Probability Theory and Related Fields manuscript No.
(will be inserted by the editor)

N. Fountoulakis · B.A. Reed

Faster Mixing and Small Bottlenecks

Received: 2 November 2005 / Revised version: 17 February 2006

Abstract. We prove a new bound on the mixing time of a Markov chain by considering the conductance of its connected subsets.

1. Introduction

Probabilistic algorithms which use sampling are extremely important tools. Perhaps the oldest example of such an algorithm is the Metropolis-Hastings algorithm [22], which is used for sampling from a given distribution on a finite set. M. Dyer, A. Frieze and R. Kannan showed that using random sampling we can determine the volume of a convex body up to a $1 + o(1)$ factor in polynomial time (with an exponentially small probability of error) (see [7] and [16], [17], [18] for further improvements). In contrast, in d-dimensional space a deterministic algorithm to approximate the volume to within a factor of 2^d requires exponentially many samples. Another application is the estimation of the permanent of a matrix by M. Jerrum and A. Sinclair in [13], where perfect matchings on a suitably defined graph are randomly sampled. Random sampling is also extremely important in approximating the partition function of various statistical mechanics models, in polynomial-time. This is done for the Ising model by M. Jerrum and A. Sinclair in [14]. Moreover, probabilistic algorithms are used to sample from the Gibbs distribution of a statistical mechanics model, such as the Ising model (see [27]) or the Potts model at zero temperature, via the random sampling of proper colourings of a graph (see [8]). For further references the interested reader is referred to the survey [26].

In all the examples cited above, the size of the probability space from which we sample is exponentially large in terms of the input size. Thus, choosing the random sample is itself a non-trivial algorithmic task. A common technique is to define a graph on the sample space and carry out a random walk on this graph. If we choose the graph carefully, then after our walk has made enough steps, the element of the sample space we are currently at can be used as a random sample. In order for this technique to be useful, we need to bound the number of steps we are required to take

School of Computer Science, McGill University, 3480 University Street, Montreal, PQ, Canada, H3A 2A7, e-mail: {fountoul,breed}@cs.mcgill.ca.

Research supported by NSERC.

to ensure that the current state is sufficiently random. This is the theory of *rapidly mixing Markov chains.*

A (finite) Markov chain $\mathcal{M}$ can be thought of as a digraph G on vertex set $V_n = \{1, \dots, n\}$ (this is the set of states) together with a collection of probabilities $p_{ij} \geq 0$ that we move to state j given that we are at state i, such that $p_{ij} > 0$ precisely if ij is an edge of G. Thus, for every $i \in V_n$ we have $\sum_j p_{ij} = 1$. Let $P = (p_{ij})_{i,j \in V_n}$ be the *transition matrix.* If we start at vertex i, then the state that the Markov chain has reached after the t-th step has distribution $e_i P^t$, where e_i is the vector with a 1 in the ith coordinate and 0 elsewhere. We use P_i^t to denote this distribution.

If we want to sample from a distribution π using a walk started at an arbitrary state then it must be the case that for every i, $P_i^t \to \pi$ as $t \to \infty$, where the convergence here is with respect to the *total variation distance* d_{TV} between two probability distributions on V_n defined as:

$$d_{TV}\left(p^{(1)}, p^{(2)}\right) = \max_{A \subseteq V_n} \left| p^{(1)}(A) - p^{(2)}(A) \right|,$$

If such a distribution exists then it is called the *stationary distribution* for $\mathcal{M}$. It is easy to see that in this case, π must be the unique distribution satisfying $\pi P = \pi$.

It is natural to consider chains for which every state has a non-zero stationary probability, such chains are called *ergodic.* Clearly, the underlying digraph G for such a chain must be strongly connected. Furthermore, there cannot be any $k > 1$ such that every cycle of G has length 0 mod k, for then $P_i^t(i) = 0$ unless k divides t. In fact, a Markov chain is ergodic precisely if these two conditions are satisfied, i.e., if it is *irreducible* and *aperiodic.*

We focus on chains which are *reversible*, in the sense that in the stationary distribution, the probability of moving from i to j equals the probability of moving from j to i. That is, $\pi(i)p_{ij} = \pi(j)p_{ji} \quad \forall i, j \in V_n$. In this case, ij is in $E(G)$ precisely if ji is, and so we can think of G as an undirected graph.

