Markov Chains for Sampling Matchings

James Matthews

Doctor of Philosophy

Laboratory for Foundations of Computer Science
School of Informatics
University of Edinburgh

2008

Abstract

Markov Chain Monte Carlo algorithms are often used to sample combinatorial structures such as matchings and independent sets in graphs. A Markov chain is defined whose state space includes the desired sample space, and which has an appropriate stationary distribution. By simulating the chain for a sufficiently large number of steps, we can sample from a distribution arbitrarily close to the stationary distribution. The number of steps required to do this is known as the mixing time of the Markov chain.

In this thesis, we consider a number of Markov chains for sampling *matchings*, both in general and more restricted classes of graphs, and also for sampling *independent sets* in claw-free graphs. We apply techniques for showing rapid mixing based on two main approaches: coupling and conductance. We consider chains using single-site moves, and also chains using large block moves.

Perfect matchings of bipartite graphs are of particular interest in our community. We investigate the mixing time of a Markov chain for sampling perfect matchings in a restricted class of bipartite graphs, and show that its mixing time is exponential in some instances. For a further restricted class of graphs, however, we can show sub-exponential mixing time.

One of the techniques for showing rapid mixing is coupling. The bound on the mixing time depends on a contraction ratio β. Ideally, $\beta < 1$, but in the case $\beta = 1$ it is still possible to obtain a bound on the mixing time, provided there is a sufficiently large probability of contraction for all pairs of states. We develop a lemma which obtains better bounds on the mixing time in this case than existing theorems, in the case where $\beta = 1$ and the probability of a change in distance is proportional to the distance between the two states. We apply this lemma to the Dyer-Greenhill chain for sampling independent sets, and to a Markov chain for sampling 2Δ-colourings.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr Mary Cryan and Prof. Mark Jerrum, for their advice and support over the past four years.

I can't easily enumerate all the other people who have helped me over the years, whether academically or simply by providing support, encouragement, or distraction when things were going badly. However, Tef, Dragal and Steve (in no particular order) have been a great help in all of the above. I trust nobody is offended by my failure to name them.

Finally, I would like to thank my family for their continued support, and Mareile for her patience and understanding.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*James Matthews*)

Table of Contents

List of Figures

Chapter 1

Introduction

Traditional complexity analysis is typically concerned with the complexity of decision problems - we are given a problem instance, and wish to determine whether some condition holds. The natural context for studying the complexity of decision problems is the complexity class NP. NP can be defined in terms of a polynomial-time witness-checking predicate: given a problem instance, the condition of interest holds if and only if there is at least one input for which the witness-checking predicate is true. For example, if we consider the question "does the graph G contain an independent set of size at least k", then a hypothetical witness might be an independent set of at least that size, and we would look for witnesses among the set of all subsets of the vertex set of G. The witness-checking predicate in this case would answer the question "is w an independent set of G of size at least k".

This approach to defining NP for combinatorial problems naturally leads to the question of *how many* verifiable witnesses exist. Instead of finding an independent set, for example, we wish to know how many independent sets exist. The complexity class #P is the class of counting problems for which the witnesses may be checked by a polynomial-time predicate. We say that a problem f is #P-complete if every problem in #P is Turing reducible to f, and f is itself a member of #P. Many NP-complete decision problems lead to corresponding #P-complete counting problems. However, there are also polynomial-time decision problems whose counting analogues are #P-complete. One such problem is that of counting perfect matchings of graphs, as shown by Valiant in [41].

While it is generally believed that we cannot efficiently count the number of witnesses to an instance of a #P-complete problem exactly, in some cases it is possible to *approximately* count the number of witnesses in polynomial time, to within a constant

approximation factor. In many cases, it is possible to reduce the problem of approximately counting the number of witnesses to the problem of almost-uniform sampling from the set of witnesses. In this thesis, we consider methods of sampling matchings and independent sets in graphs, focusing on the Markov Chain Monte Carlo (MCMC) method.

The Markov Chain Monte Carlo method can be described as follows: To sample from the solution set of a problem instance, we define a Markov chain whose state space includes the desired sample space, and which has the desired stationary distribution (often the uniform distribution). We can then sample from a distribution arbitrarily close to the stationary distribution by simulating the chain for a sufficiently large number of steps. The time taken before the distribution of the chain is sufficiently close to its stationary distribution is known as the *mixing time*. If the mixing time is polynomial in the instance size, then the chain is referred to as a *rapidly mixing* chain.

There are several ways of bounding the mixing time of Markov chains. We focus on techniques based on two main approaches: coupling and conductance.

In Chapter 2 we will introduce some of the methods for bounding the mixing time of Markov chains. To illustrate the application of these techniques, we will consider the insert-delete chain for sampling independent sets, which was shown to be rapidly mixing by Luby and Vigoda [31], and the Jerrum-Sinclair chain for sampling matchings, which was shown to be rapidly mixing by Jerrum and Sinclair [25]. Both of these chains sample structures from a graph $G = (V,E)$ with maximum degree Δ, weighted according to a parameter λ. The sizes of the vertex and edge sets of G are $n = |V|$ and $m = |E|$, respectively. We will present the argument that the insert-delete chain is rapidly mixing for $\lambda \leq \frac{1}{\Delta - 1}$ using path coupling, and that the Jerrum-Sinclair chain is rapidly mixing for any constant λ using canonical paths.

One of the techniques for showing rapid mixing of Markov chains is coupling - indeed, a Markov chain is rapidly mixing if and only if there exists a coupling that converges in polynomial time (finding such a coupling, however, may be difficult [20]). A particular technique used to obtain coupling results is *path coupling* [7]. When we use path coupling arguments, we attempt to find a bound on a contraction ratio β between two states, from which we can infer a bound on the mixing time. The relationship between β and the mixing time is given by Theorem 2.15 of Chapter 2. Ideally, we wish to show that $\beta < 1$, but in the case where $\beta = 1$ it may still be possible to obtain a bound on the mixing time, provided there is a strictly positive probability of contraction for all pairs of states.

In Chapter 3 we will develop Theorem 3.3, which allows us to obtain better bounds on the mixing time in the case $\beta = 1$ than existing theorems for path coupling, in cases where the probability that there is a change in distance is proportional to the distance between the two states. We will apply this theorem to the Dyer-Greenhill chain for sampling independent sets [14], and also to a simple Markov chain for sampling 2Δ-colourings. We will also prove a slightly more general version of Theorem 2.15, which can be applied in cases where the coupling uses a non-integer metric.

In Chapter 4, we will introduce a new Markov chain for sampling matchings in bounded degree graphs. This new chain is based on Dyer and Greenhill's chain for sampling independent sets. We will use path coupling to show that the mixing time of our chain is $O(n^3)$ provided $\lambda \leq \frac{\Delta+\sqrt{3\Delta^2-5\Delta+2}}{2\Delta^2-5\Delta+2}$. We will also consider a family of Markov chains for sampling matchings in lattice graphs using block moves, and show that these have a mixing time of $O(n\log(n))$ provided the choice of block size is sufficiently large. We use comparison techniques to show that the Jerrum-Sinclair chain for sampling matchings mixes in time $O(nm)$ for the two-dimensional torus, an improvement of $\log(n)$ over their result using canonical paths (but for a more restricted class of graphs).

In Chapter 5, we will consider the problem of sampling independent sets in claw-free graphs. In particular, we rely on the fact that the symmetric difference of two independent sets in a given graph gives a set of alternating paths and cycles, provided the underlying graph is claw-free. We adapt the Jerrum-Sinclair chain for sampling matchings to sample independent sets of claw-free graphs, and show using canonical paths that the mixing time of our adapted chain is $O(\Delta n^3)$ for general claw-free graphs, where n is the number of vertices. We also adapt our family of chains using block moves from Chapter 4 to sample independent sets in claw-free graphs, and combine this result with comparison techniques to show that our adaptation of the Jerrum-Sinclair chain mixes in time $O(\Delta n^2)$ for claw-free lattices such as the triangular and kagome lattices.

A problem which has received much interest in our community is that of sampling perfect matchings in bipartite graphs. The number of perfect matchings in a bipartite graph is equal to the permanent of its adjacency matrix. Furthermore, uniformly sampling perfect matchings of a graph $G = (V_1, V_2, E)$ corresponds to the problem of uniformly sampling permutations of the vertex set V_1, where each vertex is restricted to a limited number of positions determined by the adjacency matrix of G. This type of condition may arise in a class of statistical tests known as *permutation tests*. The

problem of sampling (and approximately counting) perfect matchings was solved by Jerrum, Sinclair and Vigoda in 2001 for general bipartite graphs [26]. In Chapter 6, we study the mixing time of a simple chain using *Diaconis moves*, on a special class of bipartite graphs. This special class corresponds to instances of permutation testing where the data set is doubly truncated - that is, each dependent measurement can fall into a single contiguous range determined by the independent variable. We will show that the Diaconis chain is not rapidly mixing, even for this class of graphs. However, we will present a further restricted class of bipartite graphs for which the mixing time is sub-exponential, and which still covers many instances of doubly-truncated data. We also show that the mixing time is polynomial in cases equivalent to sampling singly-truncated data, but that in such cases it is straightforward to sample and count perfect matchings exactly.

Chapter 2

Background

In this chapter we introduce the Markov chain Monte Carlo method for sampling problems, and some of the techniques used to show rapid mixing of Markov chains. We illustrate these techniques by considering the insert-delete chain for sampling independent sets, which was originally shown to be rapidly mixing in [31], and the Jerrum-Sinclair chain for sampling matchings, which was originally shown to be rapidly mixing in [25].

2.1 Counting and sampling

Traditional complexity analysis is concerned largely with decision problems: given a problem instance, does some condition hold? Such a problem can be seen as a boolean function $\varphi : \Sigma^* \to \{0,1\}$ over some alphabet Σ. The complexity class P is the set of decision problems that can be computed by a deterministic Turing machine in polynomial time, while NP is the set of decision problems that can be computed by a nondeterministic Turing machine in polynomial time [35].

We say that a decision problem represented by a function φ over an alphabet Σ is *reducible* to another function φ' over an alphabet Σ' if there is a function $f : \Sigma^* \to \Sigma'^*$ such that for all $x \in \Sigma^*$, $\varphi(x) \Leftrightarrow \varphi'(f(x))$. If φ is a boolean function and every problem in NP is reducible to φ by some function that is computable in time polynomial in the size of the input, then φ is said to be NP-hard. If φ is also in NP, then it is said to be NP-complete. This form of reducibility is known as Karp reducibility.

An alternative way of defining NP is to consider a "witness-checking" predicate $\chi : \Sigma^* \times \Sigma^* \to \{0,1\}$. A function φ is in NP if and only if there exists such a predicate χ that is computable in polynomial time by a deterministic Turing machine, and a

polynomial p such that for all $x \in \Sigma^*$,

$$\varphi(x) \Leftrightarrow \exists w \in \Sigma^* : \chi(x,w) \wedge |w| \leq p(|x|).$$

In the theory of counting, sometimes we are interested in counting *how many* witnesses there are for a given problem instance. This question arises naturally when we consider decision problems where witnesses are combinatorial structures such as independent sets, matchings or graph colourings. These counting problems can be represented as a function $f : \Sigma^* \to \mathbb{N}$. We say that $f \in$ FP if f can be computed by a deterministic Turing machine transducer (that is, a Turing machine equipped with a write-only output tape [35]) in polynomial time. Using a witness-checking characterisation, the complexity class #P can be seen as a counting analogue of NP. Instead of determining whether a witness exists, we are interested in how many witnesses exist: $f \in$ #P if and only if there exists a polynomial-time checkable predicate χ and a polynomial p as before, such that for all $x \in \Sigma^*$,

$$f(x) = |\{w \in \Sigma^* : \chi(x,w) \wedge |w| \leq p(|x|)\}|.$$

It is immediately clear that there are problems in #P that are at least as hard as those in NP: if we know how many witnesses exist for a given problem instance, we can certainly say whether at least one witness exists.

We can construct a hierarchy of counting problems in a similar way to decision problems, by demonstrating reducibility between problems. Instead of Karp reducibility, we use Turing reducibility. A counting problem f is *Turing reducible* to a problem f' if there exists a Turing machine that computes f, given an oracle for f'. The key difference between Karp reducibility and Turing reducibility is that a Turing reduction is allowed to make many calls to f', whereas a Karp reduction is allowed to make only one call to φ'.

A function f is said to be #P-hard if every problem in #P is Turing reducible to f in time polynomial in the input size, and #P-complete if, additionally, $f \in$ #P. The standard reduction of Cook, which shows how to reduce any problem in NP to SAT [18], preserves the number of witnesses of the original problem instance. Therefore the counting analogue #SAT is #P-complete. Generally speaking, if there is a Turing reduction from any #P-complete problem to some problem f that preserves the number of witnesses, then f is #P-hard. Such witness-preserving reductions are said to be *parsimonious* in the literature.

Many other NP-complete decision problems give rise to #P-complete counting problems, although it is an open problem whether all NP-complete problems do

so [24]. However, there also exist #P-complete problems that arise from non-NP-complete decision problems. One such problem is that of counting perfect matchings of a graph.

Definition 2.1. *Given a graph $G = (V, E)$, a* matching *is a subset of edges $M \subseteq E$ such that no two edges in M share a common endpoint. A* perfect matching *is a matching that covers all vertices of G, that is, $\bigcup_{(u,v)\in M}\{u, v\} = V$.*

There are polynomial-time algorithms to determine whether a perfect matching exists in a graph G, for both bipartite and general graphs [15]. However, counting perfect matchings exactly is #P-complete [40].

A polynomial-time algorithm for any #P-complete problem would imply that #P=FP. It is generally believed that no such algorithm exists. However, for some #P-complete problems it is possible to *approximately* count the number of solutions to a problem, to within a constant factor, in polynomial time.

Definition 2.2. *A* fully polynomial randomised approximation scheme *(FPRAS) for a problem f is a randomised algorithm such that, for every problem instance x, and for $\varepsilon > 0$, the value X returned by the algorithm satisfies $e^{-\varepsilon} f(x) \leq X \leq e^{\varepsilon} f(x)$ with probability at least $\frac{3}{4}$, and such that the running time is a polynomial in ε^{-1} and the size of the problem instance, $|x|$.*

A problem related to that of counting solutions to a problem instance is that of sampling uniformly from the set of solutions. It is often possible to reduce from approximating the number of solutions to a given problem to almost-uniform sampling from its solution set [27].

