Analyzing Glauber dynamics by comparison of Markov chains

Dana Randall[1] and Prasad Tetali[2]

[1] School of Mathematics and College of Computing,
Georgia Institute of Technology, Atlanta GA 30332, USA
randall@math.gatech.edu*

[2] School of Mathematics, Georgia Institute of Technology,
Atlanta, GA 30332, USA
tetali@math.gatech.edu**

Abstract. A popular technique for studying random properties of a combinatorial set is to design a Markov chain Monte Carlo algorithm. For many problems there are natural Markov chains connecting the set of allowable configurations which are based on local moves, or "Glauber dynamics." Typically these single site update algorithms are difficult to analyze, so often the Markov chain is modified to update several sites simultaneously. Recently there has been progress in analyzing these more complicated algorithms for several important combinatorial problems.
In this work we use the comparison technique of Diaconis and Saloff-Coste to show that several of the natural single point update algorithms are efficient. The strategy is to relate the mixing rate of these algorithms to the corresponding non-local algorithms which have already been analyzed. This allows us to give polynomial bounds for single point update algorithms for problems such as generating tilings, colorings and independent sets.

1 Introduction

Random sampling of combinatorial structures such as lozenge tilings of a triangular lattice, Eulerian orientations and 3-colorings of planar lattices, and independent sets of a graph has attracted the attention of researchers in combinatorics, theoretical computer science and statistical physics in recent years. The Markov chain Monte-Carlo method has played a crucial role in establishing efficient algorithms for almost uniform sampling of such structures and in yielding fully polynomial randomized approximation schemes for the corresponding counting problem (see e.g. [8],[13],[15],[10]).

First we consider planar tiling problems, namely lozenge tilings on regions of the triangular lattice and domino tilings on the Cartesian lattice. For each problem the tiles cover two adjacent cells in the lattice and the tiling is a way

* Research supported by NSF Grant No. CCR-9703206.
** Research supported by NSF Grant No. CCR-9503952.

of covering the region so that each cell is covered by exactly one tile. The set of tilings (or matchings of the dual lattices) corresponds to dimer configurations in statistical physics. In addition, lozenge tilings are well known to correspond to plane partitions which have combinatorial interest.

The second problem is three-coloring subregions of the Cartesian lattice. A proper three-coloring assigns a color to each vertex in the region so that neighbors are differently colored. Again the primary motivation for the generation problem comes from statistical physics where colorings relate to the ice model and the dominant coefficient of the partition function of the three-state Potts model [2].

The final problem we consider is that of generating a random independent set of a graph. The weighted version namely, sampling the independent sets according to a Gibbs measure, was motivated by the interest in estimating the partition function of the hard-core distribution (see e.g. [3]).

Each of these generation problems have natural Markov chain algorithms based on single site updates, also known as *Glauber dynamics*. For example an obvious algorithm for generating a random independent set can be described as follows. The state space is the set of all (valid) independent sets of the given graph. Starting with an arbitrary independent set of the graph, at each step pick a vertex uniformly at random. If it is contained in the independent set then remove it with a certain probability, and otherwise add it to the independent set, also with an appropriate probability, if this new set remains an independent set. The probabilities are chosen carefully so that the stationary distribution of the Markov chain is exactly the distribution we would like to sample from.

In this work we analyze the single site update/replacement algorithms for the above problems, and we provide low-degree polynomial bounds for the rates of mixing of the corresponding Markov chains. Our analysis makes use of the results on the chains using the fancier moves inspired by heat bath algorithms – the "tower moves" in the case of tilings and the "edge moves" in the case of independent sets.

The proof technique uses a comparison theorem due to Diaconis and Saloff-Coste [5]. Their theorem yields a geometric comparison inequality that gives bounds on the eigenvalues of a reversible Markov chain in terms of the eigenvalues of a second chain. The main application in [5] was to give a sharp upper bound on the second eigenvalue of the symmetric exclusion process on a graph. The symmetric exclusion process on a graph is a certain generalization of a simple random walk on a graph and can be described as follows. Start with an arbitrary placement of r particles on r vertices of a graph. At each discrete time step, a particle is chosen at random, and then one of its neighboring vertices is chosen at random. If the neighbor is unoccupied (by a particle) then the chosen particle is moved there, otherwise the system stays as it was. The special case of $r = 1$ corresponds to the simple random walk on a graph. Diaconis and Saloff-Coste bound the second eigenvalue of this chain by comparing it to a well studied chain (the Bernoulli–Laplace model for diffusion) whose eigenvalues are known.

