a path argument exists.

We are able to answer this question in the affirmative provided the definition of ρ is generalised in a natural way to allow multiple rather than canonical paths between pairs of states. This leads us to consider a multicommodity flow problem in the graph G describing the Markov chain, in which a certain quantity of some commodity (x, y) is to be transported from x to y for all pairs $x, y \in X$. For a given flow, ρ may then be interpreted as the maximum total flow through any edge e as a fraction of its weight, or capacity. Minimising over all possible flows, we get a quantity which we call the *resistance* $\rho \equiv \rho(G)$ of the Markov chain.

The main result of this section states that, if a reversible Markov chain is close to equilibrium after τ steps, then its resistance cannot exceed $O(\tau)$. Thus the resistance, like the conductance, does indeed characterise the rapid mixing property. We also observe that the quantities Φ^{-1} and ρ are in fact equal up to a factor $O(\log N)$. This is actually an approximate max-flow min-cut theorem for the multicommodity flow problem, and is a natural generalisation of a result obtained in a different context by Leighton and Rao [17].

2 Bounds on the mixing rate

We assume familiarity with the elementary theory of Markov chains: see, e.g., [15] for a more detailed treatment. Let X be a finite set, and P the transition matrix of a discrete-time Markov chain on state space X. We assume throughout that P is irreducible (i.e., that all states communicate) and *reversible* with respect to the probability distribution π on X, i.e., it satisfies the detailed balance condition

$$Q(x, y) \equiv \pi(x)P(x, y) = \pi(y)P(y, x) \qquad \text{for all } x, y \in X. \tag{1}$$

Condition (1) implies that π is a stationary or equilibrium distribution for P, i.e., $\pi P = \pi$. If in addition P is aperiodic, the distribution of the state at time t converges pointwise to π as $t \to \infty$, regardless of the initial state. In this case the chain is called *ergodic*. Simulating an ergodic chain for sufficiently many steps starting from an arbitrary initial state, and noting the final state, provides an algorithm for sampling elements of X from a distribution that is arbitrarily close to π.

We note that the above framework is quite general for the purposes of the combinatorial applications mentioned in the previous section. In particular, it is usually a straightforward matter to make condition (1) hold using as transitions simple random perturbations of the structures in X, such as those employed in the spanning tree example given earlier.

It is convenient to identify a reversible Markov chain with a weighted undirected graph G on vertex set X, with an edge of weight $Q(x,y)$ connecting vertices x and y iff $Q(x,y) > 0$. (Thus the graph may contain self-loops.) Note that this graph is always connected and uniquely specifies the chain.

As is well known, P has real eigenvalues $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N-1} \geq -1$, where $N = |X|$; P is ergodic iff $\lambda_{N-1} > -1$. For an ergodic chain, the rate of convergence to π is governed by the second-largest eigenvalue in absolute value, $\lambda_{max} = \max\{\lambda_1, |\lambda_{N-1}|\}$. To make this statement precise, let x be the state at time $t = 0$ and denote by $P^t(x,\cdot)$ the distribution of the state at time t. The *variation distance* at time t with initial state x is

$$\Delta_x(t) \ = \ \max_{S \subseteq X} |P^t(x,S) - \pi(S)| \ = \ \tfrac{1}{2} \sum_{y \in X} |P^t(x,y) - \pi(y)|.$$

We will measure rate of convergence using the function τ_x defined for $\epsilon > 0$ by

$$\tau_x(\epsilon) \ = \ \min\{t : \Delta_x(t') \leq \epsilon \text{ for all } t' \geq t\}.$$

Proposition 1 *The quantity $\tau_x(\epsilon)$ satisfies*

(i) $\tau_x(\epsilon) \leq (1 - \lambda_{max})^{-1} \left(\ln \pi(x)^{-1} + \ln \epsilon^{-1} \right)$;

(ii) $\max_{x \in X} \tau_x(\epsilon) \geq \frac{1}{2} \lambda_{max} (1 - \lambda_{max})^{-1} \ln(2\epsilon)^{-1}$. $\square$

Part (i) follows from [8, Proposition 3] and gives an upper bound on the time to reach equilibrium from a given initial state x in terms of λ_{max} and $\pi(x)$. The converse, part (ii), which is a discrete-time version of [1, Proposition 8], says that convergence cannot be rapid unless λ_{max} is bounded away from 1. (Note that in the latter bound there is a maximisation over initial states: it is possible for a chain to converge fast from certain states even when λ_{max} is close to 1. However, even if such a state exists, finding it requires more detailed information about the chain than is usually available in the more complex examples of interest to us.) Results analogous to Proposition 1 hold for measures other than the variation distance. For example, [23, 24] give bounds in terms of the relative pointwise distance, defined by $\Delta_x^{rpd}(t) = \max_{y \in X} |P^t(x,y) - \pi(y)|/\pi(y)$.

In the remainder of this paper, we will ignore the technical issues arising from the choice of initial state. Proposition 1 then shows that we can identify the rapid mixing property with a large value of the *spectral gap* $1 - \lambda_{max}$. Moreover, in practice the smallest eigenvalue λ_{N-1} is unimportant: a crude approach is to add a holding probability of $\frac{1}{2}$ to every state, i.e., replace P by $\frac{1}{2}(I + P)$, where I is the $N \times N$ identity matrix. This ensures that all eigenvalues are non-negative while decreasing the spectral gap $1 - \lambda_1$ only by a factor of 2. We therefore focus attention on the second eigenvalue λ_1.

As indicated in the previous section, the first upper bounds on λ_1 for complex Markov chains were based on the *conductance* [23, 24], defined by

$$\Phi \equiv \Phi(G) = \min_{\substack{S \subset X \\ 0 < \pi(S) \le 1/2}} \frac{Q(S, \overline{S})}{\pi(S)}, \tag{2}$$

where G is the weighted graph describing the chain and $Q(S, \overline{S})$ denotes the sum of $Q(x, y)$ over edges $\{x, y\}$ in G with $x \in S$ and $y \in \overline{S} = X - S$. The conductance may be viewed as a weighted version of the *edge expansion* of G. Alternatively, since $Q(S, \overline{S}) = \sum_{x \in S, y \in \overline{S}} \pi(x) P(x, y)$, the quotient in (2) is just the conditional probability that the chain in equilibrium escapes from the subset S of the state space in one step, given that it is initially in S. Thus Φ measures the ability of the chain to escape from any small region of the state space, and hence to make rapid progress to equilibrium. The following result formalising this intuition is from [23, 24]; see also [2, 3, 4, 16, 19, 21] for related results.