As an example, suppose we wish to approximately count the number of matchings of a graph $G = (V, E)$, with $|E| = m$. If we choose an (arbitrary) ordering of edges $e_1, \ldots, e_m$, then we can define a set of subgraphs:

$$G_i = (V, \{e_1, \ldots, e_i\}).$$

For any $i > 1$, we know that the matchings of G_i are a superset of the matchings of G_{i-1}, and that for any matching M of G_i, M is a matching of G_{i-1} if and only if $e_i \notin M$. Define Ω_i to be the set of matchings of G_i. We can therefore estimate the ratios $\frac{|\Omega_{i-1}|}{|\Omega_i|}$, by taking a sufficiently large number of samples from the set of matchings of G_i and observing the proportion that contain e_i. In order to obtain a good estimate of this ratio, we also need to know a lower bound on the ratio value, for every i. This can be shown

to be $\frac{1}{2}$, by noting that the set of matchings of G_i that do not contain e_i is exactly the set of matchings of G_{i-1}, and that there is an injective function from matchings M of G_i that do contain e_i to matchings $M \setminus \{e_i\}$ of G_{i-1}. Given this ratio, and the approximate number of matchings of G_{i-1}, we can estimate the number of matchings of G_i. Observing that G_0 has exactly one matching (the empty set), if we know the ratios $\frac{|\Omega_{i-1}|}{|\Omega_i|}$ for all i, we can estimate the number of matchings of G, $|\Omega|$:

$$|\Omega| = \left\{ \frac{|\Omega_{m-1}|}{|\Omega_m|} \cdots \frac{|\Omega_0|}{|\Omega_1|} \right\}^{-1}.$$

The accuracy of this estimate of the number of matchings of G depends on the number of samples taken to estimate $\frac{|\Omega_{i-1}|}{|\Omega_i|}$, and the accuracy of our sampling procedure (if we do not have a mechanism for exact sampling). In order to analyse almost-uniform sampling algorithms, we need to be able to measure the distance of a probability distribution from the uniform distribution.

Definition 2.3. *Let π and π' be two probability distributions on a set Ω. The* total variation distance $\|\pi - \pi'\|_{\mathrm{TV}}$ *between π and π' is defined as*

$$\|\pi - \pi'\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\omega \in \Omega} |\pi(\omega) - \pi'(\omega)| = \max_{A \subseteq \Omega} |\pi(A) - \pi'(A)|. \tag{2.1}$$

We can now use the total variation distance to define a class of algorithms for almost-uniform sampling.

Definition 2.4. *Let x be an instance of a sampling problem, and $W(x)$ be the set of witnesses to x, that is,*

$$W(x) = \{w \in \Sigma^* : \chi(x,w) \wedge |w| \leq p(|x|)\}.$$

A fully polynomial almost uniform sampler *(FPAUS) is a randomised algorithm that returns a random sample from a distribution whose total variation distance from the uniform distribution on $W(x)$ is no more than ε, and which runs in time polynomial in ε^{-1} and the size of the problem instance $|x|$.*

Returning to the problem of approximately counting matchings, if we can find an FPAUS for approximate uniform sampling of matchings in a graph G, then it is possible to estimate the ratios $\frac{|\Omega_{i-1}|}{|\Omega_i|}$, and thus approximate the number of matchings of G. This reduction was first shown by Valiant in [40], and the following calculations will summarise the steps of Jerrum's presentation of this proof in [24]. For each ratio, let Z_i be the indicator variable of the event that a sample M_i from Ω_i (selected according

to the distribution of our FPAUS) belongs to Ω_{i-1}. Let $\mu_i = \mathbb{E}[Z_i]$. If we set $\delta = \frac{\varepsilon}{6m}$, then from Definition 2.1 we have

$$\frac{|\Omega_{i-1}|}{|\Omega_i|} - \frac{\varepsilon}{6m} \leq \mu_i \leq \frac{|\Omega_{i-1}|}{|\Omega_i|} + \frac{\varepsilon}{6m}.$$

If we take some set of s samples from Ω_i, then the sample mean $\overline{Z_i}$ that we obtain from these samples will approximate μ_i. Specifically, let $s = \lceil 74\varepsilon^{-2}m \rceil$. With this number of samples, we obtain the following bound on the variance of $\overline{Z_i}$, using the previous bound that $\frac{|\Omega_{i-1}|}{|\Omega_i|} \geq \frac{1}{2}$:

$$\frac{\mathrm{Var}[\overline{Z_i}]}{\mu_i^2} \leq \frac{2}{s} \leq \frac{\varepsilon^2}{37m}.$$

We now consider the variance of the product of the $\overline{Z_i}$ values, and obtain the bound

$$\frac{\mathrm{Var}[\overline{Z_1 Z_2} \cdots \overline{Z_m}]}{(\mu_1 \mu_2 \cdots \mu_m)^2} \leq \frac{\varepsilon^2}{36}.$$

Therefore, by Chebyshev's inequality,

$$\left(1 - \frac{\varepsilon}{3}\right) \mu_1 \mu_2 \cdots \mu_m \leq \overline{Z_1 Z_2} \cdots \overline{Z_m} \leq \left(1 + \frac{\varepsilon}{3}\right) \mu_1 \mu_2 \cdots \mu_m,$$

with probability at least $\frac{3}{4}$. Applying our bound on the μ_i values with respect to $\frac{\Omega_{i-1}}{\Omega i}$, we find that with probability at least $\frac{3}{4}$,

$$e^{-\varepsilon} |\Omega|^{-1} \leq \overline{Z_1 Z_2} \cdots \overline{Z_m} \leq e^{\varepsilon} |\Omega|^{-1}.$$

Therefore the algorithm for computing $(\overline{Z_1 Z_2} \cdots \overline{Z_m})^{-1}$ is an FPRAS for $|\Omega|$. The runtime of this algorithm is $sm \leq 75\varepsilon^{-2}m^2$ times the runtime of our FPAUS for sampling matchings.

In section 2.3.5 we will demonstrate that the Jerrum-Sinclair chain for sampling matchings is an FPAUS, and can therefore be used to estimate $|\Omega|$ in polynomial time.

One method of approximate sampling is Markov chain simulation. In Markov chain Monte Carlo (MCMC), we define a Markov chain with a unique stationary distribution π equal to the desired distribution. By simulating the chain for a sufficiently large number of steps, we can obtain a sample from a distribution very close to π. When we apply Markov chain simulation, we need to prove that the chain approaches the stationary distribution fairly quickly.

2.2 Markov chain Monte Carlo algorithms

Many sampling problems can be approximated using the Markov chain Monte Carlo method.

Definition 2.5. *A* Markov chain $\mathcal{M}$ *on some state space* Ω *is a stochastic process* $X_1, X_2, \ldots$, *where* $X_i \in \Omega$ *for all i, such that for all n,*

$$\Pr(X_{n+1} = x \mid X_n = x_n, \ldots, X_1 = x_1) = \Pr(X_{n+1} = x \mid X_n = x_n).$$

Informally, for any $t \geq 0$, X_{t+1} depends only on X_t.

Definition 2.6. *A Markov chain* $\mathcal{M}$ *is said to be* aperiodic *if, for all possible states x,*

$$\gcd\{n : \Pr(X_n = x \mid X_0 = x) > 0\} = 1.$$

$\mathcal{M}$ *is said to be* irreducible *if for every pair of states x and y, there is a sequence of transitions, each with non-zero probability, from x to y. A Markov chain that is both irreducible and aperiodic is called* ergodic.

A *stationary distribution* of a Markov chain with transition matrix P is a probability distribution π, such that $\pi P = \pi$. If a Markov chain is ergodic, then it has a unique stationary distribution [24]. Furthermore, for any initial probability distribution π_0, the induced distribution after many steps of an ergodic Markov chain approaches π:

$$\forall \pi_0 \lim_{t \to \infty} \pi_0 P^t = \pi.$$

We can now use these definitions to define Markov chains for sampling from specific distributions. To sample from a distribution π on a sample space Ω', we define an ergodic Markov chain whose state space Ω contains Ω' and whose stationary distribution is the desired distribution π. Usually Ω is exactly the sample space, but in some cases we choose a chain where Ω is larger than Ω', and $\pi(\Omega)$ is larger than $\pi(\Omega')$ by no more than a polynomial factor. Usually these conditions are equivalent, but there are some chains for sampling perfect matchings where Ω is exponentially larger than Ω', whereas $\pi(\Omega)$ is only polynomially larger than $\pi(\Omega')$ [26]. By simulating the chain for a sufficiently large number of steps, we can obtain a random sample distributed (almost) according to the stationary distribution. As we will see, in some cases - such as sampling matchings and independent sets - we work in a more general setting than just the world of uniform distributions, and sample from weighted distributions that are not uniform.

Definition 2.7. *Let P be the transition matrix of a Markov chain* $\mathcal{M}$ *on state space* Ω*, and let* π' *be a probability distribution on* Ω*. If for all* $x, y \in \Omega$*,*

$$\pi'(x)P(x,y) = \pi'(y)P(y,x), \tag{2.2}$$

then π' is a stationary distribution of the Markov chain. This condition is known as detailed balance. *If Condition* (2.2) *holds for a given π' and for all $x, y \in \Omega$, and the Markov chain is ergodic, then the chain is said to be* time reversible.

Lemma 2.8. *[24, Lemma 3.7] Let $\mathcal{M}$ be a time-reversible Markov chain with respect to some distribution π'. π' is a stationary distribution of $\mathcal{M}$. If $\mathcal{M}$ is ergodic, then π' is the unique stationary distribution π.*

If we know that a chain is time-reversible with respect to a distribution π', then we can easily check whether it samples from the desired distribution. We will generally only use time reversible Markov chains for sampling.

As an example, consider the problem of sampling independent sets of a graph $G = (V, E)$.

Definition 2.9. *An* independent set *of a graph $G = (V, E)$ is a subset of vertices $I \subseteq V$ such that there is no pair of vertices $u, v \in I$ for which $(u, v) \in E$.*

We will consider a Markov chain $\mathcal{M}_{ID}$ whose state space Ω is the set of all independent sets of G, and which samples independent sets weighted according to a parameter λ. This chain was proposed by Luby and Vigoda [30], and analysed by Dyer and Greenhill [14]. Our goal is that the stationary distribution should be

$$\pi(I) = \frac{\lambda^{|I|}}{\sum_{I' \in \Omega} \lambda^{|I'|}}. \tag{2.3}$$

The denominator $\sum_{I' \in \Omega} \lambda^{|I'|}$ is known as the partition function. Note that if $\lambda = 1$ then π is the uniform distribution, and the partition function is equal to the number of independent sets of G. We now define the transitions of the chain $\mathcal{M}_{ID}$.

Definition 2.10. *$\mathcal{M}_{ID}$ is referred to as the insert-delete chain for sampling independent sets. Let X_t be the state of $\mathcal{M}_{ID}$ at time t. The state X_{t+1} at time $t+1$ is determined by the following process:*

(ID1) Select a vertex $v \in V$ uniformly at random.

(ID2) With probability $\frac{\lambda}{1+\lambda}$, let $I = X_t \cup \{v\}$; with the remaining probability $\frac{1}{1+\lambda}$, let $I = X_t \setminus \{v\}$.

(ID3) If I is an independent set, then let $X_{t+1} = I$; otherwise, let $X_{t+1} = X_t$.

We will now show that the stationary distribution of $\mathcal{M}_{ID}$ is the distribution π defined in Equation (2.3).

Lemma 2.11. *The unique stationary distribution of $\mathcal{M}_{ID}$ is*

$$\pi(I) = \frac{\lambda^{|I|}}{\sum_{I' \in \Omega} \lambda^{|I'|}}.$$

Proof. Let Ω be the state space of $\mathcal{M}_{ID}$ - that is, the set of all independent sets of G. We will demonstrate that our desired distribution π satisfies the detailed balance condition (2.2).

Let $n = |V|$ be the number of vertices of G. Consider a pair of independent sets $(I, I') \in \Omega \times \Omega$. If I and I' differ at more than one vertex, then clearly $P(I,I') = P(I',I) = 0$, and so Condition (2.2) holds. Likewise, if $I = I'$ then trivially $\pi(I) = \pi(I')$ and $P(I,I') = P(I',I)$. Now suppose that I and I' differ at exactly one vertex v, and assume without loss of generality that $v \in I$ and $v \notin I'$. Then $\pi(I) = \lambda\pi(I')$, $P(I,I') = \frac{1}{n(1+\lambda)}$, and $P(I',I) = \frac{\lambda}{n(1+\lambda)}$. Now

$$\begin{aligned}
\pi(I)P(I,I') &= \lambda\pi(I')\frac{1}{n(1+\lambda)} \\
&= \pi(I')\frac{\lambda}{n(1+\lambda)} \\
&= \pi(I')P(I',I).
\end{aligned}$$

Condition (2.2) therefore holds for the distribution π for all pairs of states.

The uniqueness of π follows from the fact that every independent set is reachable by a sequence of transitions from the empty set, and that there is a non-zero self-loop probability for all states (note that for every $x \in \Omega$, $P(x,x)$ is at least $\min\{\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\}$). Therefore the chain is ergodic, and since Condition (2.2) holds, it is time-reversible with respect to π. By Lemma 2.8, π is the unique stationary distribution of $\mathcal{M}_{ID}$. □

By definition, any ergodic Markov chain will approach its stationary distribution as the number of steps $t \to \infty$. We are interested, however, in *how long* it takes for the total variation distance from π to fall below some $\varepsilon > 0$.

Definition 2.12. *Let $\mathcal{M}$ be an ergodic Markov chain with transition matrix P and stationary distribution π. The* mixing time *of $\mathcal{M}$, for a given initial state x, is the number of steps required for the total variation distance to fall below some value $\varepsilon > 0$:*

$$\tau_x(\varepsilon) = \min\{t : \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\}.$$

We are usually interested in the mixing time of a Markov chain for an arbitrary start state. This is obtained by considering the maximum mixing time over all states:

$$\tau(\varepsilon) = \max_{x \in \Omega} \min\{t : \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\}. \tag{2.4}$$

If the mixing time $\tau(\varepsilon)$ is polynomial in the size of the problem instance, then we refer to the Markov chain as *rapidly mixing*.

In the next section we will present a proof of rapid mixing for $\mathcal{M}_{ID}$, in the case of bounded degree and small λ, in order to demonstrate the use of the coupling technique.

2.3 Methods for bounding mixing times

While there are several methods of bounding mixing times, we focus here on those based on two main techniques: those based on coupling arguments, and those based on the conductance and spectral gap of Markov chains.

2.3.1 Coupling

Coupling was discovered by Doeblin in the 1930s, and its use for bounding the mixing time of Markov chains was introduced by Aldous [1].

Let $\mathcal{M}$ be a Markov chain on state space Ω, and P be the transition matrix of $\mathcal{M}$. We consider two copies of $\mathcal{M}$, X and Y. A coupling of X and Y is a stochastic process $Z_t = (X_t, Y_t)$ on the state space $\Omega^\omega \times \Omega^\omega$, such that the marginal distribution on X_t (and, respectively, Y_t) is identical to that of $\mathcal{M}$, for every t.

If Z_t is itself a Markov chain, then we say that Z_t is a Markovian coupling. In this instance, Z_t can be defined on the state space $\Omega \times \Omega$, and the following conditions hold:

$$\Pr(X_{t+1} = x' \mid X_t = x, Y_t = y) = P(x, x'),$$
$$\Pr(Y_{t+1} = y' \mid X_t = x, Y_t = y) = P(y, y').$$

In this thesis we will consider only Markovian couplings. However, there are also techniques for finding non-Markovian couplings and using them to show rapid mixing [21, 8].

The coupling lemma allows us to use coupling to bound the mixing time of Markov chains.