Our approach is somewhat different in this paper, since the Markov chains we work with are much more combinatorial in nature; in particular, it is very hard to determine the second eigenvalue of these chains or of related chains with the same stationary distributions. However, the "known" chains in our applications are chains whose mixing times are known. Using the existing literature on relating the time to reach equilibrium and the second eigenvalue (e.g. [4],[8]), together with the comparison theorem, we derive an inequality relating the mixing times of two chains. This allows us to estimate the rate of mixing of the "unknown" chains (based on single site updates) mentioned above. Direct analysis of any of these chains seems challenging and might yield tighter bounds on the mixing times.

In Section 2 we describe relevant results from the theory of rapidly mixing Markov chains, including relations between mixing times and eigenvalues, and comparison inequalities. In Section 3 we describe in detail the application to tilings, colorings and independent sets.

2 Preliminaries

Let (Ω, P, π) denote an ergodic (i.e. irreducible and aperiodic) Markov chain with finite state space Ω, transition probability matrix P, and stationary distribution π. Furthermore, we assume that the chain is reversible, i.e. that we have the detailed–balance conditions, $\pi(x)P(x,y) = \pi(y)P(y,x)$, for all $x, y \in \Omega$. Assuming we are dealing with discrete-time Markov chains, for $x, y \in \Omega$, $t \in \mathcal{Z}^+$, let $P^t(x,y)$ denote the t-step probability of going from x to y. Then the time a Markov chain takes to be close to equilibrium can be measured using the *variation distance* between P^t and π, where the variation distance is given by

$$\Delta_x(t) = \frac{1}{2} \sum_{y \in \Omega} |P^t(x,y) - \pi(y)|.$$

We also denote by $\Delta(t)$ the variation distance starting from the worst state, i.e. $\Delta(t) = \max_{x \in \Omega} \Delta_x(t)$.

• **mixing time and the second eigenvalue**

For $\epsilon > 0$, the *mixing time*, starting from state x, is defined by

$$\tau_x(\epsilon) = \min\{t : \Delta_x(t') \leq \epsilon, \qquad \forall t' \geq t\}.$$

Once again we denote by $\tau(\epsilon)$, the mixing time starting from the worst state, i.e. $\tau(\epsilon) = \max_{x \in \Omega} \tau_x(\epsilon)$. For the rest of the paper, when we refer to *mixing time*, we always mean $\tau(\epsilon)$.

Let $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\Omega|-1} > -1$ denote the eigenvalues of P. The following result of Welsh [22] (which is an extension of a key result from [16]; see also [4], [17]) shows the relationship between mixing times and the maximum eigenvalues. Strictly speaking, λ_1 in the following theorem should be replaced by $\lambda_{max} = \max(\lambda_1, |\lambda_{|\Omega|-1}|)$, but in all our applications below we make sure that $\lambda_1 > |\lambda_{|\Omega|-1}| > 0$ by adding self-loops with weight $1/2$.

Theorem 1. *For $\epsilon > 0$, we have*
(i) for all $x \in \Omega$, $\tau_x(\epsilon) \le \frac{1}{1-\lambda_1} \log\left(\frac{1}{\pi(x)\epsilon}\right)$;
(ii) $\max_x \tau_x(\epsilon) \ge \frac{\lambda_1}{2(1-\lambda_1)} \log\left(\frac{1}{2\epsilon}\right)$.

• mixing time and coupling time

Another method for bounding the mixing time $\tau(\epsilon)$ is to construct a *coupling* for the Markov chain. A coupling is a new Markov chain on the state space $\Omega \times \Omega$ (where Ω is the original state space, e.g., the set of 3-colorings) with the following properties. Rather than updating two configurations independently, the coupled process correlates the random coin flips while maintaining that each configuration, when observed in isolation, is just performing transitions of the original Markov chain. In addition, we need that if the two configurations agree, the coupled process will force them to agree at all future times. Coupling is a crucial ingredient in all of the applications in section 3.