Theorem 2 *The second eigenvalue λ_1 of a reversible Markov chain satisfies*

$$1 - 2\Phi \le \lambda_1 \le 1 - \frac{\Phi^2}{2}. \qquad \square$$

Note that Φ characterises the rapid mixing property: a Markov chain is rapidly mixing, in the sense of the previous section, if and only if $\Phi \ge 1/poly(n)$, where n is the problem size.

In order to apply Theorem 2, it is necessary to estimate the conductance Φ. Since we are usually more interested in positive results, lower bounds on Φ are generally of greater interest and we focus on them for most of the rest of this paper. (We shall consider negative results in Section 4.) In some cases such a bound can be obtained directly, using elementary arguments [16, 24] or geometric ideas [9, 14]. However, in many important applications the only known handle on Φ is via the canonical path approach sketched in the previous section. Thus we attempt to construct a family $\Gamma = \{\gamma_{xy}\}$ of simple paths in G, one between each ordered pair of distinct states x and y, such that no edge is overloaded by paths. The maximum loading of any edge is measured by the quantity

$$\rho \equiv \rho(\Gamma) = \max_e \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y), \tag{3}$$

where the maximum is over oriented edges e in G (i.e, transitions of the Markov chain), and $Q(e) = Q(u, v)$ if $e = (u, v)$. Note that we may view the Markov chain as a flow network, in which $\pi(x)\pi(y)$ units of flow travel from x to y along γ_{xy}, and $Q(e)$ plays the role of the capacity of e. The quantity ρ then measures the maximum flow along any edge as a fraction of its capacity. We shall pursue this analogy further in Section 4.

The following simple result confirms our intuition that the existence of a good choice of paths should imply a large value for the conductance.

Theorem 3 *For any reversible Markov chain, and any choice of canonical paths,*

$$\Phi \geq \frac{1}{2\rho}.$$

Proof: Let $S \subset X$ be a subset with $0 < \pi(S) \leq \frac{1}{2}$ which minimises the quotient $Q(S, \overline{S})/\pi(S)$. For any choice of paths, the total net flow crossing the cut from S to $\overline{S}$ is $\pi(S)\pi(\overline{S})$; moreover, the aggregated capacity of the cut edges (x, y), with $x \in S$ and $y \in \overline{S}$, is just $Q(S, \overline{S})$. Hence there must exist a cut edge e with

$$\frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \geq \frac{\pi(S)\pi(\overline{S})}{Q(S, \overline{S})} \geq \frac{\pi(S)}{2Q(S, \overline{S})} = \frac{1}{2\Phi}. \qquad \square$$

Theorems 2 and 3 immediately yield the following bound on λ_1:

Corollary 4 *For any reversible Markov chain, and any choice of canonical paths, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \leq 1 - \frac{1}{8\rho^2}. \qquad \square$$

In recent work [8], Diaconis and Stroock observed that bounds on λ_1 can be obtained directly in terms of canonical paths, without appealing to the conductance bound of Theorem 2. This latter bound is potentially rather weak because of the appearance of the square, so a direct approach may lead to sharper estimates for λ_1. We now present a modified version of Diaconis' and Stroock's bound which is apparently more useful than theirs in many combinatorial applications. In the next section, we will illustrate the effectiveness of the bound by obtaining improved estimates for the second eigenvalue of several important Markov chains.

To state the new bound, we modify the measure ρ to take into account the lengths of the paths. For a given collection $\Gamma = \{\gamma_{xy}\}$ of canonical paths, the key quantity is now

$$\overline{\rho} \equiv \overline{\rho}(\Gamma) = \max_e \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)|\gamma_{xy}|, \tag{4}$$

where $|\gamma_{xy}|$ denotes the length (i.e., number of edges) of the path γ_{xy}.

Theorem 5 *For any reversible Markov chain, and any choice of canonical paths, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \leq 1 - \frac{1}{\overline{\rho}}.$$

Proof: Let $L = I - P$, so that the eigenvalues of L are $\mu_i = 1 - \lambda_i$. Following [8], the variational characterisation of μ_1 is

$$\mu_1 = \inf_\psi \frac{\sum_{x,y \in X} (\psi(x) - \psi(y))^2 Q(x, y)}{\sum_{x,y \in X} (\psi(x) - \psi(y))^2 \pi(x)\pi(y)}, \tag{5}$$

where the infimum is taken over all non-constant functions $\psi : X \to \mathbb{R}$. (The constant functions are the only eigenfunctions of L with eigenvalue $\mu_0 = 0$.) Now for any ψ, and any choice of canonical paths Γ, the denominator of (5) may be written as follows:

$$\begin{aligned}
\sum_{x,y} (\psi(x) - \psi(y))^2 \pi(x)\pi(y) &= \sum_{x,y} \pi(x)\pi(y) \Big(\sum_{e \in \gamma_{xy}} (\psi(e^+) - \psi(e^-)) \Big)^2 \\
&\leq \sum_{x,y} \pi(x)\pi(y)|\gamma_{xy}| \sum_{e \in \gamma_{xy}} (\psi(e^+) - \psi(e^-))^2 \\
&= \sum_{e} (\psi(e^+) - \psi(e^-))^2 \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)|\gamma_{xy}| \\
&\leq \sum_{e} (\psi(e^+) - \psi(e^-))^2 Q(e)\overline{\rho}(\Gamma) \\
&= \overline{\rho}(\Gamma) \sum_{x,y} Q(x,y) (\psi(x) - \psi(y))^2 .
\end{aligned}$$

Here e^- and e^+ denote the start and end vertices of the oriented edge e, and the first inequality is Cauchy-Schwarz. The result now follows from (5). □

The following simplified form of Theorem 5 is often useful.

Corollary 6 *For any reversible Markov chain, and any choice of canonical paths Γ, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \;\leq\; 1 - \frac{1}{\rho \ell},$$

where $\ell \equiv \ell(\Gamma)$ is the length of a longest path in Γ. □

Corollary 6 may be applied in the same situations as Corollary 4, by constructing paths and estimating the quantity ρ. Frequently, however, the maximum path length ℓ will be significantly less than the estimate obtained for ρ; in such cases, Corollary 6 will give a sharper bound than Corollary 4. The improved bounds presented in the next section are all based on this observation.