The mixing time T_{mix} of Markov chain $\mathcal{M}$ with stationary distribution π is

$$T_{mix} = \sup_i \min\{t \ : \ d_{TV}(P_i^t, \pi) < 1/e\}.$$

It is easy to prove that $\min\{t \ : \ d_{TV}(P_i^t, \pi) < (2/e)^l\} \leq l T_{mix}$. So, T_{mix} measures not only how long it takes to set to within $1/e$ of π, but also bounds how long it takes to get arbitrarily close to π. Thus, it is the standard measure of the rate at which the Markov chain mixes.

Some chains may take a long time to mix because they are nearly periodic. Consider for example a chain for which G is obtained from a digraph in which all cycles are even by adding a loop at some vertex x. Then until we visit x, the probability mass at consecutive time steps is concentrated on disjoint sets, one of which has probability at most $\frac{1}{2}$ in the limit distribution, so we cannot yet have mixed. Thus, the mixing time is at least the time until we hit x, which may be quite a while.

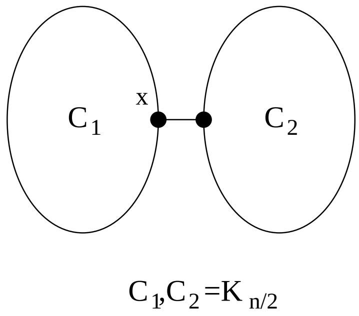

Fig. 1. A bottleneck in a graph

To avoid this problem, we consider a slightly different *lazy* Markov chain where at each time step, we stay where we are with probability $\frac{1}{2}$ and take a step of the original chain with probability $\frac{1}{2}$. Note that this does not change the limit distribution and at most doubles the mixing time. In what follows, we assume implicitly that our chain is lazy.

For lazy reversible Markov chains, the mixing time is essentially determined by bottlenecks like the one depicted in Figure 1. The figure depicts a graph H obtained from two cliques of size $n/2$ by adding an edge xy between them (a clique is a graph in which every two vertices are joined by an edge). We will consider the *simple random walk*, $\mathcal{M}(H)$ on H. This is the Markov chain obtained from H by setting $p_{ij} = 0$ if ij is not an edge of H and $p_{ij} = \frac{1}{|\{j:(i,j)\in E(H)\}|}$ otherwise.

If we start $\mathcal{M}(H)$ in C_1, then we expect to visit x $n/2$ times before we move to C_2 and each time we leave x and enter $C_1 - x$ we expect to take $\frac{n}{2}$ steps before revisiting x. So, we expect to take $n^2/4$ steps before leaving C_1. This implies that there is some vertex y of C_1 such that the probability we leave C_1 within $t = \lfloor \frac{n^2}{40} \rfloor$ steps if we start at y is at most $1/10$ (since otherwise the expected time until we leave C_1 is at most $t\sum_{i=0}^{\infty}(\frac{9}{10})^i$). So $P_y^t(C_1) > \frac{9}{10}$. But symmetry implies that $\pi(C_1) = \frac{1}{2}$ and hence $T_{mix}(\mathcal{M}(H)) > t$.

A bound of the same kind can be obtained by considering any set S of states of a Markov chain. We define $Q(S)$ to be the probability we leave S when we are in the steady state, so $Q(S) = \sum_{i\in S, j\notin S} \pi(i)p_{i,j}$ and $\frac{Q(S)}{\pi(S)}$ is the probability that we leave S given we are in it. Thus, $\frac{\pi(S)}{Q(S)}$ is the expected length of a sojourn in S when we are in the steady state. It follows as above that the mixing time of any Markov chain is at least $\max_{\{S:\ 0<\pi(S)\leq\frac{1}{2}\}} \frac{\pi(S)}{10Q(S)}$.

As in [13], we define the *conductance* of S, denoted $\Phi(S)$,

$$\Phi(S) = \frac{Q(S)}{\pi(S)\pi(V_n \setminus S)}$$

and the conductance $\Phi = \Phi(\mathcal{M})$ of $\mathcal{M}$ to be $\min_{\{S:\ 0<\pi(S)<1\}} \Phi(S)$. We note that for reversible chains, $\Phi(S) = \Phi(V_n \setminus S)$, so

$$T_{mix} \geq \frac{1}{10\Phi}.$$

We remark that $\Phi(S)$ compares the probability of moving from S to $V_n \setminus S$ in the steady state to the probability of doing so if we simply take two independent samples using $\pi(S)$.

At first blush, it might appear that there are other obstacles to rapid mixing. For example, if we have a state i such that the set S of states within distance t of i satisfies $\pi(S) \leq \frac{1}{2}$, then T_{mix} is at least $t+1$ (since $P_i^t(S) = 1$). Thus for a reversible chain, if there are two states i and j such that the shortest walk between them has length at least $2t+1$ then T_{mix} is at least t. I.e., $T_{mix} \geq \frac{diam(G)-1}{2}$ where $diam(G)$ is the diameter of G. However, as we shall see below, if the diameter is high then conductance is low. Specifically, letting $\pi_{min} = \min_{i \in V_n} \pi(i)$, we have $diam(G) = O\left(\Phi^{-1} \log \pi_{min}^{-1}\right)$.