The following theorem states that the coupling time, which is the expected time it takes for two configurations to meet starting from the worst starting point, provides a good bound on the mixing time. More formally, let x and y be the starting configurations. Then

$$T^{x,y} = \min\{t : X_t = Y_t \mid X_0 = x,\, Y_0 = y\},$$

and define the *coupling time* to be $T = \max_{x,y} \mathrm{E}T^{x,y}$. The following result relates the mixing time and the coupling time (see [1]).

Theorem 2. $\tau(\epsilon) \le 6T(1 + \ln \epsilon^{-1})$.

• comparison of eigenvalues (via Dirichlet forms)

Let $\tilde{P}$ and P denote two reversible Markov chains on the same state space Ω with the same stationary distribution π. Then Diaconis and Saloff-Coste (see [5]) provide the following geometric bound between the two eigenvalues $\lambda_1(\tilde{P})$ and $\lambda_1(P)$. (Strictly speaking, the result in [5] compares the Dirichlet forms associated with $\tilde{P}$ and P thus yielding the following comparison result between *all* nontrivial eigenvalues, and not just the second eigenvalue.)

First we need some more notation. As we shall see, in applications, $\tilde{P}$ is the chain with known eigenvalues (or known mixing time), and P is the chain whose mixing time we would like to bound by comparing with $\tilde{P}$. Let $E(P) = \{(x,y) : P(x,y) > 0\}$ and $E(\tilde{P}) = \{(x,y) : \tilde{P}(x,y) > 0\}$ denote the sets of edges of the two chains, viewed as directed graphs. For each x, y with $\tilde{P}(x,y) > 0$, define a *path* γ_{xy} using a fixed sequence of states, $x_0 = x, x_1, \ldots, x_{k-1}, x_k = y$ with $P(x_i, x_{i+1}) > 0$. The length ($= k$) of such a path will be denoted by $|\gamma_{xy}|$. Further let

$$\Gamma(z,w) = \left\{(x,y) \in E(\tilde{P}) \text{ such that } (z,w) \in \gamma_{xy}\right\}$$

denote the set of paths which use the transition (z,w).

Theorem 3. *With the above notation, we have*

$$(1-\lambda_1(P)) \geq \frac{1}{A}(1-\lambda_1(\tilde{P})),$$

where

$$A = \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}|\pi(x)\tilde{P}(x,y) \right\}.$$

It is worth noting that the quantity A above depends on our choice of paths $\{\gamma_{xy}\}$; thus these paths play a role akin to that of the *canonical paths* in bounding the *conductance* of a Markov chain. However, the crucial difference, as pointed in [5], is that we need only define these paths between pairs of states which are adjacent in the known chain.

In the following our strategy is as follows. We begin with a bound on the mixing time of a chain established, say, via the coupling method and Theorem 2. We then use part (ii) of Theorem 1 above to lower bound the spectral gap of such a chain. Next we use the comparison theorem (Theorem 3) to lower bound the spectral gap of an unknown chain by carefully bounding the parameter A. This in turn provides us with a bound on the mixing time of the unknown chain in view of part(i) of Theorem 1. The following technical proposition makes precise the aforementioned strategy, and is thus crucial to our results of the next section.

Let $\tilde{\tau}(\epsilon)$ and $\tau(\epsilon)$ denote the mixing times of $\tilde{P}$ and P respectively. Then with A as in Theorem 3, we have the following comparison result relating the mixing times. Let π_* denote $\min_{x\in\Omega} \pi(x)$.

Proposition 4. *For $0 < \epsilon < 1$, and for all $x \in \Omega$, we have that*

$$\tau(\epsilon) \leq \frac{4\log(1/(\epsilon\pi_*))}{\log(1/2\epsilon)} A\tilde{\tau}(\epsilon).$$

Proof. For $0 < \epsilon < 1$, from part (ii) of Theorem 1, we have,

$$\tilde{\tau}(\epsilon) \geq \frac{\lambda_1(\tilde{P})}{2(1-\lambda_1(\tilde{P}))} \log\left(\frac{1}{2\epsilon}\right).$$

This implies that

$$(1-\lambda_1(\tilde{P})) \geq \frac{1}{4\tilde{\tau}(\epsilon)} \log\left(\frac{1}{2\epsilon}\right),$$

wherein we also used the trivial bound, $\lambda_1(\tilde{P}) \geq 1/2$. Now using the comparison theorem, we get that

$$1-\lambda_1(P) \geq \frac{1}{A}(1-\lambda_1(\tilde{P})) \geq \frac{1}{A}\frac{1}{4\tilde{\tau}(\epsilon)} \log\left(\frac{1}{2\epsilon}\right).$$

Finally, using part(i) of Theorem 1 we can bound the mixing time of P, starting from any state x,

$$\tau_x(\epsilon) \leq \frac{4\log(1/(\epsilon\pi(x)))}{\log(1/2\epsilon)} A\tilde{\tau}(\epsilon),$$

completing the proof of the proposition.