Remark: Diaconis and Stroock [8] give a bound which is similar to that of Theorem 5 but which uses a different measure of path length. To get their bound we replace $|\gamma_{xy}|$ in the definition (4) of $\overline{\rho}$ by the quantity

$$|\gamma_{xy}|_{Q,e} \;=\; \sum_{e' \in \gamma_{xy}} \frac{Q(e)}{Q(e')},$$

with everything else defined as before. Let $\overline{\rho}_{DS}$ be the measure obtained in this way. Diaconis' and Stroock's bound [8, Proposition 1], may be stated as

$$\lambda_1 \leq 1 - \overline{\rho}_{DS}{}^{-1}. \tag{6}$$

The examples in the next section indicate that $\overline{\rho}$, or $\rho\ell$, may be a more useful quantity to work with in practice that $\overline{\rho}_{DS}$. The reason seems to be that $\overline{\rho}$ has a more "local" nature than $\overline{\rho}_{DS}$. To see this, note that the contribution of a path $\gamma_{xy} \ni e$ to $\overline{\rho}$ is just

$$Q(e)^{-1}|\gamma_{xy}|\pi(x)\pi(y),$$

which depends only on the path length and on $Q(e)$, while its contribution to $\overline{\rho}_{DS}$ is

$$Q(e)^{-1}|\gamma_{xy}|_{Q,e}\,\pi(x)\pi(y) \;=\; \sum_{e'\in\gamma_{xy}} Q(e')^{-1}\pi(x)\pi(y),$$

which depends on the capacities $Q(e')$ of all path edges. As we shall see, this makes the former quantity easier to use in applications where weights play a significant role, so that the capacities $Q(e)$ vary considerably: the problem with $\overline{\rho}_{DS}$ is that a path γ_{xy} may pass through edges of very small capacity, so that $|\gamma_{xy}|_{Q,e}$ is much larger than $|\gamma_{xy}|$.
If the Markov chain under consideration is random walk on a graph $G = (X, E)$ then $Q(e) = 1/2|E|$ for all e, so $|\gamma_{xy}|_{Q,e} = |\gamma_{xy}|$; hence the quantities $\overline{\rho}$ and $\overline{\rho}_{DS}$ coincide in this case. Most of the examples discussed by Diaconis and Stroock [8] are in fact random walks on graphs, so our bound yields identical results for them. Note also that $|\gamma_{xy}|_{Q,e} \geq 1$ for all γ_{xy} and e, so certainly $\overline{\rho}_{DS}$ is bounded below by ρ. Hence the bounds of Theorem 5 and Corollary 6 can be worse than (6) by at most a factor ℓ. Moreover, there are examples for which $\overline{\rho}$ is provably significantly better than $\overline{\rho}_{DS}$; one such is the Ehrenfest urn model, discussed by Diaconis and Stroock [8]. However, the two quantities seem to be incomparable in general. $\square$

The examples in the next section and in [8] indicate that $\overline{\rho}$ frequently leads to sharper bounds on λ_1 than does ρ itself. By way of contrast, here is a simple example where Corollary 4 provably does better than Theorem 5 and Corollary 6. Consider asymmetric random walk on the line $[0, N-1]$ with reflecting barriers, i.e., $X = \{0, 1, \ldots, N-1\}$ and the transition probabilities are given by $P(i-1, i) = \beta$, $P(i, i-1) = 1-\beta$ for $0 < i < N$, and $P(0,0) = 1-\beta$, $P(N-1, N-1) = \beta$, where $\beta \in (0, \frac{1}{2})$ is a constant. This chain is reversible and ergodic, with stationary distribution $\pi(i) \propto r^i$, where $r = \beta/(1-\beta)$. In this case there is a unique simple path between each pair of states i and j, so our choice of canonical paths is forced. Elementary calculations show that the quantity $Q(e)^{-1}\sum_{\gamma_{xy}\ni e}\pi(x)\pi(y)$ is maximised on the edge $e = (\lceil\frac{N}{2}\rceil - 1, \lceil\frac{N}{2}\rceil)$ and that its value there is $\frac{1+r}{1-r}(1 + O(r^{N/2}))$. Hence Corollary 4 gives the bound

$$\lambda_1 \leq 1 - \frac{(1-r)^2}{8(1+r)^2}\left(1 + O(r^{N/2})\right). \tag{7}$$

The value of λ_1 for this chain is known exactly: it is $2(\beta(1-\beta))^{1/2}\cos(\frac{\pi}{N}) = \frac{2r^{1/2}}{1+r}(1 + O(N^{-2}))$. Hence (7) differs from the true value asymptotically by only a constant factor. On the other hand, a similar calculation considering the edge $(N-2, N-1)$ shows that $\overline{\rho} \geq (1+r)N + O(1)$. Thus Theorem 5 gives the bound

$$\lambda_1 \leq 1 - \frac{1}{(1+r)N} + O\left(\frac{1}{N^2}\right),$$

which is asymptotically much worse than (7).

3 Applications

In this section we discuss a series of complex Markov chains used in combinatorial applications whose mixing rate is currently estimated using the conductance-based bounds of Theorem 2 or Corollary 4. In each case, we indicate the improvement in the lower bound on the spectral gap $1-\lambda_1$ obtained using Corollary 6. By Proposition 1, this translates directly to a similar improvement in the mixing rate. As the precise arguments are combinatorially delicate, our present treatment will necessarily be very sketchy. For full details, the reader is urged to consult the stated references.

The sharpened analysis of these Markov chains is of interest in its own right and is our main concern here. However, it also leads to improved estimates for the runtimes of various polynomial-time algorithms that make use of the chains. In fact, since the runtime is dominated by the time needed to sample some number of structures from the stationary distribution, each algorithm is immediately speeded up by exactly the factor saved in the spectral gap. (The runtimes of the algorithms can be readily computed from the spectral gap and are not given explicitly here; details may be found in the references.) These improvements, though significant, are in most cases not sufficient to make the algorithms genuinely practical for large inputs. However, they do represent a tightening of the most intricate part of the analysis. There is undoubtedly room for refinement of other aspects of these algorithms, but such an investigation is beyond the scope of this paper.