Lemma 2.13 (Coupling lemma [1])**.** *Let $\mathcal{M}$ be a Markov chain on state space Ω, and $Z = (X, Y)$ be a coupling of two copies of $\mathcal{M}$. Let $t(\varepsilon)$ be a function such that for all $x, y \in \Omega$,*

$$\Pr(X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y) \leq \varepsilon.$$

Then the mixing time of $\mathcal{M}$ is bounded above by $t(\varepsilon)$.

We can use Lemma 2.13 to bound the mixing time of a Markov chain $\mathcal{M}$ if we are able to define a coupling where the two copies X_t and Y_t tend to converge with increasing t. If we work with the basic coupling lemma above, it is necessary to define and analyse the coupling over all possible pairs of states of $\mathcal{M}$. However, other techniques such as path coupling allow us to show rapid mixing by considering a coupling on a subset of pairs of states.

When we apply coupling, and especially techniques such as path coupling, we often use a metric to measure the distance between the two copies X_t and Y_t. We say that the distance between a pair of states x and y *contracts* if, for some $\beta < 1$,

$$\mathbb{E}[d(X_{t+1}, Y_{t+1})] < \beta d(X_t, Y_t) \text{ when } X_t = x, Y_t = y.$$

A Markov chain $\mathcal{M}$ is rapidly mixing if and only if there exists a coupling that contracts in time polynomial in the input size [20]. However, such a coupling need not be Markovian, does not have to contract in every step, and may be difficult to find. In Section 2.3.4 we will introduce alternative techniques that may be applied even when a Markovian coupling does not exist.

2.3.2 Path coupling

Path coupling is a method introduced by Bubley and Dyer that allows us to use coupling without the requirement of showing contraction for every possible pair of states [7]. Instead of considering all pairs of states, we define an adjacency relation S. We require that every pair of states is connected by a path in the adjacency graph formed by S. The path coupling lemma states that if we have a coupling for which the value of some metric can be shown to contract for all pairs of adjacent states, then a coupling exists that contracts for all (not necessarily adjacent) pairs of states.

Lemma 2.14 (Path coupling lemma [7])**.** *Let $S \subseteq \Omega \times \Omega$ be an adjacency relation on the state space of $\mathcal{M}$, such that for any two states X_t and $Y_t \in \Omega$, there exists a path from X_t to Y_t using only transitions in S. Let d be the path metric defined on $\Omega \times \Omega$ induced by some metric on S. Suppose that for every pair $(X_t, Y_t) \in S$,*

$$\mathbb{E}[d(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq \beta d(X_t, Y_t).$$

Then the contraction condition $\mathbb{E}[d(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq \beta d(X_t, Y_t)$ holds for all pairs $(X_t, Y_t) \in \Omega \times \Omega$.

The path coupling lemma gives a bound on the expected distance at time $t+1$ given the distance at time t. Theorem 2.15 allows us to bound the mixing time of $\mathcal{M}$ given such a bound.

Theorem 2.15 ([13]). *Let (X,Y) be a coupling on $\mathcal{M}$, d an integer-valued metric defined on $\Omega \times \Omega$, and D the maximum value that d can take for any pair of states in Ω. Suppose that for some $\beta \leq 1$,*

$$\mathbb{E}[d(X_{t+1}, Y_{t+1})] \leq \beta d(X_t, Y_t)$$

for all t. If $\beta < 1$ then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{1-\beta}.$$

If $\beta = 1$ and there exists some $\alpha > 0$ such that

$$\Pr(d(X_{t+1}, Y_{t+1}) \neq d(X_t, Y_t)) \geq \alpha$$

for all t, then the mixing time satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

The case where $\beta = 1$ of Theorem 2.15 concerns the situation where the coupling cannot be shown to contract in all cases, but can be shown not to expand. In this situation, we additionally need to show that there is a minimum probability of a change in distance *for all pairs of states*. This is often easier than showing contraction over all pairs of states, but finding good bounds on the probability of d changing may be difficult. However, in more recent work, Bordewich and Dyer have shown that it is often sufficient to show that there is variance only for adjacent states [4]. Their theorem requires that, for every pair of adjacent states (v,w), a single instance of a chain $\mathcal{M}$ starting at state v has a minimum probability of moving at least some distance δ towards state w.

Theorem 2.16 ([4]). *Let $\mathcal{P}$ be a path coupling for a Markov chain $\mathcal{M}$, and let $S \subseteq \Omega^2$ be the adjacency relation on which $\mathcal{P}$ is defined. Let d be the path metric defined on $\Omega \times \Omega$ induced by some metric on S. Let X_t represent the state of $\mathcal{M}$ at time t. Define the function p for all $\delta > 0$ as follows:*

$$p(\delta) = \min_{(v,w)\in S} \Pr(d(v, X_{t+1}) \geq \delta, d(v, X_{t+1}) + d(X_{t+1}, w) = d(v,w) \mid X_t = v),$$

That is, $p(\delta)$ represents the minimum probability over all pairs of adjacent states $(v,w) \in S$ that a single copy of $\mathcal{M}$, starting from state v, moves at least distance δ towards state w.

If $\beta(\mathcal{M},\mathcal{P}) \leq 1$ and there exists a $\delta > 0$ such that $p = p(\delta) > 0$, then there exists a modified Markov chain $\mathcal{M}^*$ with the same stationary distribution as $\mathcal{M}$, and a coupling $\mathcal{C}$ of $\mathcal{M}^*$ such that $\beta(\mathcal{M}^*,\mathcal{C}) \leq 1$ and $\sigma^2(\mathcal{M}^*,\mathcal{C}) \geq \frac{p\delta^2}{1+p}$.

The transition probabilities of $\mathcal{M}^*$ are given by:

$$\Pr_{\mathcal{M}^*}(X_{t+1} = x' \mid X_t = x) = \begin{cases} \frac{\Pr_{\mathcal{M}}(X_{t+1}=x|X_t=x)+p}{1+p} & \text{if } x' = x \\ \frac{\Pr_{\mathcal{M}}(X_{t+1}=x'|X_t=x)}{1+p} & \text{otherwise.} \end{cases}$$

The bound on σ^2 given by Theorem 2.16, combined with the knowledge that $\beta \leq 1$, allows us to obtain a bound on the mixing time of $\mathcal{M}^*$. It is not necessarily the case, however, that we can obtain a bound on the mixing time of the original chain $\mathcal{M}$.

In Chapter 3 we will prove a slightly more general form of Theorem 2.15 in the case where $\beta = 1$, in which we will permit metrics that take non-integer values. We will also show that we can achieve a better bound on the mixing time when $\beta = 1$, if we have a lower bound on the probability that the distance changes that is linear in the distance at time t.

We now return to the insert-delete chain for sampling independent sets. We will demonstrate basic path coupling, using Theorem 2.15, but we will use Theorem 2.16 in Chapter 4. We can use path coupling to show that $\mathcal{M}_{ID}$ is rapidly mixing for small values of λ. Dyer and Greenhill stated this result, but omitted the details of the proof [14]. The result we will prove in Theorem 2.17 is quite weak in terms of the range of λ values covered. Luby and Vigoda also obtained a proof of rapid mixing for $\mathcal{M}_{ID}$, for a larger range of values of λ than we consider [31]. The intent here is to illustrate the technique of path coupling.

Theorem 2.17. *Consider a graph $G = (V,E)$ with maximum degree Δ and $|V| = n$. $\mathcal{M}_{ID}$ is rapidly mixing on G provided $\lambda \leq \frac{1}{\Delta-1}$.*

Proof. We say that two states I and I' are adjacent if and only if I and I' differ at exactly one vertex v:

$$S = \{(I,I') : \exists v \text{ s.t. } I' = I \cup \{v\}; I \text{ is an independent set }\}.$$

There is certainly a path between any arbitrary pair of states, because it is possible to reach the empty set from any state using only transitions between adjacent states. We

will use Hamming distance, that is, the number of vertices that differ between I and I', as our metric. Note that Hamming distance is the path metric that we obtain if we consider states in S to be at distance 1.

Suppose we have two copies of $\mathcal{M}_{ID}$, I_t and I'_t, and that at time t their states differ at a unique vertex v. Assume without loss of generality that $I'_t = I_t \cup \{v\}$. Our path coupling will update the same vertex u in each copy of the chain. If $u = v$, then with probability $\frac{\lambda}{1+\lambda}$ we add u to I_t and make no change to I'_t; with the remaining probability $\frac{1}{1+\lambda}$, we remove u from I'_t and make no change to I_t. Our path coupling therefore couples with probability 1 in this case, and the expected distance at time $t+1$ decreases by 1. If we select a vertex which is neither v nor adjacent to v, then we can make the same transition in each copy of the chain, but the Hamming distance will not change. Finally, if we select a vertex adjacent to v, then there are two cases:

1. If u is adjacent to some other v' which is present in both copies of the chain, then neither copy can change at time $t+1$ and so the Hamming distance is unchanged.

2. If there is no such v', then we can insert u in one copy only, with probability $\frac{\lambda}{1+\lambda}$. The expected distance therefore increases by $\frac{\lambda}{1+\lambda}$.

There are at most Δ choices of u for which case 2 may occur. In the worst case, therefore, the expected distance increases by $\frac{1}{n}(\frac{\Delta\lambda}{1+\lambda} - 1)$.

Our choice of adjacency relation means that whenever the pair of states I_t and I'_t are adjacent, $d(I_t, I'_t) = 1$. Therefore, by Lemma 2.14,

$$\mathbb{E}[d(I_{t+1}, I'_{t+1})] \leq \left(1 + \frac{\lambda(\Delta-1)-1}{(1+\lambda)n}\right) d(I_t, I'_t),$$

and so the contraction ratio β is

$$\beta = 1 + \frac{\lambda(\Delta-1)-1}{(1+\lambda)n}.$$

The contraction condition holds whenever $\beta \leq 1$, and this occurs whenever $\lambda \leq \frac{1}{\Delta-1}$.

We now apply Theorem 2.15 for the case when $\lambda < \frac{1}{\Delta-1}$, that is, when $\beta < 1$. The mixing time of the insert-delete chain in this case satisfies

$$\tau(\varepsilon) \leq \frac{n\log(n\varepsilon^{-1})(1+\lambda)}{1-\lambda(\Delta-1)},$$

observing that the maximum value of $d(I_t, I'_t)$ is the number of vertices n.

In the case $\lambda = \frac{1}{\Delta-1}$, observe that if $d(I_t, I'_t) \neq 0$ then there is always at least one vertex v that differs between I_t and I'_t, and that if v is selected then the distance changes with probability at least $\frac{1}{1+\lambda}$. In this boundary case we therefore have

$$\begin{aligned}\alpha &= \frac{1}{n(1+\lambda)} \\ &= \frac{1}{n(1+\frac{1}{\Delta-1})} \\ &= \frac{\Delta-1}{n\Delta}.\end{aligned}$$

Applying the second part of Theorem 2.15, we obtain the following bound on the mixing time when $\lambda = \frac{1}{\Delta-1}$:

$$\tau(\varepsilon) \leq \left\lceil \frac{en^3\Delta}{\Delta-1} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

□

Dyer and Greenhill considered a more complex chain, introducing an additional *drag* move, and showed that their chain mixes rapidly for a wider range of values of λ than we show here [14]. The drag move removes one vertex and inserts an adjacent vertex in a single transition. They further showed that their results imply that the insert-delete chain mixes rapidly for the same values of λ as they showed for their chain. Dyer and Greenhill's drag move allows us to make a move in both copies of the chain in the case of the proof of Theorem 2.17 where the expected distance increases. The probability of the drag move in their chain is optimised to minimise the expected increase in distance [14]. Dyer and Greenhill were able to show that their chain is rapidly mixing where $\lambda \leq \frac{2}{\Delta-2}$. In Chapter 4 we will mimic their approach and introduce a similar transition in a Markov chain for sampling matchings, to obtain improved mixing times for matching chains.

In Chapter 3 we will develop a path coupling theorem that can be applied when we can show that the probability of a change in distance is proportional to the current distance, as is the case for $\mathcal{M}_{ID}$ when $\lambda = \frac{1}{\Delta-1}$. Our new theorem can be applied to obtain a better bound on the mixing time than Theorem 2.15 in these cases.

2.3.3 Spatial mixing

Spatial mixing is a property that can be shown to hold on spin systems, which can be used to infer rapid mixing of Markov chains on states of the spin systems. We will derive spatial mixing results on a certain set of spin systems in Chapters 4 and 5.

Definition 2.18. *A* spin system *is a stochastic system consisting of a graph $G = (V,E)$ and a set of* spins *Q, such that each vertex $v \in V$ is assigned a spin $q_v \in Q$. The set of spins that each vertex can take may be constrained, and such constraints may be* hard constraints *(the spins that v can take, conditioned on the spins of neighbouring vertices, are a subset of Q) or* soft constraints *(v can take any spin, but with varying probabilities, which may depend on the spins of the vertices in its neighbourhood).*

We can think of a spin system as being a Markov chain on the state space $Q^{|V|}$ that chooses a vertex uniformly at random, and replaces its spin with one chosen from the distribution induced by the spins of its neighbours. Spin systems with both hard and soft constraints can be modelled in this way. Examples of spin systems include graph colourings and the Ising model. We will define proper colourings of a graph G, which are an example of a spin system with hard constraints, in Chapter 3, and show that a simple Markov chain for sampling proper colourings is rapidly mixing when $|Q| = 2\Delta$, where Δ is the maximum degree of G. The Ising model, which we will not consider further in this thesis, is an example of a spin system with soft constraints. In the Ising model, the set of spins is $Q = \{1,-1\}$, and the probability that a vertex has a given spin is a function of the numbers of adjacent vertices with the same and opposite spins, referred to as the *energy function*.

Strong spatial mixing is a property which can be shown to hold for some spin systems, whereby the effect of a single-site discrepancy at a vertex y on any set of vertices Λ decays exponentially with the distance of Λ from y [44]. We describe this property formally in what follows:

Consider a graph $G = (V,E)$. We will refer to a finite subset of vertices $R \subseteq V$ as a *region* of G. The *boundary* of a region R is the set of vertices that belong to R and are adjacent to at least one vertex of $V \setminus R$.

Definition 2.19 ([44])**.** *Let $G = (V,E)$ be a graph. Let $R \subseteq V$ be a non-empty finite region of G, $\Lambda \subseteq R$ a subset of R. Let ∂R be the set of vertices on the boundary of R, and $\Delta \subseteq \partial R$ be a subset of these vertices. Let $\mathcal{B}$ and $\mathcal{B}'$ be a pair of configurations of ∂R differing only at the vertices of Δ. Let $\pi_{\mathcal{B},\Lambda}$ and $\pi_{\mathcal{B}',\Lambda}$ be the stationary distributions of the region Λ, conditioned on the configurations $\mathcal{B}$ and $\mathcal{B}'$ respectively. A spin system has* strong spatial mixing *if there exist constants $\beta, \beta' > 0$ such that for every R, Λ, Δ, $\mathcal{B}$ and $\mathcal{B}'$,*

$$\|\pi_{\mathcal{B},\Lambda} - \pi_{\mathcal{B}',\Lambda}\|_{\mathrm{TV}} \leq \beta|\Lambda|e^{-\beta' d(\Delta,\Lambda)}, \tag{2.5}$$

where $d(\Delta,\Lambda)$ is the minimum distance from Δ to Λ.