Remark 1. The above proposition illustrates the fact that the comparison argument is effective as long as we can control the factor A, which depends on the choice of paths in the unknown chain. The dependence on π_*, albeit not as crucial, can affect the mixing time by another factor involving the size of the input, since $1/\pi_*$ in most cases is at most exponential in the size of the input.

3 Applications

3.1 Lozenge tilings

Let R be a region of the triangular lattice. A lozenge tiling of R is a covering of the region with lozenges tiles, where each lozenge covers two adjacent cells in R and no two lozenges overlap. Just looking at a lozenge tiling causes a three dimensional surface to appear – in fact the set of lozenge tilings correspond bijectively with the surfaces formed by placing unit cubes in a larger three-dimensional frame such that each cube is supported on its back three sides. The shape of the frame is uniquely determined from the region R (see Fig. 1).

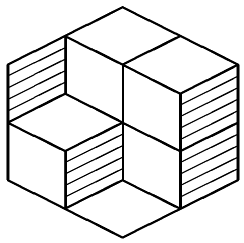

Fig. 1. A lozenge tiling viewed as a surface

Given this equivalence, there is an obvious Markov chain $\mathcal{M}_{ssu}$ for generating tilings. Namely, connect any two tilings whose surfaces differ by the addition or removal of a single cube. In the two dimensional picture of a tiling this corresponds to chosing a hexagonal window and if it is comprised of three tiles, rotate them by 60^o (see Fig. 2). More precisely, the transition matrix $P(\cdot,\cdot)$ of $\mathcal{M}_{ssu}$ is defined as

$$P(x,y)=\begin{cases} 1/2N, & \text{if } x \oplus y \text{ is a cube (or hexagon)} \\ 1-\sum\limits_{z\neq x}\tilde{P}(x,z), & \text{if } y=x, \end{cases}$$

In [13] a modified algorithm based on tower moves was analyzed. Again the state space is the set of all lozenge tilings. Two tilings differ by a tower of height k if they differ by the addition or removal of a $1\times 1\times k$ vertical column of cubes. Let x and y be lozenge tilings of the region R. Let $\mathcal{M}_{loz}$ represent the Markov chain in which there is a move from x to y if and only if the symmetric difference

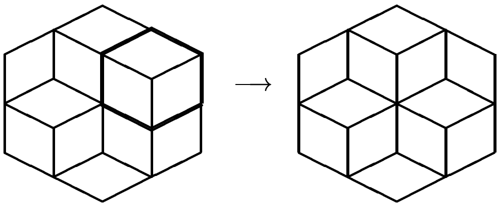

Fig. 2. A move in the Markov chain $\mathcal{M}_{ssu}$

of the edges of x and y is a tower. Recall that the transition probabilities $\tilde{P}(\cdot\,,\cdot)$ of $\mathcal{M}_{loz}$ are defined by

$$\tilde{P}(x,y) = \begin{cases} 1/2Nh, & \text{if } x \oplus y \text{ is a} \\ & \text{tower of height } h; \\ 1 - \sum\limits_{z \neq x} \tilde{P}(x,z), & \text{if } y = x, \end{cases}$$

where N is the area of the region being tiled. Note that both $\mathcal{M}_{ssu}$ and $\mathcal{M}_{loz}$ have the uniform distribution as the stationary distribution. Wilson improves the analysis given in [13] to show that the Markov chain based on tower moves mixes in time $\mathrm{O}(W^2 N \log N \log(2/\epsilon))$, where W is the width and N is the area of the triangular lattice region to be tiled [23]. However, neither approach gives the mixing time of $\mathcal{M}_{ssu}$.