(i) The monomer-dimer or all-matchings chain

Let $H=(V,A)$ be a weighted graph with positive edge weights $\{c(a) : a \in A\}$, and consider the Markov chain whose state space X consists of all matchings in H, i.e., subsets $M \subseteq A$ such that no two edges in M share an endpoint. Transitions from M are made as follows: select an edge $a=\{u,v\}$ of A uniformly at random, and then

(i) if $a \in M$, move to $M-a$ with probability $1/(1+c(a))$;

(ii) if u and v are both unmatched in M, move to $M+a$ with probability $c(a)/(1+c(a))$;

(iii) if $a' = \{u,w\} \in M$ for some w, and v is unmatched in M, move to $(M+a)-a'$ with probability $c(a)/(c(a)+c(a'))$;

(iv) in all other cases, do nothing.

It is easy to check using (1) that this chain is reversible with stationary distribution $\pi(M) \propto w(M)$, where $w(M) = \prod_{a \in M} c(a)$ is the *weight* of matching M. Simulating the chain therefore enables one to sample matchings randomly with probabilities approximately proportional to their weights. This has several important applications to the design of polynomial-time approximation algorithms for hard combinatorial enumeration problems. In statistical physics, H describes a monomer-dimer system whose *partition function* is given by

$$Z(H) \;=\; \sum_{M} w(M). \tag{8}$$

Weighted sampling of matchings enables Z to be estimated accurately. The special case of (8) in which all edge weights are 1 corresponds to counting all matchings in H. Moreover, by varying the edge weights in H in a suitable fashion and sampling matchings as above, it is possible to estimate the number of matchings in H of a given size. In particular, for most graphs the number of *perfect matchings* can be estimated, a problem which corresponds to evaluating the permanent of a 0-1 matrix. Approximate counting of various other structures may be reduced to this problem [12, 20]. Finally, weighted sampling of matchings also enables a matching of nearly maximum cardinality to be found with high probability, an example of stochastic search by simulated annealing. Details of these applications may be found in [11, 23].

In order for the resulting algorithms to be efficient (in the sense of having polynomially bounded runtime), accurate sampling must be possible in time bounded by a polynomial in the size of H and $c_{max} = \max\{1, \max_{a \in A} c(a)\}$. By Proposition 1, this requires a bound on second eigenvalue of the form $\lambda_1 \le 1 - 1/poly(|H|, c_{max})$. We now present a brief sketch of the canonical path argument used to obtain such a bound. In doing so, our aim will be to illustrate how the quantity ρ arises naturally from a combinatorial encoding technique. For the details the reader is referred to [11, 23].

Let I and F be matchings in H, and consider the symmetric difference $S = I \oplus F$. The connected components of S are paths and cycles in H whose edges belong alternately to I and F. The canonical path γ_{IF} from I to F is determined as follows:

- order the components of S according to a fixed underlying ordering on the paths and cycles of H;

- "unwind" each component, removing edges of I and adding edges of F using transitions of the Markov chain, in an obvious way.

Now let $e = (M, M')$ be an arbitrary oriented edge (transition) in the graph describing the Markov chain, and denote by $paths(e)$ the set of paths which pass through e. The key idea is to enumerate $paths(e)$ using the states of the chain themselves. Specifically, we set up an *injective mapping*

$$\sigma_e : paths(e) \to X,$$

so that each $\gamma_{IF} \in paths(e)$ is encoded by a unique matching $\sigma_e(I, F)$. Moreover, we do this in a way that preserves weights, i.e.,

$$w(M)w(\sigma_e(I, F)) \approx w(I)w(F). \tag{9}$$

(Essentially, we just take $\sigma_e(I, F)$ to be the complement of M in the multiset $I \cup F$, though we have to take care to ensure that $\sigma_e(I, F)$ is indeed a matching.) Now summing (9) over pairs I, F such that $\gamma_{IF} \in paths(e)$, and recalling that $\pi(\cdot) \propto w(\cdot)$, we get

$$\sum_{\gamma_{IF} \ni e} \pi(I)\pi(F) \approx \pi(M) \sum_{\gamma_{IF} \ni e} \pi(\sigma_e(I, F)) \leq \pi(M), \tag{10}$$

since σ_e is injective. But $Q(e) = \pi(M)P(M, M')$, so (10) gives us an upper bound on the crucial quantity $Q(e)^{-1} \sum_{\gamma_{IF} \ni e} \pi(I)\pi(F)$, and hence on ρ.

Precisely, the bound derived in [11, 23] by this method is $\rho \leq 4|A|c_{max}^2$. Corollary 4 therefore yields

$$\lambda_1 \leq 1 - 1/128|A|^2 c_{max}^4 .$$

On the other hand, the maximum length of any canonical path is easily seen to be at most $|V| = n$, so Corollary 6 gives the much sharper bound

$$\lambda_1 \leq 1 - 1/4n|A|c_{max}^2 .$$

The improvement in the spectral gap $1 - \lambda_1$, and hence in the mixing rate and the runtime of the algorithms mentioned above, is a factor of $32|A|c_{max}^2 n^{-1}$. In the application to approximating the permanent, the largest value of c_{max} is the ratio of the number of "near-perfect" matchings to the number of perfect matchings in H. (A near-perfect matching is a matching in which precisely two vertices of H are unmatched.) This quantity is at least $n/2$, and can be quite large in interesting cases: for example, for dense graphs (with minimum vertex degree at least $n/2$), the ratio is about n^2 and $|A| \geq n^2/2$, leading to an improvement of $O(n^5)$; and [11] gives a bound on the ratio of n^{10} for random graphs of low density. (The ratio can in fact be exponentially large, but then the chain no longer converges in polynomial time.)

(ii) Broder's chain for the dense permanent

This chain, which was proposed in [5] and analysed in [11, 23], is a restricted version of Example (i); it again allows the number of perfect matchings in a graph to be estimated in polynomial time provided the ratio of the number of near-perfect matchings to the number of perfect matchings is polynomially bounded. Let H

be an (unweighted) graph with n vertices; the states of the chain are all perfect matchings and all near-perfect matchings in H. Transitions are made in similar fashion to Example (i) but without weights; the stationary distribution is uniform. Using canonical paths similar to those in Example (i), and the same encoding technique, it can be shown that $\rho = O(n^6)$, whence $\lambda_1 \leq 1 - O(n^{-12})$ by Corollary 4. However, since the maximum path length is at most $2n$, Corollary 6 yields the sharper bound $\lambda_1 \leq 1 - O(n^{-7})$. The mixing rate is therefore reduced by a factor $O(n^5)$. This is exactly the same improvement as that discussed in Section 4 of [8]: in this case the Diaconis-Stroock bound (6) is equivalent to Theorem 5 because there are no weights, i.e., $Q(e)$ is uniform.