Moreover, M. Jerrum and A. Sinclair proved in [13] that the mixing time of an irreducible, aperiodic and reversible Markov chain satisfies:

$$T_{mix} \leq \frac{C}{\Phi^2} \log \pi_{min}^{-1},$$

for some constant C. Thus, the mixing time of a reversible Markov chain is approximately determined by its conductance.

Treading the path blazed by L. Lovász and R. Kannan in [16], we prove a strengthening of this result which can be used to tie down the mixing time of many Markov chains more precisely. We prove this bound in the next section. We close this section with a statement of our main result. First we give a brief intuition as to why it is true and how it can be useful.

To warm up, we prove the aforementioned upper bound on the diameter in terms of the conductance. For a node i let S_t^i be the set of nodes at exactly distance t from i and let $S_{\leq t}^i$ be the set of nodes at distance at most t from i. Thus, $S_{\leq t+1}^i = S_{\leq t}^i \cup S_{t+1}^i$. Now, whenever we leave $S_{\leq t}^i$ via an edge of G we enter S_{t+1}^i. Furthermore, in equilibrium, $Q(S_{\leq t}^i)$ is the probability that we are in $V_n \setminus S_{\leq t}^i$ but we were in $S_{\leq t}^i$ at the last step. So, $\pi(S_{t+1}^i) \geq Q(S_{\leq t}^i)$, which yields

$$\pi(S_{\leq t+1}^i) = \pi(S_{\leq t}^i) + \pi(S_{t+1}^i) \geq \pi(S_{\leq t}^i)\left(1 + \Phi(S_{\leq t}^i)\pi(V_n \setminus S_{\leq t}^i)\right).$$

If $\pi(S_{\leq t}^i) \leq \frac{1}{2}$ then this yields: $\pi(S_{\leq t+1}^i) \geq (1 + \frac{\Phi}{2})\pi(S_{\leq t}^i)$. Hence,

$$\pi(S_{\leq_t}^i) \geq \min\left\{\frac{1}{2}, \pi_{min}(1 + \frac{\Phi}{2})^t\right\}.$$

In particular, $\pi(S_{\leq t}^i) > \frac{1}{2}$ for t bigger than $2\Phi^{-1} \log\ \pi_{min}^{-1}$ and so the diameter of G is at most $4\Phi^{-1} \log\ \pi_{min}^{-1} + 1$.

We can strengthen this bound by using a more precise estimate of each $\Phi(S^i_{\leq t})$. For $p > \pi_{min}$, we let

$$\Phi(p) = \min_{\substack{S \text{ connected}, \\ \frac{p}{2} \leq \pi(S) \leq p}} \Phi(S)$$

and if there is no such a set we define $\Phi(p) = 1$ (the dual use of letter Φ should cause no confusion). Since the $S^i_{\leq t}$'s are connected sets, we can show: $diam(G) = O\left(\sum_{j=1}^{\lceil \log\ \pi_{min}^{-1} \rceil - 1} \Phi^{-1}(2^j \pi_{min})\right)$. This is the case because for any i, using the above argument we can show that for all $p \leq 1/2$ there are at most $2\Phi^{-1}(p)$ values of t with $\frac{p}{2} \leq \pi(S^i_{\leq t}) \leq p$.

In the same vein, we can strengthen the result of Jerrum and Sinclair, proving:

Theorem 1. *For an irreducible, reversible and aperiodic lazy Markov chain on V_n we have*

$$T_{mix} \leq C \sum_{j=1}^{\lceil \log \pi_{min}^{-1} \rceil} \Phi^{-2}(2^{-j}),$$

for some constant C that does not depend on the chain.

Remark: The above sum can be approximated within a constant factor by the integral $\int_{\pi_{min}}^{1/2} \frac{dx}{x\Phi^2(x)}$. Related results were obtained independently in [15] and [24].

This bound is quite useful if the connectivity faults in the chain are small and widely spaced. For in this case, it is only the terms for large j which contribute significantly to the sum. This is a common situation in random structures, and so our result should be particularly useful in this context. To illustrate this point we state the following theorem which we will prove in a companion paper. It concerns $G_{n,p}$ random graphs on V_n, where each edge is present independently with probability p and, more specifically, the mixing time of a simple random walk on the largest component of $G_{n,p}$, which we denote by $T_{mix}(G_{n,p})$.