We now define a set of paths and then bound A corresponding to these paths. For each (x,y) which differ by a tower move of height h_{xy} (and are thus adjacent in the known chain), there is a unique minimum length sequence of single site update moves of length h_{xy} which transforms x into y. Such a sequence defines a path γ_{xy}, in a natural way, using transitions of $P(\cdot,\cdot)$. Note that the length of the path is h_{xy} and $\tilde{P}(x,y) = 1/2Nh_{xy}$. Consider an arbitrary (z,w) where z and w are lozenge tilings which differ by a single cube. Note that $P(z,w) = 1/2N$. Furthermore, for a given (z,w), the number of (x,y) such that the path γ_{xy} uses (z,w) is at most H^2, where H is the maximum height of a tower. (This is because the bottom and the top of any tower containing a particular single site can be chosen in at most H ways.) Thus yielding the following bound for the quantity A from the comparison theorem.

$$\begin{aligned} A &= \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}| \pi(x) \tilde{P}(x,y) \right\} \\ &= \max_{(z,w)\in E(P)} \left\{ \frac{2N}{\pi(z)} \sum_{\Gamma(z,w)} h_{xy} \pi(x)(1/2Nh_{xy}) \right\} \\ &\leq \sum_{\Gamma(z,w)} 1 \leq H^2. \end{aligned}$$

Theorem 5. *Let R be a region in the triangular lattice whose convex hull has area N. Then the mixing time of $\mathcal{M}_{ssu}$ for generating a lozenge tiling of R is given by*

$$\tau_{ssu} = \mathrm{O}\left(N^4 \log N + N^3 \log N \log(1/\epsilon)\right).$$

Proof. Clearly the number of lozenge tilings of a region of size N is at most 3^N (since we can overcount by replacing each triangle in the underlying region with a triangle with one side identified – those configurations where identified edges line up are the set of valid tilings). The bound on the mixing time of $\mathcal{M}_{loz}$ given by Wilson [23] is $\tau(\epsilon) = \mathrm{O}(W^2 N \log N \log(1/\epsilon))$, where W is the width of the region. Therefore, by proposition 4,

$$\begin{aligned}\tau_{ssu} &\leq \frac{\log(1/(\epsilon\pi_*))}{\log(2/\epsilon)}(H^2)(W^2 N \log N \log(1/\epsilon)) \\ &= \mathrm{O}\left(N^4 \log N + N^3 \log N \log(1/\epsilon)\right),\end{aligned}$$

since $H * W = \mathrm{O}(N)$.

• **Domino tilings**

The single site algorithm for domino tilings follows exactly the same analysis. Starting from any domino tiling, choose a 2×2 window; if there are two parallel dominoes, rotate them by 90^o. This simple algorithm is motivated by the linear time tiling algorithm of Thurston [19] for generating a single tiling. In [13] a tower algorithm is presented which achieves a mixing time of $\mathrm{O}(N^{3.5}(1 + \log(1/\epsilon)))$, where N is the area of the Cartesian lattice region to be tiled. We can show that $A \leq H^2$, where H is the size of the maximal tower; for square regions this is $\mathrm{O}(N^{1/2})$ and in general is at most N. In addition, the number of domino tilings is trivially bounded by 4^N. Thus, the comparison theorem establishes the efficiency of this local algorithm.

3.2 Three-colorings and Eulerian orientations

The second application we consider is generating three-colorings of a subregion of the 2-dimensional Cartesian lattice. If we fix the colors around the boundary of the region, the set of three-colorings corresponds to Eulerian orientations of a region in the dual lattice – an Eulerian orientation is a way of directing the edges within a region so that each vertex has its indegree equal to its outdegree (see, e.g. [2]). In [13] an algorithm based on tower moves was analyzed which is similar to the algorithm for lozenges. We now show that a single point update algorithm is also rapidly mixing. Using the ideas in [15] this can be generalized further to show that the single point update algorithm is rapidly mixing on the set of *all colorings* (allowing the boundary colors to vary) but we leave the details for the full paper.

The single point update algorithm is defined as follows. Starting from a valid three-coloring (e.g., the 2-coloring arising from the bipartition of the Cartesian lattice) choose a vertex v and a color $c \in \{0,1,2\}$ uniformly at random. If v can be recolored using c, recolor it with probability $1/2$, otherwise remain at the current coloring. In [13] it was shown that this Markov chain is ergodic. Jerrum showed that this single point update algorithm is rapidly mixing on the set of k-colorings (choosing $c \in \{0,...,k-1\}$) when $k \geq 2\Delta$, where Δ is the maximum degree [7]; this implies we can k-color the 2-d lattice whenever $k \geq 8$. Here we show the same algorithm is efficient when $k = 3$. We start by recalling the 3-coloring algorithm based on tower moves.