(iii) The Ising model

In this example drawn from statistical physics, the states of the Markov chain are all subgraphs of the graph (V, A) of interactions of a ferromagnetic Ising system, i.e., all graphs (V, A') where $A' \subseteq A$. (These graphs arise from the so-called high-temperature expansion of the partition function.) Transitions are made by random addition or subtraction of individual edges with appropriate probabilities. The stationary distribution assigns to each subgraph the weight $\lambda^j \mu^k$, where $\lambda, \mu \in (0,1)$ are parameters of the system, and j, k are respectively the number of edges and the number of odd-degree vertices in the subgraph. By sampling from this distribution, various important quantities, such as the *partition function* of the system, can be effectively approximated; the details are in [13].

In [13] a choice of canonical paths is presented for which it can be shown, again using the encoding technique sketched in Example (i), that $\rho \leq 2|A|\mu^{-4}$. This leads to the bound $\lambda_1 \leq 1 - \mu^8/32|A|^2$, from Corollary 4. The length of paths here is at most $|A|$, so Corollary 6 yields the sharper bound $\lambda_1 \leq 1 - \mu^4/2|A|^2$. The improvement in the spectral gap is therefore a factor $16\mu^{-4}$. In the applications discussed in [13], the parameter μ is taken down to n^{-1}, where $n = |V|$ is the number of sites in the system. Hence the improvement in the runtime is a factor $O(n^4)$.

(iv) Approximate counting and uniform generation

The Markov chain considered here is of a different flavour from those of Examples (i)–(iii). It is based on a tree which reflects an inductive construction of a class of combinatorial structures; the structures themselves correspond to leaves of the tree. The transition probabilities are determined by weights attached to the edges of the tree, which in turn are derived from crude estimates of the number of structures in the subtree below the edge. Simulation of the Markov chain allows the structures to be sampled from an almost uniform distribution, and indirectly enables one to bootstrap the crude counting estimates to arbitrarily precise estimates of the number of structures. For the details and some applications, see [23, 24].

In [23, 24] a direct argument gives the bound $\Phi \geq (4r^2d)^{-1}$ for the conductance, where d is the depth of the tree and $r \geq 1$ is the error factor allowed in the crude counting estimates. This in turn yields, by Theorem 2, $\lambda_1 \leq 1 - (32r^4d^2)^{-1}$. On the other hand, using (the only possible) canonical paths we get $\rho \leq 8r^2d$ and

$\ell \leq 2d$, which by Corollary 6 implies $\lambda_1 \leq 1 - (16r^2d^2)^{-1}$. The improvement in the spectral gap is thus a factor $2r^2$.

4 Multicommodity flow

In this section we present a natural generalisation of the path-counting ideas of Section 2. We consider a multicommodity flow problem in the graph G describing a reversible Markov chain, and obtain upper bounds on λ_1 in terms of a measure on flows which is analogous to the measure ρ of Section 2. Moreover, there is also a matching *lower* bound on λ_1 in terms of this measure, so that it, like the conductance Φ, captures the mixing rate of a Markov chain rather closely.

As in Section 2, let G be the weighted graph describing a reversible Markov chain with stationary distribution π. Let us view G as a flow network by assigning to each oriented edge e of G the capacity $Q(e)$. Now imagine that, for each ordered pair of distinct vertices x and y, a quantity $\pi(x)\pi(y)$ of some commodity (x, y) is to be transported from x to y along the edges of the network. The object is to construct a flow which minimises the total throughput through any oriented edge e as a fraction of its capacity $Q(e)$. This is entirely analogous to our previous measure ρ, except that we are now allowing multiple paths between states rather than canonical paths. Thus it is natural to suppose that our new measure will yield similar bounds on the mixing rate.

Formally, a *flow* in G is a function $f : \mathcal{P} \to \mathbb{R}^+$ which satisfies

$$\sum_{p \in \mathcal{P}_{xy}} f(p) = \pi(x)\pi(y) \qquad \text{for all } x, y \in X, \quad x \neq y,$$

where $\mathcal{P}_{xy}$ is the set of all simple directed paths from x to y in G and $\mathcal{P} = \bigcup_{x \neq y} \mathcal{P}_{xy}$. Now extend f to a function on oriented edges by setting

$$f(e) = \sum_{p \ni e} f(p),$$

i.e., $f(e)$ is just the total flow routed by f through e. By analogy with the definition (3) of ρ, the quality of a flow f is measured by the quantity $\rho(f)$, which is the maximum value over oriented edges e of the ratio $f(e)/Q(e)$.

Theorem 3 and Corollary 4 carry over immediately to this more general setting.

Theorem 3′ *For any reversible Markov chain, and any flow f,*

$$\Phi \geq (2\rho(f))^{-1}. \qquad \square$$

Corollary 4′ *For any reversible Markov chain, and any flow f, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \leq 1 - \frac{1}{8\rho(f)^2}. \qquad \square$$

In order to generalise the measure $\overline{\rho}$ from Section 2 to a flow f, define a function $\overline{f}$ on oriented edges by

$$\overline{f}(e) = \sum_{p \ni e} f(p)|p|,$$

where $|p|$ is the number of edges in the path p. (We may think of $\overline{f}(e)$ as the *elongated flow* through e.) Now set $\overline{\rho}(f) = \max_e \overline{f}(e)/Q(e)$. The proof of Theorem 5 carries over almost unchanged, and the analogue of Corollary 6 is then immediate.