The reason that we are interested in strong spatial mixing is because of its connection to the mixing time of Markov chains. If the vertex neighbourhood of every $v \in V$ grows sub-exponentially with increasing distance, strong spatial mixing implies the existence of a rapidly mixing Markov chain for sampling configurations of the spin system [19]. The proof of this connection is done by constructing a chain $\mathcal{M}$ with the same distribution as the original spin system. The chain $\mathcal{M}$ selects a vertex v uniformly at random, and replaces the configuration of a fixed radius region R around v with one selected from the uniform distribution, conditioned on the configuration of ∂R. We can use path coupling to show that the chain is rapidly mixing if the region R is large enough.

In our applications in Chapters 4 and 5, we will demonstrate that the spatial mixing property holds as part of a path coupling argument, where adjacent states differ at a single vertex, y. We therefore need to consider only the case where $\Delta = \{y\}$. Weitz notes that if G is an integer lattice $\mathbb{Z}^d$, then this weaker definition where Δ contains only one element is equivalent to Definition 2.19, but that this is not the case in general [44].

The proof that strong spatial mixing implies rapid mixing proceeds along the following lines:

Let X_t and Y_t be a pair of states of $\mathcal{M}$, differing at a single vertex w, and consider X_{t+1} and Y_{t+1}. We will use Hamming distance as a metric. If w falls within the region R that is updated, then the configurations of ∂R in X_t and Y_t are identical and we can select $X_{t+1} = Y_{t+1}$. Likewise, if w is outside R and is not on the boundary ∂R, then we can update R with the same configuration in each copy of the chain, and so the Hamming distance does not increase. Finally, if w lies on ∂R, the spatial mixing property ensures that the increase in Hamming distance is small. If $|R|$ is sufficiently large compared to $|\partial R|$, then we can show that $\mathbb{E}[d(X_{t+1}, Y_{t+1})] \leq d(X_t, Y_t)$, and hence that $\mathcal{M}$ is rapidly mixing.

In practice, the large block transitions required to show rapid mixing of $\mathcal{M}$ can be difficult to compute. The number of configurations of the boundary ∂R grows exponentially as the size of R increases, so it becomes infeasible to enumerate these directly. Van den Berg and Brouwer used a second Markov chain to update the configuration of R, using single-site moves, to show rapid mixing of a Markov chain for sampling matchings [42]. This results in the new configuration of R being chosen from an almost uniform distribution, which they showed is sufficient for the upper-level chain to mix rapidly. They also noted that it is possible to use comparison arguments, which we

will see in Section 2.3.6, to bound the mixing time of a single-site chain with the same stationary distribution, given a bound on the mixing time of a chain using block moves. We follow this approach in Chapters 4 and 5, and use bounds on the mixing time of Markov chains using block moves to bound the mixing time of single-site chains.

2.3.4 Conductance and the spectral gap

When a coupling with contraction ratio $\beta \leq 1$ can be found for a Markov chain, coupling generally gives good bounds on the mixing time of the chain. Indeed, for every rapidly mixing Markov chain, there exists some coupling that bounds its mixing time [20]. However, such a coupling may not be Markovian, and it is not always possible to find a Markovian coupling that contracts. For example, Kumar and Ramesh have shown that there is no Markovian coupling that contracts for the Jerrum-Sinclair chain for sampling matchings [28]. Conductance provides an alternative way of bounding mixing times. However, we require that a Markov chain be time-reversible if we are to use conductance to bound its mixing time. All of the Markov chains we consider will satisfy the detailed balance condition (2.2), and will therefore be time-reversible.

The conductance of a Markov chain gives an indication of how easy it is for the chain to leave any subset of states. Intuitively, a high conductance implies a low probability that the chain can get "stuck" in a small part of the state space.

Definition 2.20. *Let $\mathcal{M}$ be an ergodic time-reversible Markov chain on state space Ω, and let P and π be the transition matrix and stationary distribution of $\mathcal{M}$, respectively. For any non-empty subset $S \subset \Omega$, let $C_S = \sum_{x \in S} \pi(x)$ be the total probability of S, and $F_S = \sum_{x \in S, y \notin S} \pi(x)P(x,y)$ be the ergodic flow out of S. Now let*

$$\Phi_S = \frac{F_S}{C_S} = \frac{\sum_{x \in S, y \notin S} \pi(x)P(x,y)}{\pi(S)}.$$

Then the conductance of $\mathcal{M}$ is

$$\Phi = \min_{S : 0 < \pi(S) \leq \frac{1}{2}} \Phi_S.$$

The use of conductance to bound mixing time is justified by its relationship to the *spectral gap* of a Markov chain. The spectral gap of an ergodic Markov chain $\mathcal{M}$ is $1 - |\lambda_1|$, where λ_1 is the second largest eigenvalue in absolute value of its transition matrix (note that since $\mathcal{M}$ has a unique stationary distribution, the largest eigenvalue is 1). If we can bound the spectral gap, then it is possible to bound the mixing time of the Markov chain.

Theorem 2.21 ([38]). *Let $\mathcal{M}$ be an ergodic Markov chain on state space Ω, and let $\lambda_{\min} < \cdots < \lambda_2 < \lambda_1 < 1$ be the eigenvalues of the transition matrix of $\mathcal{M}$. Let $\pi_* = \min_{x\in\Omega} \pi(x)$ be the smallest probability of any state in the stationary distribution π of $\mathcal{M}$. Then the mixing time of $\mathcal{M}$ satisfies:*

$$\tau(\varepsilon) \leq \frac{1}{1-\max\{|\lambda_{\min}|,|\lambda_1|\}} \log\left(\frac{1}{\pi_*\varepsilon}\right)$$
$$\tau(\varepsilon) \geq \frac{|\lambda_1|}{2(1-\max\{|\lambda_{\min}|,|\lambda_1|\})} \log\left(\frac{1}{2\varepsilon}\right).$$

Sinclair and Jerrum showed that the spectral gap, and hence the mixing time, of a Markov chain can be bounded in terms of the conductance.

Theorem 2.22 ([39, 11]). *Let $\mathcal{M}$ be an ergodic time-reversible Markov chain with transition matrix P, and let Φ be the conductance of $\mathcal{M}$. The spectral gap $1-\lambda_1$ satisfies*

$$\frac{\Phi^2}{2} \leq 1-\lambda_1 \leq 2\Phi.$$

Note that we assume that λ_1 is positive. If this is not the case, then we can alter the chain to make λ_1 positive, by introducing a uniform self-loop probability to each state. This does not affect the stationary distribution, and increases the mixing time of the chain by only a small constant factor.

Theorems 2.21 and 2.22 provide both upper and lower bounds on the spectral gap and mixing time. We can obtain a lower bound on the mixing time of a chain if we are able to show that the conductance is small. To show this, it is sufficient to find a single set S such that Φ_S is small. We will use this approach in Chapter 6 to show that the mixing time of a particular Markov chain for sampling perfect matchings is exponential, for certain types of graph. We will not generally use conductance directly to find upper bounds on the mixing time. Instead, we use a related value, the congestion, which tends to give better upper bounds on the mixing time than those obtained by attempting to bound conductance directly.

2.3.5 Congestion and canonical paths

We can bound the conductance of a Markov chain $\mathcal{M}$ by considering a suitable multicommodity flow problem. However, we will instead use this approach to bound a different value, the congestion. In most cases, the congestion allows us to obtain a tighter upper bound on the mixing time than that obtained by using canonical paths to bound conductance [38].

The canonical paths approach works as follows:

For each pair of states x and y in Ω, we define a *canonical path* γ_{xy} from x to y, using the transitions of $\mathcal{M}$. We define

$$\Gamma = \{\gamma_{xy} \mid x, y \in \Omega\}$$

to be the set of canonical paths, and for each transition t of $\mathcal{M}$, we define the function $\mathrm{cp}(t)$ to be the set of canonical paths that contain t. Note that there may be many possible choices of γ_{xy} for a particular pair of states. Hence there is a lot of freedom in how we design a set of canonical paths for a Markov chain $\mathcal{M}$. Given a particular set of canonical paths Γ, we can now define the congestion:

Definition 2.23. *Let $\mathcal{M}$ be an ergodic time-reversible Markov chain with stationary distribution π and transition matrix P, and Γ be a set of canonical paths for $\mathcal{M}$. The congestion $\rho(\Gamma)$ of $\mathcal{M}$ with respect to Γ is defined as*

$$\rho(\Gamma) = \max_{t=(u,v)} \left\{ \frac{1}{\pi(u)P(u,v)} \sum_{(x,y)\in \mathrm{cp}(t)} \pi(x)\pi(y)|\gamma_{xy}| \right\}, \tag{2.6}$$

where $|\gamma_{xy}|$ denotes the length of the path γ_{xy}, and the maximum is over all pairs of states (u,v) such that $P(u,v) > 0$.

To compute the congestion, we need to determine how many paths use each transition of $\mathcal{M}$. We can do this by defining a function η_t for each transition, mapping the start and end states x and y of each path using t to a single state in Ω. We will refer to this function η_t as an *encoding*. If we can show that each η_t is injective, or at least bound the number of paths mapping to each state, then we can give a bound on the congestion [24].

Theorem 2.24 ([12, 10])**.** *Let $\mathcal{M}$ be an ergodic time-reversible Markov chain, and let Γ be any set of canonical paths for $\mathcal{M}$. If the congestion of $\mathcal{M}$ with respect to Γ is ρ, then the mixing time of $\mathcal{M}$ is bounded above by*

$$\tau_x(\varepsilon) \leq 2\rho(\Gamma)(2\log\varepsilon^{-1} + \log\pi(x)^{-1}).$$

We now illustrate the canonical paths method by considering the Jerrum-Sinclair chain for sampling matchings of a graph $G = (V,E)$ [25]. This presentation follows the original proof of mixing of Jerrum and Sinclair [25]. As with independent sets, we are interested in sampling from a distribution weighted according to a parameter λ, with $\lambda = 1$ corresponding to the uniform distribution:

$$\pi(M) = \frac{\lambda^{|M|}}{\sum_{M'\in\Omega} \lambda^{|M'|}}. \tag{2.7}$$

Definition 2.25. *The Jerrum-Sinclair chain $\mathcal{M}_{JS}$ for sampling matchings in a graph $G = (V,E)$ is defined as follows: Let X_t be the state of the chain $\mathcal{M}_{JS}$ at time t. Then the subsequent state X_{t+1} is determined by the following sequence of steps:*

(JS1) Select an edge $e \in E$ uniformly at random.

(JS2) If $e \in X_t$ then let $M = X_t \setminus \{e\}$. If $e \notin X_t$ and $X_t \cup \{e\}$ is a matching then let $M = X_t \cup \{e\}$. If $e \notin X_t$ and there is exactly one $e' \in X_t$ such that e' is adjacent to e, then let $M = X_t \cup \{e\} \setminus \{e'\}$. We refer to this third type of transition as a slide move.

(JS3) With probability $\min\left\{1, \frac{\pi(M')}{\pi(X_t)}\right\}$, let $X_{t+1} = M$. Otherwise, let $X_{t+1} = X_t$. This method of selecting X_{t+1} is known as a Metropolis filter.

The Jerrum-Sinclair chain was inspired by Broder, who first proposed the use of a Markov chain for sampling matchings [5, 6].

In order to bound the congestion of $\mathcal{M}_{JS}$, we analyse a modified version of the chain in which there is an extra self-loop probability of $\frac{1}{2}$ for each state. This modified version is referred to as the *lazy* version of the Jerrum-Sinclair chain. This allows us to easily show that the chain is aperiodic and that all the eigenvalues of its transition matrix are positive. In Theorem 2.26 below, we will show how to construct a set of canonical paths for $\mathcal{M}_{JS}$, and an encoding η_t for this set of paths. In Chapter 5 we will apply a very similar argument to a Markov chain for sampling independent sets in claw-free graphs, and will include the details of the calculations following the definition of η_t.

Theorem 2.26. *Let $G = (V,E)$ be a graph. Let $n = |V|$ and $m = |E|$. The mixing time of $\mathcal{M}_{JS}$ (with the extra self-loop probabilities) on G is bounded above by*

$$\tau(\varepsilon) \leq nm\overline{\lambda}^2(4\log\varepsilon^{-1} + 2n\log n + n|\log\lambda|),$$

where $\overline{\lambda} = \max\{1,\lambda\}$.

Proof. Given a pair of matchings I and F, we define a canonical path γ_{IF} from I to F by considering the symmetric difference $I \oplus F$. This consists of a set of alternating paths and even-length cycles. We define the path from I to F by processing $I \oplus F$ on a component basis, taking components in order of the minimum vertex contained in each component (given some ordering on the vertices of G). For each component, we define a start vertex - either the smallest vertex in the case of a cycle, or the smaller endpoint

in the case of an alternating path. Each cycle will have one edge in I incident to the start vertex: we remove this edge, then proceed around the cycle using slide moves until we reach the other edge incident to the start vertex. We complete the processing of the cycle by adding this last edge. For each path, if there is an edge in I incident to the start vertex then we remove it. We then proceed along the path using slide moves. Once we have processed the rest of the path, if there is an edge in F incident to the end vertex, then we add this edge.

We now have a set of canonical paths for $\mathcal{M}_{JS}$, and wish to bound the congestion of this set of paths. We begin by defining the set $\mathrm{cp}(t)$ for each transition $t = (M, M')$, which contains all pairs of states (I, F) for which t is a transition on the canonical path from I to F:

$$\mathrm{cp}(t) = \{(I,F) \mid t \in \gamma_{IF}\}.$$

For each transition t, we construct an encoding η_t as follows:

- If t is a slide move, $(I,F) \in \mathrm{cp}(t)$, and the component of γ_{IF} that contains the edges affected by t is a cycle, then let $e_{IF,t}$ be the edge in I that is adjacent to the start vertex of the current cycle, and let $\eta_t = (I \oplus F \oplus (M \cup M')) \setminus \{e_{IF,t}\}$.
- Otherwise, let $\eta_t = I \oplus F \oplus (M \cup M')$.

We can show that for all transitions and pairs of states $(I,F) \in \mathrm{cp}(t)$, $\eta_t(I,F)$ is a matching. Furthermore, we can recover I and F if we know t and $\eta_t(I,F)$. It follows that the range of η_t is no larger than Ω, and that η_t is injective. We can now show that

$$\pi(I)\pi(F) \leq m\overline{\lambda}^2 \pi(M)P(M,M')\pi(\eta_t(I,F)),$$

where $\overline{\lambda} = \max\{1, \lambda\}$, and compute the congestion:

$$\begin{aligned}
\rho &= \max_{t=(M,M')} \left\{ \frac{1}{\pi(M)P(M,M')} \sum_{(I,F)\in \mathrm{cp}(t)} \pi(I)\pi(F) \right\} |\gamma_{IF}| \\
&\leq m\overline{\lambda}^2 \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F))|\gamma_{IF}| \\
&\leq nm\overline{\lambda}^2.
\end{aligned}$$

The bound on the mixing time follows, noting that $\log \pi(x)^{-1} \leq n \log n + \frac{1}{2} n |\log \lambda|$. □

In Chapter 4, we will analyse a modified version of this chain using path coupling to show that the modified chain is also rapidly mixing for small values of λ. Our aim is

to obtain an improved bound on the mixing time of our modified chain over the bound given by canonical paths for the Jerrum-Sinclair chain. As we will see in Chapter 4, however, the bound we are currently able to obtain is asymptotically worse than that of Theorem 2.26.