Let x and y be three-colorings of R, and let $c_x(v)$ and $c_y(v)$ be the respective colors of vertex v. We say x and y differ by a tower of height k if there exists a vertical or horizontal set of vertices where we can add 1 to each of the colors of x (or y) to get the other coloring. More precisely, it implies that there exists a vertex v and a unit basis vector e such that $c_y(v) = c_x(v) + 1,\ c_y(v+e) = c_x(v+e) + 1$,..., $c_y(v + (k-1)e) = c_x(v + (k-1)e) + 1$, or $c_y(v) = c_x(v) - 1, ..., c_y(c + (k-1)e) = c_x(c + (k-1)e) - 1$, (where all sums are taken mod 3), and $c_y(v') = c_x(v')$ for all other vertices v'. Then we define our transition matrix $\tilde{P}_{col}$ of $\mathcal{M}_{col}$ to be

$$\tilde{P}_{col}(x,y) = \begin{cases} 1/2Nh, & \text{if } x \oplus y \text{ is a} \\ & \text{tower of height } h; \\ 1 - \sum_{z \neq x} \tilde{P}_{col}(x,z), & \text{if } y = x, \end{cases}$$

Again, the crucial ingredient is bounding A, the factor arising in the comparison theorem. As before, we define our paths between x and y which are connected by a tower move of height k by decomposing the moves into k single site update moves in the obvious way. It is not difficult to see that for any (z,w) which differ by a single vertex that $\Gamma(z,w) \leq 2H^2$, where H is the maximal height of a tower. Following exactly the same analysis as in the previous subsection yields

$$A \leq 2H^2.$$

Since the number of three-colorings is bounded by 3^N, this, together with the polynomial bound on the mixing time of $\mathcal{M}_{col}$ shows that the single site update algorithm is rapidly mixing as well.

3.3 Independent sets

Consider the problem of sampling from the set $\mathcal{I}$ of all independent sets of a graph $G = (V,E)$ with the following distribution. Let $\lambda > 0$ be an arbitrary real. Then we are interested in the probability distribution π on $\mathcal{I}$ which associates with each $S \in \mathcal{I}$,

$$\pi(S) = \frac{\lambda^{|S|}}{Z},$$

where $Z = \sum_{S \in \mathcal{I}} \lambda^{|S|}$, is the normalizing factor. This problem is motivated by the hard-core lattice gas model in statistical physics, and in that context Z is referred to as the *partition function*.

A natural single site replacement algorithm for this problem is the one in which we start with an arbitrary independent set of G, and in each step, we pick a vertex v uniformly at random from V, and with appropriate probability either (i) include it in the current independent set S, if none of its neighbors are in S, or (ii) remove it from S, if v is already in S. The precise formulation is as follows. Let $|V| = n$ and $|E| = m$. Also, let $\Delta = \Delta(G)$ denote the maximum degree in G. Let x denote the current state (i.e. independent set). Then the transitions $P(x, y)$ of such a chain $\mathcal{M}_v$, based on single replacements, can be described as follows.

- Choose $u \in V$ uniformly at random
- If $u \in x$ then $y = x \setminus \{u\}$ with probability $1/(1+\lambda)$ and $y = x$ with the remaining probability.
- If $u \notin x$ and if $x \cup \{u\}$ is an independent set, then $y = x \cup \{u\}$ with probability $\lambda/(1+\lambda)$; otherwise $y = x$.

Recently, Luby and Vigoda considered [14] the following slightly different chain $\mathcal{M}_e$ which works by choosing *edges* uniformly at random. The transition probabilities $\tilde{P}(\cdot, \cdot)$ of $\mathcal{M}_e$ can be described as follows. Let x be the current state and y the next state.

- Choose $e = (u, v) \in E$ uniformly at random
- Let

$$Y = \begin{cases} x \setminus \{u, v\} & \text{with probability } 1/(1+2\lambda) \\ x \cup \{u\} \setminus \{v\} & \text{with probability } \lambda/(1+2\lambda) \\ x \cup \{v\} \setminus \{u\} & \text{with probability } \lambda/(1+2\lambda) \end{cases}$$

- If Y is a (valid) independent set, then set $y = Y$, else $y = x$.