Theorem 5′ *For any reversible Markov chain, and any flow f, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \leq 1 - \frac{1}{\overline{\rho}(f)}.$$

Corollary 6′ *For any reversible Markov chain, and any flow f, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \leq 1 - \frac{1}{\rho(f)\ell(f)},$$

where $\ell(f)$ is the length of a longest path p with $f(p) > 0$. □

There are examples in which the extra flexibility provided by flows (as opposed to canonical paths) is necessary in order to achieve good bounds on the mixing rate. Consider random walk on the complete bipartite graph $K_{2,N-2}$, with vertex set $X = \{0, 1, \ldots, N-1\}$ and edges $\{0, i\}$, $\{1, i\}$ for $i = 2, 3, \ldots, N-1$, in which transitions from each vertex are made by choosing a neighbour uniformly at random. (This process is periodic, but can be made ergodic by adding a holding probability of 1/2 to every vertex.) In the stationary distribution each vertex occurs with probability proportional to its degree, and $Q(e) = 1/4(N-2)$ for all edges e. It is easy to construct a flow f with $\rho(f) = O(1)$, by distributing flow evenly over all shortest paths between each pair of vertices. By Corollary 4′ this gives an estimate for the spectral gap which is correct to within a constant factor. However, since $\pi(0)\pi(1) = 1/16$, it is clear that the best value for ρ (or $\overline{\rho}$) obtainable using canonical paths is $\Omega(N)$, leading to the weak bound $\lambda_1 \leq 1 - \Omega(N^{-2})$ (or $\lambda_1 \leq 1 - \Omega(N^{-1})$).

Here is a further example that illustrates the use of Theorem 5′. Consider the Bernoulli-Laplace diffusion model, whose state space X is the set of all k-element subsets of $[n] = \{0, 1, \ldots, n-1\}$. Transitions are made from a given subset $x \in X$ by selecting uniformly at random an element i of x and an element j of $[n] - x$ and replacing i by j in x. The stationary distribution here is uniform, $\pi(x) = N^{-1}$ for all $x \in X$, where $N = \binom{n}{k}$. Now let x, y be distinct elements of X, with $|x \oplus y| = 2m$. We define a flow f by routing $(N^2 m!^2)^{-1}$ units of flow from x to y along each of the $m!^2$ shortest paths (of length m) from x to y.

(Each such path corresponds to an ordering of the m elements of $x-y$ and an ordering of the m elements of $y-x$.)

Now let $e=(z,z')$ be an arbitrary transition, with $z'=z\cup\{j\}-\{i\}$. To bound the flow through e, we again use the encoding technique sketched in Example (i) of Section 3. Let $paths(e)$ denote the set of paths $p\ni e$ with $f(p)>0$. We define a many-to-one mapping $\sigma_e : paths(e)\to X$ as follows: if p is a path from x to y, set $\sigma_e(p)=x\oplus y\oplus z'$. Note that $\sigma_e(p)\oplus z'=x\oplus y$, so all paths p with a given image under σ_e have the same length m and carry the same flow $(N^2 m!^{\,2})^{-1}$, where $2m=|\sigma_e(p)\oplus z'|$. Moreover, the number of such paths is

$$\sum_{r=0}^{m-1}\binom{m-1}{r}^2 r!^{\,2}(m-r-1)!^{\,2} \;=\; m(m-1)!^{\,2}.$$

(Here r corresponds to the distance along the path from x to z.) Thus the total elongated flow through e contributed by these paths is $m^2(m-1)!^{\,2}(N^2m!^{\,2})^{-1}=N^{-2}$. Finally, summing over images $\sigma_e(p)$, and noting that the range of σ_e consists of all subsets that contain i and do not contain j, we see that

$$\overline{f}(e)=\frac{1}{N^2}\binom{n-2}{k-1}=\frac{k(n-k)}{Nn(n-1)}.$$

But since $Q(e)=(Nk(n-k))^{-1}$ for all edges e, we deduce that $\overline{\rho}\le\frac{k^2(n-k)^2}{n(n-1)}$. Theorem $5'$ therefore yields the bound $\lambda_1\le 1-\frac{n(n-1)}{k^2(n-k)^2}$. In the case $n=2k$, the exact value is $\lambda_1=1-2/k$, so the estimate is correct within a factor of about $k/2$. It seems difficult to get this close using canonical paths. A similar bound was obtained by a slightly different method in [8]. The above idea of using all geodesic paths first appeared in the analysis of a Markov chain on matchings by Dagum *et al* [6]. $\square$

By analogy with the conductance Φ, we may define the *resistance*† of a reversible Markov chain described by a graph G by

$$\rho\equiv\rho(G)=\inf_f\rho(f),$$

where the infimum is taken over all valid flows in G. Corollary $4'$ indicates that ρ provides a lower bound on the spectral gap of the form $\Omega(\rho^{-2})$. Thus in particular a family of Markov chains will be rapidly mixing if ρ is bounded above by a polynomial in the problem size. From Theorem 2 we already know that Φ characterises the rapid mixing property, since it measures the spectral gap up to a square. It is natural to ask whether ρ provides a similar characterisation.

†This terminology is chosen because, as we shall see shortly, ρ is almost the inverse of the conductance Φ. It should not be confused with the resistance of a graph familiar from electrical network theory.

This is indeed the case. We will show that, if a reversible Markov chain is close to equilibrium after τ steps, then it supports a flow f for which $\rho(f) = O(\tau)$, where $N = |X|$ is the number of states. Therefore, the chain is rapidly mixing *if and only if* ρ is bounded above by a polynomial in the problem size.

In order to state this result (Theorem 8), we need to formalise the notion of the time τ for a chain to become close to equilibrium. With τ_x defined as in Section 2, set $\tau = \max_{x\in X} \tau_x(1/4)$, i.e., τ is the time taken for the variation distance from an arbitrary initial state to fall to 1/4. As will become clear, the value 1/4 is not significant and could be replaced by any sufficiently small constant. First we need a simple technical lemma.

Lemma 7 *For any $t \geq 2\tau$ and all $x, y \in X$,*

$$\frac{P^t(x,y)}{\pi(y)} \geq \frac{1}{8}.$$

Proof: Fix arbitrary states $x, y \in X$, and consider the set $Z = \{z \in X : \frac{P^\tau(y,z)}{\pi(z)} \geq \frac{1}{2}\}$. It is not hard to see that $\pi(Z) \geq \frac{1}{2}$, which by definition of τ ensures that $P^t(x, Z) \geq \frac{1}{4}$ for all $t \geq \tau$.
Now we have, for any $t \geq 2\tau$,

$$\begin{aligned}
P^t(x,y) &\geq \sum_{z\in Z} P^{t-\tau}(x,z)P^\tau(z,y) \\
&= \pi(y) \sum_{z\in Z} P^{t-\tau}(x,z)\,\frac{P^\tau(y,z)}{\pi(z)} \\
&\geq \frac{\pi(y)}{2} \sum_{z\in Z} P^{t-\tau}(x,z) \\
&\geq \frac{\pi(y)}{8},
\end{aligned}$$

where in the second line we have used the fact that $\pi(y)P^\tau(y,z) = \pi(z)P^\tau(z,y)$, which follows from reversibility. □

Theorem 8 *The resistance of an ergodic reversible Markov chain satisfies*

$$\rho \leq 16\tau,$$

where τ is defined as above.