Theorem 2. *For every $p = p(n)$ with $1 + \Theta(1) < np$, we have*

$$T_{mix}(G_{n,p}) = \Theta\left(\max\left\{\left(\frac{\ln\ n}{np}\right)^2, \frac{\ln\ n}{\ln\ np}\right\}\right),$$

with probability $1 - o(1)$.

Our bound is also useful if the only sets with low conductance are fairly large, in which case the only large terms in the sum are those for small j. This case was first dealt with by Kannan and Lovász in [16] who proved the variant of our result obtained by replacing $\Phi(p)$ by the minimum of $\Phi(S)$ over all S with $\pi(S) \leq p$. They applied this result to improve the complexity of a volume approximation algorithm. Their result was also used by I. Benjamini and E. Mossel in [3] to tie down precisely the mixing time of a simple random walk on the infinite cluster of a supercritical percolation process on the d-dimensional lattice $\mathbb{Z}^d$. However, this variant is not as widely applicable as our result. In particular if we use it to estimate the mixing time of a simple random walk on the largest component of $G_{n,p}$ we obtain an upper bound on the mixing time which is off by a factor of $\ln\ n$.

We feel that typically it is small sets which have low conductance and hence our variant will extend the reach of this approach significantly.

2. The Main Result

To prove Theorem 1 we actually bound the value of a different invariant of a Markov chain which is within a constant factor of T_{mix}.

We consider starting in some initial state i and performing a random walk stopping exactly in π. In order to do so we will need to allow ourselves to vary the time at which we stop. A *stopping rule* Γ is a procedure for doing so, where after each step we halt with some probability which depends (only) on the sequence of states we have seen so far. We are interested in stopping rules such that the distribution of the halting state is π. In [20], it was shown that such stopping rules exist. For each $i \in V_n$, we define the hitting time time from i to π, denoted $\mathcal{H}(i,\pi)$ to be the minimum over all stopping rules Γ from i to π of the expected number of steps taken by Γ.

The mixing time will be defined as

$$\mathcal{H} = \max_{i \in V_n} \mathcal{H}(i,\pi).$$

D. Aldous has shown in [1] (see also [2]) that this definition of the mixing time is equivalent to the definition of T_{mix} up to some multiplicative constants; i.e. there exist $C_1, C_2 > 0$ such that for every (lazy) Markov chain, $C_1\mathcal{H} \leq T_{mix} \leq C_2\mathcal{H}$. So, to bound T_{mix} from above it is sufficient to establish an upper bound on $\mathcal{H}$ and this is how we proceed in the proof of Theorem 1. We remark that our bound on the mixing time holds regardless of whether or not the chain is lazy, laziness is only needed to prove the equivalence of the mixing time and T_{mix}.

3. Proof of Theorem 1

We choose a state i_0 and stopping rule Γ from i_0 to π whose hitting time is $\mathcal{H}$. The exit frequency of state i, denoted x_i, is the expected number of times state i is exited before the walk halts. By definition, $\mathcal{H} = \sum_{i=1}^{n} x_i$. We prove that

$$\mathcal{H} \leq 28 \sum_{j=1}^{\lceil \log \pi_{min}^{-1} \rceil} \Phi^{-2}(2^{-j}).$$

Theorem 1 follows from the relation between T_{mix} and $\mathcal{H}$. We shall use the notion of the *scaled exit frequency* defined as $y_i = x_i/\pi(i)$ for any $i \in V_n$.

Thus, we have $\mathcal{H} = \sum_{i=1}^{n} \pi(i) y_i$. The scaled exit frequency $y_i = \frac{x_i}{\pi(i)}$ indicates whether or not we are *overvisiting* i. Now, $y_{i_0} \geq \frac{1}{\pi(i_0)}$ so we overvisit i_0. Indeed it turns out, as we show below, that we overvisit i_0 at least as much as any other state. Also crucial to our analysis is the fact, proven in [20], that there is a *halting state* h such that $x(h) = y(h) = 0$. Our proof mimics that of the diameter bound given above. The vertices i_0 and h will play the role of the two furthest apart vertices in that proof. We will grow a nested sequence of connected sets $S_1, \ldots$ around each vertex such that $\pi(S_{i+1}) \geq \pi(S_i)\left(1 + \frac{\Phi(S_i)}{4}\right)$. We bound $\sum_{i=1}^{n} \pi(i) y_i$ by a set of terms each of which is a constant over $\Phi(S_i)$ for some i in the series. The key trick is to choose the right definition for these connected sets. Intuitively, it makes sense that the more we overvisit a vertex, the closer it is to i_0. Indeed, as we show below, the set $S_\alpha = \{i : \ y_i > \alpha\}$ is a connected set which is either empty or contains i_0. The sequence of sets containing i_0 are the S_α's for various values of α. Now, the set $Z_\alpha = \{i : \ y_i < \alpha\}$ may not be connected. So the sets we grow around h are the components of Z_α containing h for various values of α. Forthwith the details.