In [14], under the assumption that $\lambda \leq 1/(\Delta - 3)$, the chain $\mathcal{M}_e$ was shown to have the mixing rate $\tilde{\tau}(\epsilon)$ of $\mathrm{O}(n^3 \ln(\epsilon^{-1}))$.

Once again, there is a natural way to define a path between pairs (x, y) with $\tilde{P}(x, y) > 0$. If (x, y) is a valid transition in P (i.e. $P(x, y) > 0$), then we let the same transition denote the (trivial) path of length one between x and y. This leaves the only case of when x and y are such that there exist adjacent vertices u and v in the graph with $u \in x$, $u \notin y$ and $v \in y$, $v \notin x$. In such a case we define the path γ_{xy} to be $x_0 = x, x_1 = x \setminus \{u\}, x_2 = x_1 \cup \{v\} = y$; note that such a path is of length two.

For $(z, w) \in E(P)$, consider

$$A(z, w) = \frac{1}{\pi(z) P(z, w)} \sum_{(x,y) \in \Gamma(z,w)} |\gamma_{xy}| \pi(x) \tilde{P}(x, y)$$

This can be bounded by considering the following possible cases.

Case 1. Suppose (z, w) is such that $z = w \cup \{u\}$ (see Fig. 3). Then $P(z, w) = 1/[(1+\lambda)n]$. Such a transition can be in a path γ_{xy} of length one or a path $\gamma_{x',y'}$ of length two. In the former case $x = z$ and $y = w$, and there are at most Δ

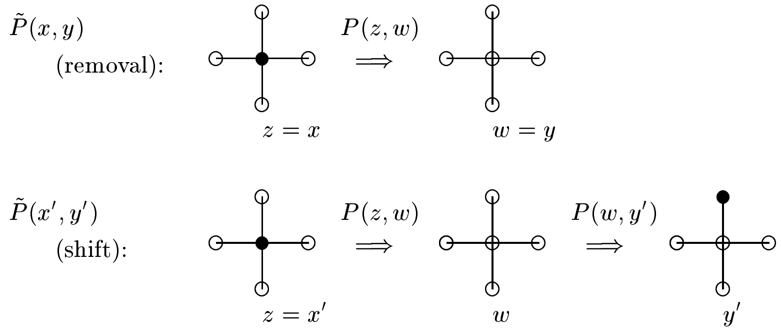

Fig. 3. Case 1: Edge moves decomposed into one or two–step paths in the single site chain.

edges moves which cause this transition. Similarly, in the latter case $x' = z$ and so $\pi(x') = \pi(z)$ and there are at most Δ such (x', y'). Letting $\Gamma_2(z,w) = \{(x',y')$ such that $\gamma_{x',y'}$ has length two$\}$, we find that

$$\begin{aligned}
A(z,w) &\leq \frac{1}{P(z,w)} \left(\tilde{P}(z,w) + \sum_{(x',y') \in \Gamma_2(z,w)} 2\tilde{P}(x',y') \right) \\
&\leq (1+\lambda)n \left(\frac{\Delta}{m} \frac{1}{(1+2\lambda)} + 2\Delta \frac{\lambda}{(1+2\lambda)m} \right) \\
&= \frac{\Delta(1+\lambda)n}{(1+2\lambda)m}(1+2\lambda) = (1+\lambda)\frac{\Delta n}{m}.
\end{aligned}$$

Case 2. Suppose (z,w) is such that $z = w \setminus \{u\}$. By reversibility, $\pi(z)P(z,w) = \pi(w)P(w,z)$. For each $(x,y) \in \Gamma(z,w)$, we have a path $(y,x) \in \Gamma(w,z)$, and once again, by reversibility, $\pi(x)P(x,y) = \pi(y)P(y,x)$. Therefore $A(z,w) = A(w,z)$, and by Case 1 we again find

$$A(z,w) \leq (1+\lambda)\frac{\Delta n}{m}.$$

Thus from both cases, we have,

$$\begin{aligned}
A = &\max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}| \pi(x) \tilde{P}(x,y) \right\} \\
&\leq (1+\lambda)\frac{\Delta n}{m}.
\end{aligned}$$

Remark 2. Vigoda recently discovered an argument for directly bounding the mixing time of the single site dynamics for the independent set problem using a clever potential function and *path coupling* [20]. This shows that the mixing time of $\mathcal{M}_v$ is bounded by $O(n \log n \log(2/\epsilon))$ for $\lambda < 2/(\Delta - 2)$, and $O(n^2 \log n \log(2/\epsilon))$ for $\lambda = 2/(\Delta - 2)$.