Proof: We show how to construct a flow f with the stated bound on $\rho(f)$. Let $t = 2\tau$. For a given state x, the choice of paths used to carry flow from x to other states is determined by the t-step evolution of the Markov chain itself, starting at x. More precisely, let $\mathcal{P}_{xy}^{(t)}$ denote the set of all (not necessarily simple) paths of length exactly t from x to y in the underlying graph G, and for $p \in \mathcal{P}_{xy}^{(t)}$ let

$prob(p)$ denote the probability that the Markov chain, starting in state x, makes the sequence of transitions defined by p. Note that Lemma 7 guarantees that $\mathcal{P}_{xy}^{(t)}$ is non-empty for all x, y. Now for each x, y and $p \in \mathcal{P}_{xy}^{(t)}$, set

$$f(p) \;=\; \frac{\pi(x)\pi(y)prob(p)}{\sum_{p\in\mathcal{P}_{xy}^{(t)}} prob(p)} \;=\; \frac{\pi(x)\pi(y)prob(p)}{P^t(x,y)}, \tag{11}$$

and set $f(p) = 0$ for all other paths $p \in \mathcal{P}$. Note that $\sum_{p\in\mathcal{P}_{xy}^{(t)}} f(p) = \pi(x)\pi(y)$ for all x, y. Strictly speaking, f is not a flow according to our definition since the paths in $\mathcal{P}_{xy}^{(t)}$ are not necessarily simple; however, we can always obtain a flow f' from f, without increasing the throughput through any edge, by simply bypassing the cycles on all paths. We now proceed to estimate $\rho(f')$.
From (11), the flow routed by f' through e is

$$f'(e) \;\le\; \sum_{x,y} \sum_{\substack{p\in\mathcal{P}_{xy}^{(t)}\\ p\ni e}} \frac{\pi(x)\pi(y)prob(p)}{P^t(x,y)} \;\le\; 8\sum_{x,y} \sum_{\substack{p\in\mathcal{P}_{xy}^{(t)}\\ p\ni e}} \pi(x)prob(p), \tag{12}$$

where the second inequality follows from Lemma 7. Now the final double sum in (12) is precisely the probability that the Markov chain, when started in the stationary distribution π over X, traverses the oriented edge e within t steps. But this probability is at most $tQ(e)$, since the probability that the stationary process traverses e in any one step is precisely $Q(e)$. Combining this observation with (12) yields

$$\rho(f') \;=\; \max_e \frac{f'(e)}{Q(e)} \;\le\; 8t \;=\; 16\tau,$$

as required. $\square$

Remarks: (a) An analogous result in terms of the elongated flow measure $\overline{\rho}$ also holds: since the proof of Theorem 8 uses only paths of length $t = 2\tau$, we have constructed a flow f for which $\overline{\rho}(f) = O(\tau^2)$.
(b) We have stated Theorem 8 in terms of the variation distance, for consistency with our earlier approach. It should be clear that similar formulations in terms of other measures are possible. For example, define $\Delta^{rpd}(t) = \max_{x,y} |P^t(x,y) - \pi(y)|/\pi(y)$, the relative pointwise distance at time t maximised over initial states x. Then the proof of Theorem 8 shows that

$$\rho \le \left(1 - \Delta^{rpd}(t)\right)^{-1} t,$$

provided t is large enough that $P^t(x,y) > 0$ for all $x, y \in X$. A similar result has been observed independently by Jim Fill [10]. $\square$

In fact, a direct comparison of Φ and ρ sheds some interesting light on these two quantities. It is convenient at this point to introduce a symmetrised version of Φ, namely

$$\Phi' = \min_{\substack{S \subset X \\ 0<\pi(S)<1}} \frac{Q(S,\overline{S})}{\pi(S)\pi(\overline{S})}.$$

Clearly $\Phi \leq \Phi' \leq 2\Phi$, so Φ and Φ' differ by at most a constant factor. Inspection of the proof of Theorem 3 reveals that the marginally stronger bound

$$\Phi' \geq \rho^{-1} \tag{13}$$

also holds. The reader will recognise this bound as nothing other than the trivial direction of the max-flow min-cut theorem for our multicommodity flow problem: the net flow across any cut cannot exceed the capacity of the cut.† In view of the well-known result for single commodity flows, one might ask whether equality holds in (13). We have already seen an example where $\Phi' = \rho^{-1}$, namely the asymmetric random walk of Section 2. The reason for this is that the underlying graph G is a tree, so there is a unique valid flow f, and it is easy to see that the only cuts we need consider in the definition of Φ' are single edges. Hence we have

$$\Phi' = \min_e \frac{Q(e)}{\sum_{x\in S}\pi(x)\sum_{y\in\overline{S}}\pi(y)} = \min_e \frac{Q(e)}{f(e)} = \rho^{-1},$$

where $(S,\overline{S})$ is the partition of X induced by the cut edge e.

The above question was extensively studied in more generality, and in a totally unrelated context, by Matula and Shahrokhi [18, 22]. The determination of ρ is a case of what Matula and Shahrokhi call the Maximum Concurrent Flow Problem, while computing Φ' is a case of the Sparsest Cut Problem. Matula and Shahrokhi show that these two problems are "near-duals" of one other, and make this statement precise. They call graphs for which equality holds in (13) *bottleneck graphs*, and identify some examples. In our language, these include tree processes (i.e., Markov chains whose underlying graph G is a tree), and random walks on complete graphs, cycles and cubes. More significantly, they also exhibit examples for which equality definitely does not hold. To see that this can happen, consider the random walk on $K_{2,N-2}$ discussed earlier. It is not hard to verify that $\Phi' = 1 + O(N^{-2})$ (and $\Phi' = 1$ when N is even), but that the resistance $\rho = \frac{5}{4} + O(N^{-1})$. Asymptotically, therefore, ρ exceeds Φ'^{-1} by a factor $\frac{5}{4}$. Thus we are led to ask by how much ρ can exceed Φ'^{-1} in general.