Assume without loss of generality that the scaled exit frequencies are such that $y_1 \leq \ldots \leq y_n$. By reordering, we can maintain this property whilst ensuring that $i_0 = n$ and $h = 1$. Our proof relies on the following, which is an important consequence of reversibility:

Lemma 1. *If Z is a set of states* not *containing i_0 then*

$$\pi(Z) = \sum_{i \in Z, j \notin Z} \pi(i) p_{ij} (y_j - y_i).$$

Proof. The expected number of times we traverse the edge (i, j) under the stopping rule is $x_i p_{ij}$, or equivalently $y_i \pi(i) p_{ij}$. So, the expected number of times we leave Z is $\sum_{i \in Z, j \notin Z} y_i \pi(i) p_{ij}$ and the expected number of times we enter Z is $\sum_{i \in Z, j \notin Z} y_j \pi(j) p_{ji}$, which by reversibility is $\sum_{i \in Z, j \notin Z} y_j \pi(i) p_{ij}$. Since we start the walk outside Z and halt in π, it follows that

$$\pi(Z) = \sum_{i \in Z, j \notin Z} (y_j - y_i) \pi(i) p_{ij}.$$

The following consequence of the above lemma is used in our proof:

Corollary 1. *For any α the set $S_\alpha = \{i : \ y_i \geq \alpha\}$ either is empty or is a connected set containing i_0. Hence $y_{i_0} \geq y_i \forall i$ as claimed.*

Proof. Let Z be a (non-empty) component of S_α, which does not contain i_0. Then the left-hand side of the above equality is positive, but the right-hand side has only negative terms.

For an $i \in V_n$, we let C_i be the connected component of the subgraph induced by the set of vertices $\{1, \ldots, i\}$ which contains 1 and let $i^* = \min\{i \ : \ \pi(C_i) > 1/2\}$. Let $m_0 = y_1 = 0$ and set $Z_0 = \{1\}$. For $i \geq 1$, we set $m_i = m_{i-1} + \frac{4}{\Phi(Z_{i-1})}$ and $j_i = \max\{j \ : \ y_j \leq m_i\}$. Then we let $Z_i = V(C_{j_i})$. Let $L = \max\{j \ : m_j < y_{i^*}\}$.

We shall bound $\sum_{i=1}^n \pi(i) y_i$, by splitting it into two sums. We first bound $y_{i^*} = \sum_{i=1}^n y_{i^*}\pi(i)$ and then bound $\sum_{i>i^*}(y_i - y_{i^*})\pi(i)$. Clearly, the sum of these two sums is at least $\sum_{i=1}^n y_i\pi(i)$. It is immediate that

$$y_{i^*} \leq m_{L+1} = \sum_{k=0}^{L}(m_{k+1} - m_k) \leq \sum_{k=0}^{L} \frac{4}{\Phi(Z_k)}. \tag{1}$$

To bound the terms in this sum, we must first bound $\pi(Z_{k+1} \setminus Z_k)$. To do so, we use Lemma 1: for $0 \leq k \leq L$

$$\pi(Z_k) = \sum_{i \in Z_k, j \notin Z_k} \pi(i) p_{ij}(y_j - y_i).$$

Now, for any edge ij with $i \in Z_k$, $j \notin Z_{k+1}$, we have $y_j > m_{k+1}$. Thus, if $\bar{A}$ denotes the complement of A and $Q(A,B) = \sum_{i \in A, j \in B} \pi(i) p_{ij}$, then applying Lemma 1 we obtain:

$$\begin{aligned}\pi(Z_k) &= \sum_{i \in Z_k, j \notin Z_k} \pi(i) p_{ij}(y_j - y_i) \geq (m_{k+1} - m_k) \sum_{i \in Z_k, j \notin Z_{k+1}} \pi(i) p_{ij} \\ &= \frac{4}{\Phi(Z_k)}\ Q(Z_k, \bar{Z}_{k+1}) \geq \frac{2\pi(Z_k)}{Q(Z_k, \bar{Z}_k)}\ Q(Z_k, \bar{Z}_{k+1}),\end{aligned}$$