4 Conclusions

In this paper we address the issue that although two Markov chains appear quite similar, it is often the case that only one admits a simple analysis using currently available tools. We envision that there are many other applications for comparison techniques.

For instance, the results from section 3.3 for independent sets can be extended to relate the mixing rate of the single site updates to the mixing rate of a dynamics which would update all the sites in a rectangle of fixed size $a \times b$. This so-called "heat bath" algorithm is used experimentally in statistical physics to study the uniqueness of the Gibbs state of the hard-core lattice gas model. Using the method given here, we derive a bound on the mixing rate of this new Markov chain which introduces a factor which depends exponentially on $\min(a, b)$.

Finally, we remark that Diaconis–Saloff-Coste have also extended their comparison theorem to compare the log-Sobolev constants of two Markov chains [6]. It would be nice to find applications in computer science for this result as well.

References

[1] Aldous, D. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités* XVII, 1981/82, Springer Lecture Notes in Mathematics 986, pp. 243–297.

[2] Baxter, R.J. *Exactly solved models in statistical mechanics.* Academic Press, London, 1982.

[3] van den Berg, J. and Steif, J.E. Percolation and the hard-core lattice gas model. *Stochastic Processes and their Applications* **49**, 1994, pp. 179-197.

[4] Diaconis, P. and Stroock, D. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probability* **1**, 1991, pp. 36-61.

[5] Diaconis, P. and Saloff-Coste, L. Comparison theorems for reversible Markov chains. *Ann. Appl. Probability* **3**, 1993, pp. 696-730.

[6] Diaconis, P. and Saloff-Coste, L. Logarithmic Sobolev Inequalities for Finite Markov Chains. *Ann. of Appl. Probab.* **6** (1996), pp. 695-750.

[7] Jerrum, M. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms* **7**, 1995, pp. 157-165.

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

[9] Jerrum, M., Valiant, L. and Vazirani, V. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science* **43** (1986), pp. 169–188.

[10] Kannan, R., Tetali, and P., Vempala, S. Simple Markov chain algorithms for generating bipartite graphs and tournaments. *Proc. of the 8th ACM-SIAM Symp. on Discrete Algorithms* January 1997.

[11] Lieb, E.H. Residual entropy of square ice. *Physical Review* **162**, 1967, pp. 162-172.

[12] Lubin, M. and Sokal A.D. Comment on "Antiferromagnetic Potts Model. *Phys. Rev. Lett.* **71**, 1993, pp. 1778.

[13] Luby, M., Randall, D. and Sinclair, A. Markov Chain Algorithms for Planar Lattice Structures. *Proc. 36th IEEE Symposium on Foundations of Computing* (1995), pp. 150-159.

[14] Luby, M., Vigoda, E. Approximately counting up to four. *Proc. 29th ACM Symposium on Theory of Computing* (1997), pp. 150-159.

[15] Madras, N. and Randall, D. Factoring Graphs to Bound Mixing Time. *Proc. 37th IEEE Symposium on Foundations of Computing* (1996).

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

[17] Sinclair, A.J. *Algorithms for random generation & counting : a Markov chain approach.* Birkhäuser, Boston, 1993, pp. 47-48.

[18] Sinclair, A.J. *Improved bounds for mixing rates of Markov chains and multicommodity flow.* Combinatorics, Probability, & Computing. **1** (1992), pp. 351-370.

[19] Thurston, W. Conway's tiling groups. *American Mathematical Monthly* **97**, 1990, pp. 757-773.

[20] Vigoda, E. personal communication.

[21] Welsh, D.J.A. *The computational complexity of some classical problems from statistical physics.* In Disorder in Physical Systems, (G. Grimmett and D. Welsh eds.). Claredon Press, Oxford, 1990, pp. 323-335.

[22] Welsh, D.J.A. *Approximate counting.* In Surveys in Combinatorics, (R.A. Bailey, ed.). Cambridge University Press, London Math Society Lecture Notes 241, 1997, pp. 287-317.

[23] Wilson, D.B. Mixing times of lozenge tiling and card shuffling Markov chains, draft of manuscript.