This question was addressed, again in a different context, by Leighton and Rao [17]. They show that, in the case of uniform flow between all pairs of vertices

---

†Note that our problem can be recast in more conventional flow maximisation terms as follows: determine the maximum value of F such that $F\pi(x)\pi(y)$ units of commodity (x,y) can be transported from x to y, for every pair x,y, and such that the flow through any edge e does not exceed its capacity $Q(e)$. The maximum such F is precisely ρ^{-1}.

(i.e., in our language, the stationary distribution π is uniform: $\pi(x) = N^{-1}$ for all $x \in X$), ρ cannot exceed Φ^{-1} by more than a factor $O(\log N)$. This yields an approximate max-flow min-cut theorem for the (uniform) multicommodity flow problem. Moreover, this bound is tight since $\rho = \Omega(\frac{\log N}{\Phi})$ for random walk on an expander graph of constant degree: here Φ is constant, but $\rho = \Omega(\log N)$ since the distance between most pairs of vertices is $\Omega(\log N)$ and the graph has only $\Theta(N)$ edges. As has been observed by Éva Tardos, Leighton and Rao's result actually holds in the more general situation where the flow between each pair x, y is of the form $v(x)v(y)$ for any fixed function $v : X \to \mathbb{R}^+$. In our setting, this leads to the following result for arbitrary stationary distributions; since the proof is essentially the same as that of [17, Theorem 1] we omit it here.

Theorem 9 *For any reversible Markov chain with N states,*

$$\rho = O\left(\frac{\log N}{\Phi}\right). \qquad \Box$$

Putting together Theorem 9 and Theorem 2, we obtain a lower bound on the second eigenvalue in terms of ρ.

Corollary 10 *For any reversible Markov chain with N states, the second eigenvalue λ_1 satisfies*

$$\lambda_1 \geq 1 - O\left(\frac{\log N}{\rho}\right). \qquad \Box$$

As far as bounds on the mixing rate are concerned, Theorem 8 is rather stronger than Corollary 10: Theorem 8 says that $\tau = \Omega(\rho)$, whereas Corollary 10, in conjunction with Proposition 1(ii), gives the weaker bound $\tau = \Omega(\rho/\log N)$. However, Theorem 9 and Corollary 10 seem to be of interest in their own right.

Acknowledgements

It is a pleasure to thank David Aldous, Persi Diaconis, Satish Rao, Alan Sokal, Éva Tardos and Umesh Vazirani for helpful discussions on the ideas presented here, and Mark Jerrum both for helpful discussions and for his comments on an earlier version of the paper.

References

[1] Aldous, D. Some inequalities for reversible Markov chains. *Journal of the London Mathematical Society (2)* **25** (1982), pp. 564–576.

[2] Aldous, D. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probability in the Engineering and Informational Sciences* **1** (1987), pp. 33–46.

[3] Alon, N. Eigenvalues and expanders. *Combinatorica* **6** (1986), pp. 83–96.

[4] Alon, N. and Milman, V.D. λ_1, isoperimetric inequalities for graphs and superconcentrators. *Journal of Combinatorial Theory Series B* **38** (1985), pp. 73–88.

[5] Broder, A.Z. How hard is it to marry at random? (On the approximation of the permanent). *Proceedings of the 18th ACM Symposium on Theory of Computing,* 1986, pp. 50–58. Erratum in *Proceedings of the 20th ACM Symposium on Theory of Computing*, 1988, p. 551.

[6] Dagum, P., Luby, M., Mihail, M. and Vazirani, U.V. Polytopes, permanents and graphs with large factors. *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science* (1988), pp. 412–421.

[7] Diaconis, P. *Group representations in probability and statistics.* Lecture Notes Monograph Series Vol. 11, Institute of Mathematical Statistics, Hayward, California, 1988.

[8] Diaconis, P., and Stroock, D. Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability* **1** (1991), pp. 36–61.

[9] Dyer, M., Frieze, A. and Kannan, R. A random polynomial time algorithm for approximating the volume of convex bodies. *Proceedings of the 21st ACM Symposium on Theory of Computing* (1989), pp. 375–381.

[10] Fill, J. Unpublished manuscript.

[11] Jerrum, M. R. and Sinclair, A. J. Approximating the permanent. *SIAM Journal on Computing* **18** (1989), pp. 1149–1178.

[12] Jerrum, M. R. and Sinclair, A. J. Fast Uniform Generation of Regular Graphs. *Theoretical Computer Science* **73** (1990), pp. 91–100.

[13] Jerrum, M. R. and Sinclair, A. J. Polynomial-time approximation algorithms for the Ising model. Technical Report CSR-1-90, Dept. of Computer Science, University of Edinburgh. To appear in *SIAM Journal on Computing*; Extended Abstract in *Proceedings of the 17th International Colloquium on Automata, Languages and Programming* (1990), Springer LNCS Vol. 443, pp. 462–475.

[14] Karzanov, A. and Khachiyan, L. On the conductance of order Markov chains. Technical Report DCS 268, Rutgers University, June 1990.

[15] Keilson, J. *Markov chain models — rarity and exponentiality.* Springer-Verlag, New York, 1979.

[16] Lawler, G.F. and Sokal, A.D. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American Mathematical Society* **309** (1988), pp. 557–580.

[17] Leighton, T. and Rao, S. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation

algorithms. *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science* (1988), pp. 422–431.

[18] Matula, D. W. and Shahrokhi, F. Sparsest cuts and bottlenecks in graphs. *Discrete Applied Mathematics* **27** (1990), pp. 113–123.

[19] Mihail, M. Conductance and convergence of Markov chains: a combinatorial treatment of expanders. *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science* (1989), pp. 526–531.

[20] Mihail, M. and Winkler, P. On the number of Eulerian orientations of a graph. *Proceedings of the 3rd ACM-SIAM Symposium on Discrete Algorithms* (1992), pp. 138–145.

[21] Mohar, B. Isoperimetric numbers of graphs. *Journal of Combinatorial Theory, Series B* **47** (1989), pp. 274–291.

[22] Shahrokhi, F. and Matula, D. W. The maximum concurrent flow problem. *Journal of the ACM* **37** (1990), pp. 318–334.

[23] Sinclair, A.J. Algorithms for random generation and counting: a Markov chain approach. PhD Thesis, University of Edinburgh, June 1988. To appear as a monograph in the series *Progress in Theoretical Computer Science*, Birkhäuser, Boston, 1992.

[24] Sinclair, A.J. and Jerrum, M.R. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* **82** (1989), pp. 93–133.