because $\pi(\bar{Z}_k) > 1/2$ by the definition of Z_k. Hence,

$$Q(Z_k, \bar{Z}_{k+1}) \leq \frac{Q(Z_k, \bar{Z}_k)}{2},$$

and therefore

$$Q(Z_k, Z_{k+1} \setminus Z_k) \geq \frac{Q(Z_k, \bar{Z}_k)}{2}.$$

Since $\pi(Z_k) \leq 1/2$, this yields

$$Q(Z_k, Z_{k+1} \setminus Z_k) \geq \frac{\pi(Z_k)\Phi(Z_k)}{4}.$$

In equilibrium, $Q(Z_k, Z_{k+1} \setminus Z_k)$ is the probability that we are in $Z_{k+1} \setminus Z_k$ and that we moved there from Z_k. So, $\pi(Z_{k+1} \setminus Z_k) \geq Q(Z_k, Z_{k+1} \setminus Z_k)$ which implies that

$$\pi(Z_{k+1}) \geq \pi(Z_k)\left(1 + \frac{\Phi(Z_k)}{4}\right).$$

Now for any $l = 1, \ldots, \lceil \log \pi_{min}^{-1} \rceil - 1$, if Z_k is such that $2^{-l-1} \leq \pi(Z_k) \leq 2^{-l}$, then $\Phi(Z_k) \geq \Phi(2^{-l})$. So

$$\pi(Z_{k+1}) - \pi(Z_k) \geq \frac{\pi(Z_k)\Phi(2^{-l})}{4} \geq \frac{2^{-l-1}\Phi(2^{-l})}{4}.$$

Therefore, the number of such Z_k's is no more than

$$4\,\frac{2^{-l-1}}{2^{-l-1}\Phi(2^{-l})}=\frac{4}{\Phi(2^{-l})}.$$

Thus, using (1)

$$y_{i^*}\le\sum_{k=0}^{L}\frac{4}{\Phi(Z_k)}\le 16\sum_{k=1}^{\lceil\log\pi_{min}^{-1}\rceil}\Phi^{-2}(2^{-k}). \tag{2}$$

Now, we shall estimate the sum $\sum_{i>i^*}(y_i-y_{i^*})\pi(i)$ in a similar way. By Corollary 1 the subgraph that is induced by the set $\{i^*+1,\ldots,n\}$ is connected and also $\pi(\{i^*+1,\ldots,n\})<1/2$. Moreover, for every j between i^*+1 and n the graph that is induced by $\{j,\ldots,n\}$ is connected.

Let $n_0=n$ and set $S_0=n$. For $i\ge 1$, we set $S_i=\{n_i,\ldots,n\}$, where $n_i=\min\{j\ :\ \pi(\{j,\ldots,n\})-\pi(j)<\pi(\{n_{i-1},\ldots,n\})\left(1+\frac{\Phi(S_{i-1})}{4}\right)\le\pi(\{j,\ldots,n\})\}$. Let L be such that $n_L\ge i^*+1>n_{L+1}$. Note that

$$\sum_{i>i^*}(y_i-y_{i^*})\pi(i)\le\sum_{k=0}^{L}\sum_{j=n_{k+1}}^{n_k-1}(y_{j+1}-y_j)\pi(\{j+1,\ldots,n\}).$$

Now, for $0\le k\le L$ we obtain:

$$\sum_{j=n_{k+1}}^{n_k-1}(y_{j+1}-y_j)\pi(\{j+1,\ldots,n\})\le\pi(S_{k+1}\setminus\{n_{k+1}\})(y_{n_k}-y_{n_{k+1}}).$$

So,

$$\sum_{i>i^*}(y_i-y_{i^*})\pi(i)\le\sum_{k=0}^{L}\pi(S_{k+1}\setminus\{n_{k+1}\})(y_{n_k}-y_{n_{k+1}}).$$

Furthermore, for each $0\le k\le L$ by Lemma 1

$$\pi(\overline{S}_k)=\sum_{i\in S_k,j\in\bar S_k}\pi(j)p_{ji}(y_i-y_j),$$

and consequently

$$\begin{aligned}\pi(\overline{S}_k)&\ge\sum_{i\in S_k,j\in\bar S_{k+1}\cup\{n_{k+1}\}}\pi(j)p_{ji}(y_i-y_j)\\&\ge(y_{n_k}-y_{n_{k+1}})\sum_{i\in S_k,j\in\bar S_{k+1}\cup\{n_{k+1}\}}\pi(j)p_{ji}\\&=(y_{n_k}-y_{n_{k+1}})\ Q(S_k,\bar S_{k+1}\cup\{n_{k+1}\}).\end{aligned}$$