2.3.6 Markov chain comparison

If we have two Markov chains $\mathcal{M}$ and $\mathcal{M}'$ on the same state space, with transition probabilities P and P' respectively, such that the transitions of $\mathcal{M}'$ can be encoded as sequences of transitions of $\mathcal{M}$, and $\mathcal{M}$ and $\mathcal{M}'$ have the same stationary distribution π, then it is often possible to bound the mixing time of $\mathcal{M}$ in terms of the mixing time of $\mathcal{M}'$ [12]. This is of particular interest where $\mathcal{M}'$ is difficult to implement efficiently, as is the case when $\mathcal{M}'$ updates large blocks of vertices.

Definition 2.27. *Let $E^*(\mathcal{M}')$ be the set of pairs (x,y) of (not necessarily distinct) states of $\mathcal{M}'$ with $P'(x,y) > 0$. For each pair $(x,y) \in E^*(\mathcal{M}')$, let $\mathcal{P}_{x,y}$ be the set of all paths from x to y using transitions of $\mathcal{M}$. Let $\mathcal{P} = \bigcup_{(x,y)\in E^*(\mathcal{M}')} \mathcal{P}_{x,y}$ be the set of all paths between all pairs of states that are adjacent in $\mathcal{M}'$. An $(\mathcal{M},\mathcal{M}')$-*flow *is a function f from $\mathcal{P}$ to the interval $[0,1]$ such that for all pairs of states $(x,y) \in E^*(\mathcal{M}')$,*

$$\sum_{\gamma\in\mathcal{P}_{x,y}} f(\gamma) = \pi'(x)P'(x,y). \tag{2.8}$$

If every path γ such that $f(\gamma) > 0$ contains an odd number of transitions, then we say that f is an odd $(\mathcal{M},\mathcal{M}')$-flow.

For each path $\gamma \in \mathcal{P}$, let $r((z,w),\gamma)$ be the number of times the transition (z,w) appears on the path γ.

For each $(z,w) \in E^*(\mathcal{M})$, the *congestion* [12, 10] of (z,w) in the flow f is

$$A_{z,w}(f) = \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} r((z,w),\gamma)|\gamma|f(\gamma).$$

The congestion of the flow f is the maximum congestion over all edges:

$$A(f) = \max_{(z,w)\in E^*(\mathcal{M})} A_{z,w}(f).$$

Note that this definition of congestion differs slightly from the one stated in Definition 2.23 and used in the canonical paths argument. The canonical paths argument is a special case of comparison in which $\mathcal{M}'$ is the trivial Markov chain in which $P'(z,w) = \pi(w)$ for all z and w. The two forms of congestion are equivalent when this substitution is made.

Theorem 2.28 ([37]). *Let $\mathcal{M}$ and $\mathcal{M}'$ be two Markov chains with the same stationary distribution π, and f be an odd $(\mathcal{M},\mathcal{M}')$-flow. Let $\tau(\mathcal{M}',\delta)$ be the mixing time $\tau(\delta)$ for the Markov chain $\mathcal{M}'$. Then for any $0<\delta<\frac{1}{2}$, the mixing time of $\mathcal{M}$ with respect to any start state x satisfies*

$$\tau_x(\mathcal{M},\varepsilon) \leq A(f)\left[\frac{\tau(\mathcal{M}',\delta)}{\log(1/2\delta)}+1\right]\log\frac{1}{\varepsilon\pi(x)}.$$

While Theorem 2.28 requires that f is an odd flow, this requirement is not necessary if the second largest eigenvalue of $\mathcal{M}$ is positive. This is the case for the lazy version of the Jerrum-Sinclair chain, and also for any other chain with a sufficiently high self-loop probability for each state.

We will use Theorem 2.28 in Chapters 4 and 5, where we demonstrate rapid mixing of Markov chains using large block updates, and deduce that simpler chains (the Jerrum-Sinclair chain for matchings and an analogous chain for sampling independent sets in claw-free graphs) also mix rapidly. In the case of the Jerrum-Sinclair chain, this comparison technique will give a tighter bound on the mixing time than that given by Theorem 2.26, for certain graphs.

2.3.7 Continuous time

While we have considered only discrete time Markov chains so far, it is sometimes easier to analyse chains that have been modified to run in continuous time.

If $\mathcal{M}$ is a discrete time chain performing single-site updates, then we can define a related continuous-time chain that makes updates according to a Poisson process with rate $\lambda = 1$. Equivalently, each site is independently updated according to a Poisson process with rate $\frac{1}{n}$, where n is the number of sites.

It is possible to analyse the Jerrum-Sinclair chain in continuous time, which avoids the need for the extra self-loop probability we introduced to ensure aperiodicity [24]. We will also use a continuous-time setting in Chapter 3 to analyse a Markov chain for sampling graph colourings.

The continuous time chain can be simulated by selecting t' from the Poisson distribution with parameter t, and simulating the underlying discrete time chain for t' steps. However, a bound on the mixing time of the continuous chain does not necessarily imply a bound on the mixing time of the underlying chain, as the discrete form of the chain may be periodic.

Chapter 3

The boundary case for coupling

When we use coupling to show that Markov chains are rapidly mixing, we aim to show that the expected distance between two copies of a Markov chain at any time $t+1$ is no greater than the distance at time t. If we can show that the distance $d(X_{t+1}, Y_{t+1}) \leq \beta d(X_t, Y_t)$, and the contraction ratio β is strictly less than 1, then we can infer a bound on the mixing time.

In this chapter, we consider the case where the contraction ratio $\beta = 1$. In such cases, the bound on the mixing time does not immediately follow. Instead, we need to show that there is a sufficiently large probability that the distance changes at each step. We begin by proving a slightly more general version of Theorem 2.15, which can be applied where the metric d can take non-integer values.

In Section 3.2, we consider the situation where the probability that the distance changes at each step is proportional to the distance at time t. We show that the bound on the mixing time in these cases is $O(n^2)$ - an improvement over the $O(n^3)$ that tends to be obtained by naïvely applying Theorem 2.15, and also the $O(n^2 \log n)$ achieved by Dyer and Greenhill using a tailored argument for their Markov chain for sampling independent sets [14]. We apply our new lemma to the Dyer-Greenhill chain, showing that its mixing time is $O(n^2)$ when $\lambda = \frac{2}{\Delta - 2}$.

Bordewich and Dyer also considered the case when $\beta = 1$. They showed that it is often sufficient to show that there is a high probability of a change in distance for pairs of adjacent states [4]. However, their result gives a constant bound on the probability of a change in distance for all pairs of states, which means the resulting bound on the mixing time tends to be $O(n^3)$. Our result differs in that we obtain a tighter bound on the mixing time, but it is more difficult to show a suitable bound on the probability of change.

In Section 3.4, we apply our new lemma to the problem of sampling 2Δ-colourings. This is not a trivial application, as there are situations where the probability of a change in distance is zero. This issue was also addressed by Bordewich and Dyer [4]. Instead of applying our lemma directly, we consider the evolution of a Markov chain over multiple steps, and show that this provides sufficient variance to obtain a bound on the mixing time.

3.1 Constant probability of change

Here we prove the second part of Theorem 2.15, in which the contraction ratio $\beta = 1$. We will in fact show a slightly more general version of the theorem, which is applicable to non-integer metrics. We will use this version of the theorem in Section 4.1.2.

Theorem 3.1. *Let (X_t, Y_t) be a coupling of a Markov chain $\mathcal{M}$ on state space Ω, and d be a metric (not necessarily integer-valued) on (X_t, Y_t), such that $\mathbb{E}(d(X_{t+1}, Y_{t+1})) \leq d(X_t, Y_t)$. Let D be the largest value that d can take. Suppose that whenever $X_t \neq Y_t$, for some α and δ,*

$$\Pr[|d(X_t, Y_t) - d(X_{t+1}, Y_{t+1})| > \delta] \geq \alpha. \tag{3.1}$$

Then the mixing time of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha\delta^2} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

Proof. From Equation (3.1), we can see that

$$\mathbb{E}[(d(X_t, Y_t) - d(X_{t+1}, Y_{t+1}))^2] \geq \alpha\delta^2,$$

and so

$$\mathbb{E}[d(X_{t+1}, Y_{t+1})^2] \geq 2d(X_t, Y_t)\,\mathbb{E}[d(X_{t+1}, Y_{t+1})] - d(X_t, Y_t)^2 + \alpha\delta^2. \tag{3.2}$$

Consider the value $\mathbb{E}[(D - d(X_{t+1}, Y_{t+1}))^2]$:

$$\mathbb{E}[(D - d(X_{t+1}, Y_{t+1}))^2] = D^2 + \mathbb{E}[d(X_{t+1}, Y_{t+1})^2] - 2D\,\mathbb{E}[d(X_{t+1}, Y_{t+1})].$$

By Equation (3.2),

$$\begin{aligned}\mathbb{E}[(D - d(X_{t+1}, Y_{t+1}))^2] &\geq D^2 - d(X_t, Y_t)^2 - 2D\,\mathbb{E}[d(X_{t+1}, Y_{t+1})] \\ &\quad + 2\,\mathbb{E}[d(X_{t+1}, Y_{t+1})]d(X_t, Y_t) + \alpha\delta^2 \\ &= D^2 - d(X_t, Y_t)^2 + 2\,\mathbb{E}[d(X_{t+1}, Y_{t+1})](d(X_t, Y_t) - D) + \alpha\delta^2.\end{aligned}$$

Now, since $D \geq d(X_t, Y_t)$ and $\mathbb{E}[d(X_{t+1}, Y_{t+1})] \leq d(X_t, Y_t)$,

$$\begin{aligned}\mathbb{E}[(D - d(X_{t+1}, Y_{t+1}))^2] &\geq D^2 - d(X_t, Y_t)^2 + 2d(X_t, Y_t)(d(X_t, Y_t) - D) + \alpha\delta^2 \\ &= D^2 + d(X_t, Y_t)^2 - 2Dd(X_t, Y_t) + \alpha\delta^2 \\ &= (D - d(X_t, Y_t))^2 + \alpha\delta^2.\end{aligned}$$

Let $T^{x,y}$ represent the first time that $X_t = Y_t$, conditioned on $X_0 = x, Y_0 = y$. Note that $T^{x,y}$ depends only on values of X_t and Y_t for $t < T^{x,y}$, and so $T^{x,y}$ is a stopping time. Define the process $Z(t)$ as follows:

$$Z(t) = (D - d(X_t, Y_t))^2 - \alpha\delta^2 \min\{t, T^{x,y}\}.$$

For any time $t < T^{x,y}$, consider the expected value of $Z(t+1)$.

$$\begin{aligned}\mathbb{E}[Z(t+1)] &= \mathbb{E}[(D - d(X_{t+1}, Y_{t+1}))^2] - (t+1)\alpha\delta^2 \\ &\geq (D - d(X_t, Y_t))^2 - t\alpha\delta^2 \\ &= Z(t).\end{aligned}$$

This same inequality holds trivially when $t \geq T$. Therefore Z is a submartingale with respect to $X_0, Y_0, \ldots, X_t, Y_t$ [33]. Note that the differences $Z(t+1) - Z(t)$ are bounded for all t. Also note that since Z is a submartingale, $-Z$ is trivially a supermartingale. By the optional stopping theorem for supermartingales [45, Theorem 10.10],

$$\begin{aligned}&\mathbb{E}[Z(T^{x,y})] \geq Z(0) \\ \Rightarrow \quad &\mathbb{E}[D^2 - T^{x,y}\alpha\delta^2] \geq (D - d(x,y))^2 && \text{since } d(X_{T^{x,y}}, Y_{T^{x,y}}) = 0 \\ \Rightarrow \quad &\mathbb{E}[T^{x,y}] \leq \frac{D^2 - (D - d(x,y))^2}{\alpha\delta^2} \\ &= \frac{d(x,y)(2D - d(x,y))}{\alpha\delta^2} \\ &\leq \frac{D^2}{\alpha\delta^2}.\end{aligned}$$

Now let $T = \frac{eD^2}{\alpha\delta^2}$. By Markov's inequality, $\Pr(T^{x,y} \geq T) \leq e^{-1}$. Suppose we run s independent trials of length T. The probability that X_t and Y_t have not coupled by time sT is at most e^{-s}. For any ε, we have $e^{-s} \leq \varepsilon$ if and only if $s \geq \log(\varepsilon^{-1})$. The number of independent trials s is an integer, and so the mixing time satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha\delta^2} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil,$$

as required. □

If we fix $\delta = 1$, then Theorem 3.1 and the case of Theorem 2.15 where $\beta = 1$ are equivalent.

Note that while Theorem 3.1 is stated in terms of requiring a minimum probability of improvement by some constant, this is for convenience in application only. We actually only require a bound on the variance $\sigma^2 = \mathrm{Var}(d(X_{t+1}, Y_{t+1}) \mid X_t, Y_t)$.

Corollary 3.2. *Let (X_t, Y_t) be a coupling of a Markov chain $\mathcal{M}$ on state space Ω, and d be a metric (not necessarily integer-valued) on (X_t, Y_t), such that $\mathbb{E}(d(X_{t+1}, Y_{t+1})) \leq d(X_t, Y_t)$. Let D be the largest value that d can take. Suppose that the variance $\sigma^2 = \mathrm{Var}(d(X_{t+1}, Y_{t+1}) \mid X_t, Y_t) \geq \alpha$. Then the mixing time of $\mathcal{M}$ satisfies*

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

Proof. Observe that $\sigma^2 = \mathbb{E}[(d(X_t, Y_t) - d(X_{t+1}, Y_{t+1}))^2]$, and so $\mathbb{E}[(d(X_t, Y_t) - d(X_{t+1}, Y_{t+1}))^2] \geq \alpha$. The result follows directly from the proof of Theorem 3.1, replacing $\alpha\delta^2$ with α throughout. □

3.2 Linear probability of change

For some couplings, it is possible to show that the probability of a change in distance is linearly proportional to the distance at time t. This is the case with the Dyer-Greenhill chain for sampling independent sets, for example. If we try to apply Theorem 3.1 directly in these circumstances, then we tend to get a bound of $O(\frac{1}{D})$ for α (by considering what happens when d takes its smallest non-zero value). The resulting bound on the mixing time is $O(n^3)$. In their analysis of their chain for sampling independent sets, Dyer and Greenhill considered the mixing time as a sum of random walks with different values of α, and obtained a bound of $O(n^2 \log n)$. In this section, we present a theorem that shows that the mixing time in these cases is $O(n^2)$. We apply this new theorem to the Dyer-Greenhill chain, in the boundary case where $\lambda = \frac{2}{\Delta - 2}$.