I.e.

$$y_{n_k}-y_{n_{k+1}}\le\frac{\pi(\overline{S}_k)}{Q(S_k,\bar S_{k+1}\cup\{n_{k+1}\})}.$$

Moreover,

$$\begin{aligned}
Q(S_k,\bar{S}_k) &= Q(S_k,\bar{S}_{k+1}\cup\{n_{k+1}\}) + Q(S_k, S_{k+1}\setminus(\{n_{k+1}\}\cup S_k)) \\
&\leq Q(S_k,\bar{S}_{k+1}\cup\{n_{k+1}\}) + \pi(S_{k+1}\setminus(\{n_{k+1}\}\cup S_k)) \\
&\leq Q(S_k,\bar{S}_{k+1}\cup\{n_{k+1}\}) + \frac{\pi(S_k)\Phi(S_k)}{4} \\
&\leq Q(S_k,\bar{S}_{k+1}\cup\{n_{k+1}\}) + \frac{Q(S_k,\overline{S}_k)}{2},
\end{aligned}$$

and this implies that

$$Q(S_k,\bar{S}_{k+1}\cup\{n_{k+1}\}) \geq \frac{Q(S_k,\bar{S}_k)}{2} > 0. \tag{3}$$

Hence,

$$y_{n_k} - y_{n_{k+1}} \leq \frac{2\pi(\overline{S}_k)}{Q(S_k,\bar{S}_k)}.$$

By the above inequality and the definition of S_{k+1} we obtain

$$\pi(S_{k+1}\setminus\{n_{k+1}\})(y_{n_k} - y_{n_{k+1}}) \leq \pi(S_k)\left(1+\frac{\Phi(S_k)}{4}\right)\frac{2\pi(\overline{S}_k)}{Q(S_k,\bar{S}_k)}.$$

Now, since $Q(S_k)/\pi(S_k)$ is a probability and $\pi(\overline{S_k}) > 1/2$, we have $\Phi(S_k) \leq 2$. This yields:

$$\pi(S_{k+1}\setminus\{n_{k+1}\})(y_{n_k} - y_{n_{k+1}}) \leq \frac{3}{\Phi(S_k)}.$$

Thus,

$$\sum_{i>i^*}(y_i - y_{i^*})\pi(i) \leq \sum_{k=0}^{L}\frac{3}{\Phi(S_k)}.$$

We now argue as above: for any $l = 1,\ldots,\lceil\log\pi_{min}^{-1}\rceil - 1$, if S_k is such that $2^{-l-1} \leq \pi(S_k) \leq 2^{-l}$, then $\Phi(S_k) \geq \Phi(2^{-l})$. Further, by the definition of S_{k+1}, we have:

$$\pi(S_{k+1}) \geq \pi(S_k)\left(1+\frac{\Phi(S_k)}{4}\right).$$

Thus,

$$\pi(S_{k+1}) - \pi(S_k) \geq \frac{\pi(S_k)\Phi(2^{-l})}{4} \geq \frac{2^{-l-1}\Phi(2^{-l})}{4}.$$

So, the number of such S_k's is no more than

$$4\,\frac{2^{-l-1}}{2^{-l-1}\Phi(2^{-l})} = \frac{4}{\Phi(2^{-l})},$$

and therefore we obtain

$$\sum_{i>i^*}(y_i - y_{i^*})\pi(i) \leq \sum_{k=0}^{L} \frac{3}{\Phi(S_k)} \leq 12 \sum_{k=1}^{\lceil \log \pi_{min}^{-1} \rceil} \Phi^{-2}(2^{-k}). \qquad (4)$$

Finally, inequalities (2) and (4) yield the upper bound on $\mathcal{H}$ and the proof of Theorem 1 is complete.

4. Concluding Remarks

We intend this paper to be a jumping off point for future research in two respects.

Firstly, the main theorem can be sharpened significantly at some cost in simplicity, by improving upon the inequality $Q(S_k, S_{k+1} \setminus (\{n_{k+1}\} \cup S_k)) \leq \pi(S_{k+1} \setminus (\{n_{k+1}\} \cup S_k))$. We hope our exposition gives readers enough feeling for the guts of the proof that they can develop their own sharpenings as needed.