Theorem 3.3. *Let (X_t, Y_t) be a coupling of a Markov chain $\mathcal{M}$ on state space Ω, and d be a metric (not necessarily integer-valued) on (X_t, Y_t), such that $\mathbb{E}(d(X_{t+1}, Y_{t+1})) \leq d(X_t, Y_t)$. Let D be the largest value that d can take. Suppose that for some α and δ,*

$$\Pr[d(X_t, Y_t) - d(X_{t+1}, Y_{t+1}) > \delta] \geq \frac{\alpha d(X_t, Y_t)}{D}. \tag{3.3}$$

Then the mixing time of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{2eD^2}{\alpha\delta^2} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

Proof. In the proof of Theorem 3.1, we showed a bound on $\mathbb{E}[(D-d(X_{t+1},Y_{t+1}))^2]$, and used this bound to construct a submartingale from which we could infer a bound on the mixing time. In this proof, we use a similar approach, but instead of bounding $\mathbb{E}[(D-d(X_{t+1}Y_{t+1}))^2)]$, we bound the expectation of a different function f, evaluated at $d(X_{t+1},Y_{t+1})$. Define the function f as follows:

$$f(x) = x\log x - (1+\log D)x + D.$$

Note that f is strictly decreasing over the range $0 < x < D$, and that $f(0) = D$ and $F(D) = 0$.

Consider the distance $d(X_t,Y_t)$ at time t. Since $\mathbb{E}[d(X_{t+1},Y_{t+1})] \leq d(X_t,Y_t)$, we have

$$f(\mathbb{E}[d(X_{t+1},Y_{t+1})]) \geq f(d(X_t,Y_t)).$$

We wish to obtain a relationship between $\mathbb{E}[f(d(X_{t+1},Y_{t+1}))]$ and $f(d(X_t,Y_t))$. However, working with f directly is difficult. Instead, let $a = d(X_t,Y_t)$, and define two further functions g_1 and g_2:

$$g_1(x) = \frac{x^2}{2a} + x(\log a - \log D - 1) + D - \frac{a}{2}$$
$$g_2(x) = x(\log a - \log D) + D - a.$$

g_1 and g_2 are chosen such that $g_1(a) = g_2(a) = f(a)$, $g_1'(a) = g_2'(a) = f'(a)$, and $g_1''(a) = f''(a)$.

Observe that provided $0 \leq a \leq D$, the following conditions hold:

$$g_1(x) \leq f(x) \qquad \text{for } 0 \leq x \leq a, \text{ and} \tag{3.4}$$
$$g_2(x) \leq f(x) \qquad \text{for } 0 \leq x \leq D. \tag{3.5}$$

We now define a function $h(x)$:

$$h(x) = \begin{cases} 0 & \text{if } x \geq a \\ g_1(x) - g_2(x) & \text{if } x < a \end{cases}$$
$$= \begin{cases} 0 & \text{if } x \geq a \\ \frac{1}{2a}(x-a)^2 & \text{if } x < a. \end{cases}$$

By Equation (3.3), we see that $\mathbb{E}[h(d(X_{t+1},Y_{t+1}))] \geq \frac{\alpha\delta^2}{2D}$.

Now we can bound $\mathbb{E}[f(d(X_{t+1},Y_{t+1}))]$:

$$\begin{aligned}\mathbb{E}[f(d(X_{t+1},Y_{t+1}))] &\geq \mathbb{E}[g_2(d(X_{t+1},Y_{t+1})) + h(d(X_{t+1},Y_{t+1}))] \\ &= g_2(\mathbb{E}[d(X_{t+1},Y_{t+1})]) + \mathbb{E}[h(d(X_{t+1},Y_{t+1}))] \\ &\geq g_2(d(X_t,Y_t)) + \frac{\alpha\delta^2}{2D} \\ &= f(d(X_t,Y_t)) + \frac{\alpha\delta^2}{2D}.\end{aligned}$$

Let T be the first time that $X_t = Y_t$. T depends only on values of X_t and Y_t for $t \leq T$, and so is a stopping time. We now define the process $Z(t)$:

$$Z(t) = f(d(X_t,Y_t)) - \frac{\alpha\delta^2 \min\{t,T\}}{2D}.$$

For any time $t < T$,

$$\begin{aligned}\mathbb{E}[Z(t+1)] &= \mathbb{E}[f(d(X_{t+1},Y_{t+1}))] - \frac{\alpha\delta^2(t+1)}{2D} \\ &\geq f(d(X_t,Y_t)) - \frac{\alpha\delta^2 t}{2D} \\ &= Z(t).\end{aligned}$$

The same inequality holds trivially when $t \geq T$. Hence, $Z(t)$ is a submartingale with respect to $X_0, Y_0, \ldots, X_t, Y_t$ [33]. Note also that the values of $Z(t)$ are bounded, and so we can apply the optional stopping theorem for supermartingales [45, Theorem 10.10] (noting again that if Z is a submartingale then $-Z$ is a supermartingale). Since the maximum possible value of $d(X_0,Y_0)$ is D, we know that $Z(0) \geq 0$. Therefore

$$\begin{aligned}& & \mathbb{E}[Z(T)] &\geq Z(0) \\ \Rightarrow \quad & & \mathbb{E}\left[f(d(X_T,Y_T)) - \frac{\alpha\delta^2 T}{2D}\right] &\geq 0 \\ \Rightarrow \quad & & \mathbb{E}[f(d(X_T,Y_T))] &\geq \frac{\alpha\delta^2\,\mathbb{E}[T]}{2D} \\ \Rightarrow \quad & & \mathbb{E}[T] &\leq \frac{2D^2}{\alpha\delta^2}.\end{aligned}$$

We can now proceed as in the proof of Theorem 3.1. Let $T' = \left\lceil \frac{2eD^2}{\alpha\delta^2} \right\rceil$. By Markov's inequality, the probability that X_t and Y_t have not coupled by time T' is at most e^{-1}. If we run s independent trials of length T', then the probability that X_t and Y_t have not coupled by time sT' is at most e^{-s}. Now let $t = sT'$. $e^{-s} \leq \varepsilon$ if and only if $s \geq \log(\varepsilon^{-1})$. Since s is an integer, we need to take

$$t \geq T' \left\lceil \log(\varepsilon^{-1}) \right\rceil$$

to ensure that the chains couple with probability at least $1-\varepsilon$. Thus, the mixing time satisfies:

$$\tau(\varepsilon) \leq \left\lceil \frac{2eD^2}{\alpha\delta^2} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

□

3.3 The boundary case for the Dyer-Greenhill chain

The Dyer-Greenhill chain is a Markov chain for sampling independent sets weighted according to a parameter λ, distributed according to Equation (2.3). Luby and Vigoda defined a Markov chain for sampling independent sets, and used coupling to show that it is rapidly mixing when $\lambda \leq \frac{1}{\Delta-3}$, provided $\Delta \geq 4$. Dyer and Greenhill used path coupling to show an improved upper bound on the mixing time of the Luby-Vigoda chain, and also defined another Markov chain for sampling independent sets. They showed that their chain is rapidly mixing when $\lambda \leq \frac{2}{\Delta-2}$ using path coupling [14]. In the case $\lambda = \frac{2}{\Delta-2}$, they used a tailored argument to show that the mixing time of the Dyer-Greenhill chain is $O(n^2 \log n)$. In this section, we will apply Theorem 3.3 to obtain an improved mixing time of $O(n^2)$ in the case $\lambda = \frac{2}{\Delta-2}$.

Definition 3.4. *The transitions of the Dyer-Greenhill chain for sampling independent sets, $\mathcal{M}_{DG}$, are defined as follows: Let X_t represent the state of $\mathcal{M}_{DG}$ at time t. The subsequent state of the chain, X_{t+1}, is determined by the following sequence of steps.*

(DG1) Select a vertex $v \in V(G)$ uniformly at random.

(DG2) If $v \in X_t$, then let $X_{t+1} = X_t \setminus \{v\}$ with probability $\frac{1}{1+\lambda}$.

(DG3) If $v \notin X_t$ and there is no $v' \in X_t$ adjacent to v, then let $X_{t+1} = X_t \cup \{v\}$ with probability $\frac{\lambda}{1+\lambda}$.

(DG4) If $v \notin X_t$ and there is exactly one $v' \in X_t$ adjacent to v, then let $X_{t+1} = (X_t \cup \{v\}) \setminus \{v'\}$ with probability $\frac{\lambda}{4(1+\lambda)}$. This is referred to as a drag *move.*

(DG5) In all other cases, $X_{t+1} = X_t$.

The probability of the drag move of the Dyer-Greenhill chain is chosen to maximise the value of λ for which the chain is rapidly mixing.

Dyer and Greenhill showed that when $\lambda = \frac{2}{\Delta-2}$, the mixing time satisfies

$$\tau(\varepsilon) \leq \left\lceil 2n^2 e(1+\lambda)(\log(n)+1) \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

We can use Theorem 3.3 to obtain a better bound on the mixing time. The analysis of the Dyer-Greenhill chain uses Hamming distance as a metric, and the coupling always updates the same vertex in each copy of the chain. Consider a vertex v that belongs to $X_t \oplus Y_t$. With probability $\frac{1}{n(1+\lambda)}$, v is deleted from one independent set, decreasing the Hamming distance. Therefore the probability that the Hamming distance changes is at least $\frac{d(X_t,Y_t)}{n(1+\lambda)}$. Applying Theorem 3.3 with $D = n$, $\delta = 1$ and $\alpha = \frac{1}{1+\lambda}$, we obtain the improved mixing time

$$\tau(\varepsilon) \leq \lceil 2en^2(1+\lambda) \rceil \lceil \log(\varepsilon^{-1}) \rceil .$$

In Chapter 4, we will use the Dyer-Greenhill chain to show that we can sample weighted matchings of a graph G, with restricted λ, by considering independent sets of the line graph of G.

3.4 Sampling 2Δ-colourings

Given a graph $G = (V,E)$ and a set of colours Q, a proper colouring is an assignment of colours to the vertices of G, such that no two adjacent vertices are assigned the same colour. We can define a Markov chain $\mathcal{M}_C$ for sampling proper colourings using local heat-bath updates.

Definition 3.5. *$\mathcal{M}_C$ is the heat-bath chain for sampling proper colourings of a graph $G = (V,E)$. Given the state X_t at time t, we determine the state X_{t+1} at time $t+1$ as follows:*

1. *Select a vertex $v \in V$ uniformly at random.*
2. *Select a random colour q that is not present in the neighbourhood of v, and recolour v with colour q.*

A coupling argument using Hamming distance as a metric [23] shows that this chain is rapidly mixing provided the maximum degree of G is bounded, and $|Q| > 2\Delta$. Jerrum notes an observation by Frieze, that the argument can also be applied where $|Q| = 2\Delta$. However, when $|Q| = 2\Delta$, the bound on the mixing time worsens by a factor of about n^2 [24].

Vigoda showed that this chain has a mixing time of $O(n^2 \log n)$ where $|Q| > \frac{11\Delta}{6}$, by using a comparison argument with a different chain [43]. Bounds on the mixing

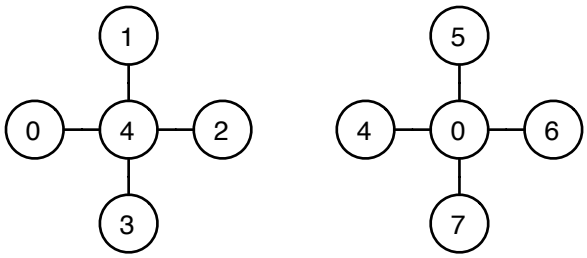

Figure 3.1: *A pair of configurations such that no update to the central vertex can result in a change in distance.*

time exist for smaller numbers of colours, but they impose additional restrictions on the girth of G [17].

In this section we will show that the mixing time of $\mathcal{M}_C$ is quadratic when the number of colours $|Q| = 2\Delta$.

Theorem 3.6. *The Markov chain $\mathcal{M}_C$ for sampling colourings mixes in time $O(n^2)$ in the case where $|Q| = 2\Delta$.*

The main difficulty we face in showing that $\mathcal{M}_C$ mixes rapidly is in bounding the probability that the distance changes in any step. Consider the situation in Figure 3.1. This shows a pair of configurations of a small section of a 2-dimensional lattice graph with $\Delta = 4$ and $|Q| = 8$. We can easily see that there is no single colour q that can be assigned to the central vertex in both copies of the graph. If we construct a coupling that updates the same vertex in each copy of the graph and uses Hamming distance as a metric, we cannot show any improvement in distance when this central vertex is selected.

Furthermore, it is possible to construct a situation in which *every* vertex behaves in this way. Figure 3.2 shows a pair of 8-colourings of a toroidal graph, in which no single site update can result in a change in Hamming distance. We will overcome this difficulty by considering a continuous-time version of $\mathcal{M}_C$, and defining a new chain, each step of which is equivalent to running the continuous chain for time $n \log 2$ (this is the time required for each vertex to be updated at least once with probability $\frac{1}{2}$). First, however, it is useful to prove a variation of Theorem 3.3. Where Theorem 3.3 requires a bound on the probability that the distance between the two copies of our chain changes, Lemma 3.7 requires a bound on the expected absolute change in distance.

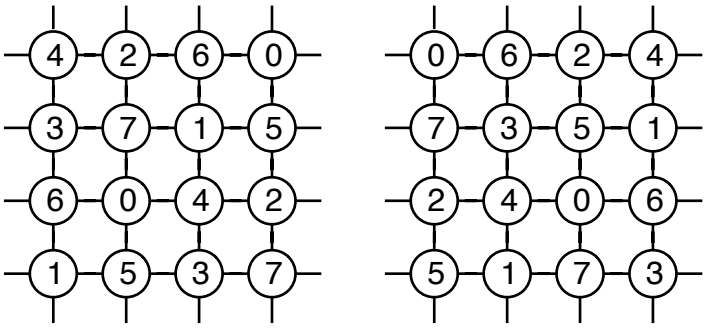

Figure 3.2: *A pair of configurations of a 2-dimensional toroidal graph, where the probability of a change in distance is zero.*

Lemma 3.7. *Let (X_t, Y_t) be a coupling of a Markov chain $\mathcal{M}$ on state space Ω, and d be a metric on (X_t, Y_t), such that $\mathbb{E}(d(X_{t+1}, Y_{t+1})) \leq d(X_t, Y_t)$. Let D be the largest value that d can take. Suppose that for some α,*

$$(\mathbb{E}[|d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)|])^2 \geq \frac{\alpha d(X_t, Y_t)}{D}. \tag{3.6}$$

Then the mixing time of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{8eD^2}{\alpha} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil.$$

Proof. As in the proof of Theorem 3.3, we will attempt to obtain a bound on $\mathbb{E}[f(d(X_{t+1}, Y_{t+1}))]$ in terms of $f(d(X_t, Y_t))$, where f is the function

$$f(x) = x \log x - (1 + \log D)x + D.$$

Let $a = d(X_t, Y_t)$, and let $X = d(X_{t+1}, Y_{t+1})$ be a random variable implicitly conditioned on X_t and Y_t. Define the functions g_1 and g_2 as follows:

$$g_1(x) = \frac{x^2}{2a} + x(\log a - \log D - 1) + d - \frac{a}{2}$$
$$g_2(x) = x(\log a - \log D) + D - a.$$

These are the same functions that we used in Theorem 3.3 to bound f, and Conditions (3.4) and (3.5) hold. Define the function $h(x)$ as follows:

$$h(x) = \begin{cases} 0 & \text{if } x \geq a \\ g_1(x) - g_2(x) & \text{if } x < a \end{cases}$$
$$= \begin{cases} 0 & \text{if } x \geq a \\ \frac{1}{2a}(x-a)^2 & \text{if } x < a. \end{cases}$$

Now we need to obtain a bound on $h(X)$. Let $X^- = \min\{X, a\}$ and $X^+ = \max\{X, a\}$. Then

$$|X - a| = X^+ - X^- = X - 2X^- + a.$$

Also observe that $h(X) = \frac{(a-X^-)^2}{2a}$. Therefore

$$\begin{aligned}
\mathbb{E}[|X-a|] &= \mathbb{E}[X] - 2\,\mathbb{E}[X^-] + a \\
&\leq 2(a - \mathbb{E}[X^-]) && \text{(since } \mathbb{E}[X] \leq a\text{)} \\
&= 2\,\mathbb{E}[a - X^-].
\end{aligned}$$

Hence

$$(\mathbb{E}[|X-a|])^2 \leq 4(\mathbb{E}[a-X^-])^2 \leq 4\,\mathbb{E}[(a-X^-)^2] = 8a\,\mathbb{E}[h(X)].$$

Substituting in the bound on $(\mathbb{E}[|X-a|])^2$ from Condition 3.6, we can obtain a bound on $\mathbb{E}[h(X)]$:

$$\mathbb{E}[h(X)] \geq \frac{\alpha}{8D}.$$

We can now bound $\mathbb{E}[f(d(X_{t+1}, Y_{t+1}))]$:

$$\begin{aligned}
\mathbb{E}[f(d(X_{t+1}, Y_{t+1}))] &\geq \mathbb{E}[g_2(d(X_{t+1}, Y_{t+1})) + h(d(X_{t+1}, Y_{t+1}))] \\
&= g_2(\mathbb{E}[d(X_{t+1}, Y_{t+1})]) + \mathbb{E}[h(d(X_{t+1}, Y_{t+1}))] \\
&\geq g_2(d(X_t, Y_t)) + \frac{\alpha}{8D} \\
&= f(d(X_t, Y_t)) + \frac{\alpha}{8D}.
\end{aligned}$$

Now let T be the first time that $X_t = Y_t$. T depends only on values of X_t and Y_t for $t \leq T$, and so is a stopping time. Define the process $Z(t)$:

$$Z(t) = f(d(X_t, Y_t)) - \frac{\alpha \min\{t, T\}}{8D}.$$

For any time $t < T$,

$$\begin{aligned}
\mathbb{E}[Z(t+1)] &= \mathbb{E}[f(d(X_{t+1}, Y_{t+1}))] - \frac{\alpha(t+1)}{8D} \\
&\geq f(d(X_t, Y_t)) - \frac{\alpha t}{8D} \\
&= Z(t).
\end{aligned}$$

This inequality also holds trivially when $t \geq T$. Therefore Z_t is a submartingale with respect to $X_0, Y_0, \ldots, X_t, Y_t$. The values of $Z(t)$ are bounded, and so we can apply the

optional stopping theorem for supermartingales to $-Z$ [45]. We know that $d(X_0, Y_0) \leq D$, and so $Z(0) \geq 0$. Therefore

$$\begin{aligned} & \mathbb{E}[Z(T)] \geq Z(0) \\ \Rightarrow \quad & \mathbb{E}\left[f(d(X_T, Y_T)) - \frac{\alpha T}{8D}\right] \geq 0 \\ \Rightarrow \quad & \mathbb{E}[f(d(X_T, Y_T))] \geq \frac{\alpha \mathbb{E}[T]}{8D} \\ \Rightarrow \quad & \mathbb{E}[T] \leq \frac{8D^2}{\alpha}. \end{aligned}$$

Now let $T' = \left\lceil \frac{8eD^2}{\alpha} \right\rceil$. By Markov's inequality, the probability that X_t and Y_t have not coupled by time T' is at most e^{-1}. If we run s independent trials of length T', then the probability that X_t and Y_t have not coupled by time sT' is at most e^{-s}. Now let $t = sT'$. $e^{-s} \leq \varepsilon$ if and only if $s \geq \log(\varepsilon^{-1})$. Since s is an integer, we require

$$t \geq T' \left\lceil \log(\varepsilon^{-1}) \right\rceil$$

to ensure that the chains couple with probability at least $1 - \varepsilon$. Therefore the mixing time satisfies

$$\tau(\varepsilon) \leq \left\lceil \frac{8eD^2}{\alpha} \right\rceil \left\lceil \log(\varepsilon^{-1}) \right\rceil,$$

as required. □

In the proof of Theorem 3.6, we will require upper and lower bound tail inequalities in order to bound the probability that the value of a random variable is at least some distance from the mean. For the lower bound, it is not sufficient to consider just the variance of our random variable for this - we also need a bound on the fourth moment. The following lemma builds on a result shown by Petrov [36], and gives a lower bound tail inequality where we have a sum of random variables with bounded second and fourth central moments.

Lemma 3.8. *Let $Y_1, \ldots, Y_n$ be a sequence of independent random variables such that for all i, and for some constants c_1, c_2, c_3,*

$$\begin{aligned} \mathbb{E}[Y_i] &= \mu_i \\ c_1 \leq \mathbb{E}[(Y_i - \mu_i)^2] &\leq c_2 \\ \mathbb{E}[(Y_i - \mu_i)^4] &\leq c_3, \end{aligned}$$

Let $Y = \sum_{i=1}^n Y_i$. Then $\mathbb{E}[Y] = \mu = \sum_{i=1}^n \mu_i$, and for any $0 \leq b \leq \sqrt{nc_1}$,

$$\Pr(|Y - \mathbb{E}[Y]| > b) \geq \frac{(nc_1 - b^2)^2}{nc_3 + 3n(n-1)c_2^2}.$$

Proof. We will use the following result by Petrov [36]: If X is a random variable and the fourth central moment of X is finite, then for any $0 \leq b \leq \sqrt{\mathrm{Var}(X)}$,

$$\Pr(|X - \mathbb{E}[X]| > b) \geq \frac{(\mathrm{Var}(X) - b^2)^2}{\mathbb{E}[X - \mathbb{E}[X]]^4}. \tag{3.7}$$

First, let $X_i = Y_i - \mu_i$ for all i, and let $X = \sum_{i=1}^n X_i$. This ensures that $\mathbb{E}[X] = \mathbb{E}[X_i] = 0$, and will simplify the proof somewhat. Since the Y_i variables are independent, we know that $\mathbb{E}[Y] = \mathbb{E}[X] + \sum_{i=1}^n \mu_i$, as required.

To apply Equation (3.7), we need to find a lower bound on the variance of X, and an upper bound on the fourth central moment of X. By standard properties of variance for independent random variables, we have

$$nc_1 \leq \mathrm{Var}(X) \leq nc_2.$$

For the fourth central moment, we have

$$\begin{aligned}
\mathbb{E}[X^4] &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] \\
&= \mathbb{E}\left[\sum_{i,j,k,l=1}^n X_i X_j X_k X_l\right] \\
&= \sum_{i,j,k,l=1}^n \mathbb{E}[X_i X_j X_k X_l]
\end{aligned}$$

Since the X_i variables are independent, each term in this expression is either of the form $\mathbb{E}[X_i^4]$ for some i, of the form $\mathbb{E}[X_i^2]\mathbb{E}[X_j^2]$ for distinct i and j, or zero, because $\mathbb{E}[X_i] = 0$ for all i. For each distinct i and j, the term $\mathbb{E}[X_i^2]\mathbb{E}[X_j^2]$ appears six times in the expansion of $\mathbb{E}[X^4]$. Let $I(i \neq j)$ be the indicator function of the event $i \neq j$. We therefore have

$$\begin{aligned}
\mathbb{E}[X^4] &= \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \sum_{i,j=1}^n I(i \neq j)\, \mathbb{E}[X_i^2]\, \mathbb{E}[X_j^2] \\
&\leq nc_3 + 3n(n-1)c_2^2.
\end{aligned}$$

We can now use these bounds on the variance and fourth moment of X to bound the probability that $|X|$ is large. For $0 \leq b \leq \sqrt{nc_1}$,

$$\Pr[|X| > b] \geq \frac{(nc_1 - b^2)^2}{nc_3 + 3n(n-1)c_2^2}$$

Substituting $Y - \mathbb{E}[Y]$ for X gives the stated result. □

The use of the fourth central moment in the proof of Lemma 3.8 is justified by the fact that a small fourth central moment indicates that much of the variance is caused by frequent small deviations, instead of infrequent large deviations. It is not possible to obtain a lower bound tail inequality by considering variance alone.

There are a number of upper bound tail inequalities. However, the most general (Markov's inequality and Chebyshev's inequality) do not provide useful values for distances less than one standard deviation from the mean. In order to obtain both upper and lower bounds with non-trivial values, we need a stronger upper bound based on additional assumptions on our random variable. Hoeffding's inequality applies where we have a sum of bounded random variables.

Lemma 3.9 (Hoeffding's inequality [22, 32])**.** *Let $Y_1, \ldots, Y_n$ be independent random variables such that, for all i, there are constants a_i and b_i such that $a_i \leq Y_i \leq b_i$. Let $Y = \sum_{i=1}^{n} Y_i$. Then for all positive t,*

$$\Pr(Y - \mathbb{E}[Y] \geq nt) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \qquad \textit{and}$$
$$\Pr(Y - \mathbb{E}[Y] \leq -nt) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Therefore

$$\Pr(|Y - \mathbb{E}[Y]| \geq nt) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Finally, where we have a random variable that is a sum of Poisson trials, the following Chernoff bound allows us to find a lower bound on the probability that the random variable is close to the mean.

Lemma 3.10 ([33])**.** *Let $Y_1, \ldots, Y_n$ be independent Poisson trials such that $\Pr(Y_i) = p_i$. Let $Y = \sum_{i=1}^{n} Y_i$, and $\mu = \mathbb{E}[Y]$. Then, for $0 < \delta < 1$,*

$$\Pr(Y \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}.$$

We now define the continuous-time version of the Markov chain, $\mathcal{M}'_C$, and the upper-level chain $\mathcal{M}^*_C$. The transitions of $\mathcal{M}'_C$ are the same as those of $\mathcal{M}_C$, but occur at times determined by a Poisson process with parameter $\lambda = 1$. Each step of $\mathcal{M}^*_C$ is equivalent to running $\mathcal{M}'_C$ for time $n\log 2$. This means that for any vertex v, the probability that v is updated at least once in a single transition of $\mathcal{M}^*_C$ is $\frac{1}{2}$, and is independent of any other updates.

We can now continue with the proof of Theorem 3.6.

Proof of Theorem 3.6. Consider a coupling on a pair of colourings, evolving according to $\mathcal{M}_C^*$. We will focus on a small set of edges K, and show that even if the vertices of these edges are blocked, there is a sufficient probability of a change in distance over one step of $\mathcal{M}_C^*$.

Let D_0 be the set of vertices that disagree in the two colourings (so the Hamming distance between the colourings is $|D_0|$), and select a set of edges $K = \{(u_i, v_i) : 1 \leq i \leq |K|\}$, satisfying the following properties:

1. For each edge $(u_i, v_i) \in K$, $v_i \in D_0$.
2. The edges in K do not share any endpoints
3. Let $V_0 = \bigcup_{(u_i,v_i)\in K}\{u_i, v_i\}$. Then for any pair of vertices $w, w' \in V_0$, either $(w, w') \in K$, or the shortest path between w and w' in G has length at least 3.

Later in this proof, we will define a set $M \subseteq K$ that depends on the set of vertices that are updated during a transition of $\mathcal{M}_C^*$. These properties of K will ensure that the events that each edge $(u_i, v_i) \in K$ also belongs to M are independent.

We now show that we can find a set K such that $|K|$ is proportional to $|D_0|$.

Assume that G is connected, and consider the number of connected components after selecting each edge $(u_i, v_i) \in K$, and removing all vertices within the ball of radius 2 around u_i and v_i from V. After selecting the first edge, there are at most $2(\Delta-1)^3 - 1$ extra connected components. After selecting ℓ edges, we have removed at most $2\ell(\Delta^2 - \Delta + 1)$ vertices, and have at most $1 + 2\ell((\Delta-1)^3) - \frac{1}{2})$ components. Therefore, if there are at least $2 + 2\ell(\Delta^2 - \Delta + 1) + 2\ell((\Delta-1)^3 - \frac{1}{2})$ vertices in D_0, then by the pigeonhole principle there is at least one component with two or more D_0 vertices, and we can select another edge.

We can therefore always find a set K with $|K| \geq \left\lceil \frac{|D_0|-2}{2\Delta^3-4\Delta^2+4\Delta-1} \right\rceil$.

We will now consider the behaviour of $\mathcal{M}_C^*$. In a single step, $\mathcal{M}_C^*$ will update a sequence of vertices selected uniformly at random, and of some length determined by a Poisson distribution. We begin by fixing the sequence of vertices that are updated, and then consider the effect of different assignments of colours to these vertices. In doing so, we obtain a bound on the probability of a change in distance, conditioned on that sequence of vertices. If we can obtain a lower bound on this probability for some sequences, and a lower bound on the probability of selecting such a "good" sequence, then we will be able to bound the mixing time of $\mathcal{M}_C^*$.

Having fixed the sequence of vertices to be updated by $\mathcal{M}_C^*$, we consider the probability of a change in distance, conditioned on that sequence of updates. If we hit both

endpoints of an edge $(u_i, v_i) \in K$ during this process, and u_i is hit at least once before v_i is hit for the last time, then there is a constant probability of unblocking v_i and reducing the distance. We will show that this implies a bound on $(\mathbb{E}[|d(X_{t+1,Y_{t+1}}) - d(X_t, Y_t)|])^2$ for $\mathcal{M}_C^*$, and use Lemma 3.7 to bound the mixing time. The bound on the mixing time of the continuous time chain $\mathcal{M}_C'$ follows from this.

Given the sequence of updates, let $M \subseteq K$ be the set of edges satisfying the following:

1. u_i and v_i are updated at least once each.
2. The last time that u_i is updated occurs before the last time that v_i is updated.
3. No other neighbours of u_i and v_i are updated during this transition.

For each edge $(u_i, v_i) \in K$, there is a constant non-zero probability that these conditions are met, independent of whether they are met for any other edge in K. This independence is a consequence of the minimum distance imposed between edges in K. If this happens, then the updates to u_i and v_i lead to a change in distance with non-zero probability.

Let ω' represent the following information:

1. The sequence of all the sites that are updated during the transition.
2. The choices of colours for those vertices that are updated and are *not* endpoints of the edges in K.

Let ω represent the choices of colours for the vertices that *are* endpoints of the edges in K. Let $X(\omega', \omega)$ be the change in distance between the two colourings in the coupling, when the sequence of updates defined by ω' and ω is made.