Secondly, we have just touched the tip of the iceberg as far as applications are concerned. For example we are currently trying to tie down the mixing time of a random walk on the giant component of a random graph with a fixed degree sequence. Random graphs of this kind were treated by C. Gantzidis, M. Mihail and A. Saberi in [11], in the case where the minimum degree is at least 3. However, as it was shown in [23], degree sequences of more general form may give random graphs having almost surely a giant component. We also intend to have a look at the mixing time of a random walk on the giant component in the random cluster model on the complete graph (see [4]). The authors in [3] raise the question about the mixing time of a simple random walk on the critical cluster of a percolation process on the d-dimensional lattice $\mathbb{Z}^d$, for $d \geq 2$. We believe that the local structure plays a more important role there and our result might be used to answer the above question. We also expect that our main theorem will have applications in a broader setting (in particular to deterministically defined chains).

Acknowledgements. We would like to thank an anonymous referee for suggesting some minor corrections in the manuscript.

References

1. D.J. Aldous, Some inequalities for reversible Markov chains, Journal of the London Mathematical Society **25**, (1982) 564–576.
2. D.J. Aldous, L. Lovász and P. Winkler, Mixing Times for uniformly ergodic Markov chains, Stochastic Processes and their Applications **71**, (1997) 165–185.
3. I. Benjamini and E. Mossel, On the mixing time of a simple random walk on the super critical percolation cluster, Probability Theory and Related Fields **125**, (2003) 408–420.

4. B. Bollobás, S. Janson and G. R. Grimmett, The random-cluster model on the complete graph, Probability Theory and Related Fields **104**, (1996) 283–317.
5. P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Annals of Appl. Prob. **6**, (1996) 695–750.
6. P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains, Annals of Appl. Prob. **1**, (1991) 36–61.
7. M. Dyer, A.M. Frieze and R. Kannan, A random polynomial time algorithm for estimating volumes of convex bodies, Proceedings of the 21st Annual ACM Symposium on Theory of Computing, (1989) 375–381.
8. M. Dyer, C. Greenhill and M. Molloy, Very rapid mixing of the Glauber dynamics for proper colourings on bounded-degree graphs, Random Structures and Algorithms **20**, (2002) 98–114.
9. N. Fountoulakis and B.A. Reed, The evolution of the conductance of a random graph. In preparation.
10. W. Feller, *An Introduction to Probability Theory and its Applications I*, (3rd edition, Wiley, 1968) xviii+509 pages.
11. C. Gantzidis, M. Mihail and A. Saberi, Conductance and congestion in power law graphs, Proceedings of the 2003 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, (2003) 148–159.
12. M. Hildebrand, Random walks on random simple graphs, Random Structures and Algorithms **8** No. 4, (1996) 301–318.
13. M. Jerrum and A. Sinclair, Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved. Proceedings of the 20th Annual ACM Symposium on Theory of Computing, (1988) 235–244.
14. M. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, SIAM Journal on Computing **22**, (1993) 1087–1116.
15. R. Kannan, L. Lovász and R. Montenegro, Blocking conductance and mixing in random walks, Combinatorics, Probability and Computing, *to appear*.
16. L. Lovász and R. Kannan, Faster mixing via average conductance, Proceedings of the 31st Annual ACM Symposium on Theory of Computing, (1999) 282–287.
17. L. Lovász, R. Kannan and M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm for convex bodies, Random Structures and Algorithms **11**, (1997) 1–50.
18. L. Lovász and M. Simonovits, Random walks in a convex body and an improved volume algorithm, Random Structures and Algorithms **4**, (1993) 359–412.
19. L. Lovász and P. Winkler, Mixing of random walks and other diffusions on a graph. In *Surveys in Combinatorics*, (P. Rowlinson, editor), (1995) 119–154.
20. L. Lovász and P. Winkler, Efficient stopping rules for Markov chains, Proceedings of the 31st Annual ACM Symposium on Theory of Computing, (1995) 76–82.
21. L. Lovász and P. Winkler, Mixing Times. In *Microsurveys in Discrete Probability*, (D. Aldous and J. Propp, eds.), DIMACS Series in Discrete Math. and Theoretical Computer Science 41, Amer. Math. Soc., Providence RI, (1998) 85–134.
22. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics **21**, (1953) 1087–1092.
23. M. Molloy and B. Reed, A critical point for random graphs with given degree sequence, Random Structures and Algorithms **6**, (1995) 161–179.
24. R. Montenegro, Faster mixing by isoperimetric inequalities, Ph.D. Thesis, Department of Mathematics, Yale University, 2002.
25. J. Propp and D.B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Structures and Algorithms **9**, (1996) 223–252.

26. D. Randall, Mixing, Proceedings the 44th Annual IEEE Symposium on Foundations of Computer Science, (2003) 4–17.
27. D. Randall and D.B. Wilson, Sampling spin configurations of an Ising system, Proceedings of the 10th ACM/SIAM Symposium on Discrete Algorithms, (1999) 959–960.