We wish to show that $\mathbb{E}[|X|]$ is large, so that we can apply Lemma 3.7 and therefore bound the mixing time of $\mathcal{M}_C^*$.

Define a random variable Y as follows:

$$Y(\omega') = \mathbb{E}[X(\omega', \cdot)].$$

Y represents the expected change in distance for a given choice of ω'. We now define a second variable U as follows:

$$U(\omega', \omega) = X(\omega', \omega) - Y(\omega').$$

U is the deviation from Y for a given choice of ω, and we have the property that $X(\omega', \omega) = Y(\omega') + U(\omega', \omega)$. We wish to show that $\mathbb{E}[|X|]$ is sufficiently large. While we do not know anything about the distribution of Y, we can consider the behaviour of U conditioned on specific values of Y.

If, for a specific ω', $|Y(\omega')|$ is large, then we wish to show that $\mathbb{E}[|X|]$ is also large. We can demonstrate this by showing that $\mathbb{E}[|U(\omega', \cdot)|]$ is small. Conversely, if $|Y(\omega')|$ is small, then we need to show that $\mathbb{E}[|U(\omega', \cdot)|]$ is large in order to show that $\mathbb{E}[|X|]$ is also large. We can show this by considering upper and lower bound tail inequalities for U.

Note first that due to the choice of K and the conditions for inclusion in M, $U(\omega', \cdot)$ is a sum of independent random variables, as required by Lemma 3.8, and is independent of updates outside of M (once M has been established, we know that no neighbours of edges in M are hit). Likewise, the choice of colours for vertices outside of M is independent of ω, because no neighbours of the edges in M are hit, and so the updates to M have no influence on the vertices outside of M. Also, for all ω', we have $\mathbb{E}[U(\omega', \cdot)] = 0$.

We first consider the case where $|Y| \geq \frac{2c}{3}\sqrt{|M|}$, where c is a suitable constant (we will give an upper bound on c later). We are interested in the cases where $|U| \leq \frac{c}{3}\sqrt{|M|}$, because in these cases we are guaranteed that $|X|$ is reasonably large. For each edge e_i in M, let U_i be the random variable representing the change in Hamming distance of the endpoints of e_i after the transition of $\mathcal{M}_C^*$. Note that each U_i can take only a limited range of discrete values (from -2 to 1). U is the sum of the U_is. To obtain a good upper bound on $\Pr(|U| \geq \frac{c}{3}\sqrt{|M|})$, we apply Lemma 3.9:

$$\Pr(U - \mathbb{E}[U] \geq nt) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

$$\Rightarrow \qquad \Pr(|U| \geq \frac{c}{3}\sqrt{|M|}) \leq 2e^{-\frac{2c^2}{81}}.$$

Therefore, when $|Y| \geq \frac{2c}{3}\sqrt{|M|}$,

$$\Pr\left(|X| \geq \frac{c}{3}\sqrt{|M|}\right) \geq 1 - 2e^{-\frac{2c^2}{81}}.$$

Now we consider the case where $|Y| \leq \frac{2c}{3}\sqrt{|M|}$. Note that since each U_i can only take a constant number of discrete values, it must therefore have finite moments. Furthermore, we can enumerate all the possible configurations of e_i and its neighbouring vertices, and therefore find constant upper and lower bounds (for a fixed Δ) on the variance and fourth moment of U_i. Let c_1, c_2, c_3 be constants such that, for all i,

$c_1 \leq \mathbb{E}[(U_i - \mathbb{E}[U_i])^2] \leq c_2$, $\mathbb{E}[(U_i - \mathbb{E}[U_i])^4] \leq c_3$, and $c_3 \geq 3c_2^2$. We can now provide bounds on the constant c: in order to obtain useful results from both Lemma 3.9 and Lemma 3.8, we require that $0 < c < \sqrt{c_1}$.

We wish to find a lower bound on the probability that $|U|$ is at least $c\sqrt{|M|}$. By Lemma 3.8,

$$\Pr(|U - \mathbb{E}[U]| > b) \geq \frac{(nc_1 - b^2)^2}{nc_3 + 3n(n-1)c_2^2}$$

$$\Rightarrow \qquad \Pr(|U| \geq c\sqrt{|M|}) \geq \frac{(c_1 - c^2)^2|M|}{c_3 + 3(|M|-1)c_2^2}.$$

Therefore, when $|Y| \leq \frac{2c}{3}\sqrt{|M|}$,

$$\Pr(|X| \geq \frac{c}{3}\sqrt{|M|}) \geq \frac{(c_1 - c^2)^2|M|}{c_3 + 3(|M|-1)c_2^2}.$$

This probability is smallest when $|M| = 1$ (using the condition that $c_3 \geq 3c_2^2$), and so, regardless of $|M|$,

$$\Pr(|X| \geq \frac{c}{3}\sqrt{|M|}) \geq \frac{(c_1 - c^2)^2}{c_3}.$$

Finally, in order to obtain a bound on $\mathbb{E}[|X|]$ from these bounds on the probability that $|X|$ is larger than some multiple of $|M|$, we need a bound on the probability that $|M|$ is large. Let $\mathcal{E}$ be the event that $|M| \geq c'|K|$, where c' is a positive constant such that $\mathbb{E}[|M|] > c'|K|$. Since the event that each edge in K belongs to M is independent, $|M|$ is a random variable given by a sum of Poisson trials, and so we can apply Lemma 3.10 to find a lower bound on $\Pr(\mathcal{E})$. The requirement that $\mathbb{E}[|M|] > c'|K|$ ensures that this lower bound is non-zero.

We can now combine these results to obtain the following bound on $\mathbb{E}[|X|]$:

$$\mathbb{E}[|X|] \geq \Pr(\mathcal{E})\frac{c}{3}\sqrt{c'|K|}\min\left\{1 - 2e^{\frac{-2c^2}{81}}, \frac{(c_1 - c^2)^2}{c_3}\right\}.$$

We will now show that the bound on $\mathbb{E}[|X|]$ is proportional to $\sqrt{D_0}$. It will follow that $(\mathbb{E}[|d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)|])^2$ is linear in $|D_0|$. Recall the lower bound on $|K|$ that we obtained earlier:

$$|K| \geq \left\lceil \frac{|D_0| - 2}{2\Delta^3 - 4\Delta^2 + 4\Delta - 1} \right\rceil.$$

Also observe that when $|D_0| \geq 4$, we have the bound

$$|K| \geq \frac{D_0}{4\Delta^3 - 8\Delta^2 + 4\Delta - 2}.$$

We can combine this with our bound on $\mathbb{E}[|X|]$ to obtain the desired bound on $(\mathbb{E}[|d(X_{t+1},Y_{t+1})-d(X_t,Y_t)|])^2$. Let $n=|V|$ be the number of vertices of the graph (and thus the maximum possible value of $|D_0|$). Subject to the assumption that $|D_0|\geq 4$, we have

$$(\mathbb{E}[|d(X_{t+1},Y_{t+1})-d(X_t,Y_t)|])^2 \geq \frac{n\Pr(\mathcal{E})^2c^2c'\min\{1-2e^{\frac{-2c^2}{81}},\frac{(c_1-c^2)^2}{c_3}\}^2}{36\Delta^3-72\Delta^2+72\Delta-18}\frac{|D_0|}{n}. \tag{3.8}$$

When $|D_0|<4$, none of the vertices of D_0 can be blocked (provided $\Delta\geq 3$), and so this bound still holds.

We now use the property that the expected distance does not increase when $|Q|=2\Delta$ to apply Lemma 3.7. Set $\alpha=\frac{n\Pr(\mathcal{E})^2c^2c'\min\{1-2e^{\frac{-2c^2}{81}},\frac{(c_1-c^2)^2}{c_3}\}^2}{36\Delta^3-72\Delta^2+72\Delta-18}$. From Lemma 3.7, we obtain the bound on the mixing time for $\mathcal{M}_C^*$:

$$\tau(\varepsilon)\leq\left\lceil\frac{8en(36\Delta^3-72\Delta^2+72\Delta-18)}{\Pr(\mathcal{E})^2c^2c'\min\{1-2e^{\frac{-2c^2}{81}},\frac{(c_1-c^2)^2}{c_3}\}}\right\rceil\left\lceil\log(\varepsilon^{-1})\right\rceil.$$

The mixing time of the continuous chain $\mathcal{M}_C'$ therefore satisfies

$$\tau(\varepsilon)\leq n\log 2\left\lceil\frac{8en(36\Delta^3-72\Delta^2+72\Delta-18)}{\Pr(\mathcal{E})^2c^2c'\min\{1-2e^{\frac{-2c^2}{81}},\frac{(c_1-c^2)^2}{c_3}\}}\right\rceil\left\lceil\log(\varepsilon^{-1})\right\rceil.$$

This is $O(n^2)$ provided c_1, c_3, c and c' are constant, and $\Pr(\mathcal{E})$ has a non-zero constant lower bound. We will now show that this is the case.

Each U_i is a discrete random variable taking values in the range $[-2,1]$, and therefore each U_i has finite moments. The distribution of each U_i depends on the number and configuration of the vertices adjacent to the edge (u_i,v_i), and by the definition of M, the configuration of these edges does not change during a transition of $\mathcal{M}_C^*$. The number of possible configurations of u_i, v_i and their neighbouring edges is a function only of Δ (and $|Q|$, which is 2Δ in this case) and does not depend on n. Therefore, it is possible to enumerate every such configuration, and find the minimum and maximum possible values of the second and fourth moments of each U_i. Therefore, the values c_1 and c_2 depend only on Δ, and are constant with respect to n. c_3 depends only on Δ and c_2, and therefore is also constant.

c is a value such that $0<c<\sqrt{c_1}$. We are free to choose any value of c within this range, and since c_1 is a constant, we can choose a constant value for c.

For each edge e in K, there is a constant non-zero probability that e belongs to M, and these probabilities are independent due to the way K is selected. In the worst case,

e has exactly $2\Delta - 2$ neighbouring vertices, and the probability that e belongs to M is $\frac{1}{2^{\Delta+1}}$. The expected size of M is therefore at least $\frac{|K|}{2^{2\Delta+1}}$. We require c' to be chosen such that $0 < c' < 1$ and $\mathbb{E}[|M|] > c'|K|$. This is satisfied if $0 < c' < \frac{1}{2^{2\Delta+1}}$. This is independent of n, and so we can choose a constant value for c'.

Finally, $\Pr(\mathcal{E})$ is the probability that $|M| \geq c'|K|$. For convenience, we will actually bound the probability of the event that $|M| > c'|K|$, which will give a lower bound on $\Pr(\mathcal{E})$. Call this event $\mathcal{E}'$. As we have established, the probability that each edge in K belongs to M is constant and at least $\frac{1}{2^{\Delta+1}}$, and these events are independent. Therefore $|M|$ is a sum of independent Poisson trials. We also know that $\mathbb{E}[|M|] \geq \frac{|K|}{2^{2\Delta+1}}$. Therefore

$$\begin{aligned}
\Pr(\mathcal{E}') &= 1 - \Pr(|M| \leq c'|K|) \\
&\geq 1 - \Pr(|M| \leq c' 2^{2\Delta+1}\, \mathbb{E}[|M|] \\
&= 1 - \Pr(|M| \leq (1 - (1 - c' 2^{2\Delta+1}))\, \mathbb{E}[|M|]).
\end{aligned}$$

By the choice of c', we know that $0 < 1 - c'2^{2\Delta+1} < 1$, and so we can apply Lemma 3.10 to obtain a lower bound on $\Pr(\mathcal{E}')$:

$$\begin{aligned}
\Pr(\mathcal{E}') &\geq 1 - e^{-\frac{\mathbb{E}|M|(1-c'2^{2\Delta+1})^2}{2}} \\
&\geq 1 - e^{-\frac{|K|(1-c'2^{2\Delta+1})^2}{2^{2\Delta+2}}}.
\end{aligned}$$

This probability is smallest when $|K| = 1$, and so we obtain a constant lower bound on $\Pr(\mathcal{E})$ of

$$\Pr(\mathcal{E}) \geq 1 - e^{-\frac{(1-c'2^{2\Delta+1})^2}{2^{2\Delta+2}}}.$$

Therefore the mixing time of $\mathcal{M}_C$ is $O(n^2)$, as required. □

While we have focused here specifically on graph colourings, the argument used to prove Theorem 3.6 could be applied to other spin systems, provided the following conditions are met:

1. Moves are local, that is, all edges updated in the coupling are within a ball of constant radius. This is clearly the case for $\mathcal{M}_C$, since updates are single-site and the coupling always updates the same vertex in each copy of the chain.

2. The metric is local: for a given move, the change in distance is independent of the state of any vertex more than a constant distance from the updated vertices. This always applies when Hamming distance is used as a metric.

3. In any situation where a vertex v disagrees in the two copies of the chain, and there is no single move that can improve the distance, there is a constant length sequence of moves (an "unlocking sequence") that results in an improvement. For $\mathcal{M}_C$, the unlocking sequence is short: we need to update any vertex adjacent to v before we update v itself. In general, though, longer unlocking sequences could be used, with a corresponding worsening of the constants in the bound on the mixing time.

Chapter 4

Sampling matchings of general and lattice graphs

Having introduced methods for bounding mixing times of Markov chains, we now apply these techniques to the problem of sampling matchings in graphs. We begin by adapting Dyer and Greenhill's chain for sampling independent sets to sample matchings in graphs with bounded degree, for restricted values of λ.

We also show that we can use block updates to sample matchings in lattice graphs for arbitrary values of λ, and that this implies that the Jerrum-Sinclair chain mixes in $O(nm)$ time for these graphs.

We are already familiar with the concept of a matching and a perfect matching from Definition 2.1. In addition, we will define a near-perfect matching as a matching with exactly two uncovered vertices (sometimes called holes) v_1 and v_2:

$$\bigcup_{(u,v)\in M} \{u,v\} = V \setminus \{v_1, v_2\}.$$

Near-perfect matchings have been important for existing Markov chains that use simulated annealing to sample perfect matchings in bipartite graphs [27]. In this chapter, we will use near-perfect matchings in Theorem 4.1 to show that the Jerrum-Sinclair chain for sampling general matchings that we introduced in Definition 2.25 of Chapter 2 cannot efficiently sample perfect matchings of general bipartite graphs.

It immediately follows from Definition 2.1 that M is a matching if and only if the degree of every vertex $v \in V$ with respect to M is either 0 or 1. M is a perfect matching if and only if the degree of every vertex $v \in V$ with respect to M is 1.

It is possible to determine in polynomial time whether a perfect matching exists in any graph $G = (V,E)$ by searching for *augmenting paths* [3, Theorem 5.1]. In this

context, an augmenting path for a matching M is a path P in G such that the edges of P alternately belong to M and $E \setminus M$, and such that the first and last vertices of P are unmatched in M. If G is bipartite, then an augmenting path (if one exists) can be found by breadth-first search. Edmonds developed a polynomial time algorithm to find a maximum matching in general graphs [15]. However, in common with many combinatorial structure decision problems, the problem of counting perfect matchings is #P-hard [40].

The problem of sampling or counting perfect matchings of a bipartite graph $G = (V_1, V_2, E)$ is of particular interest, because there are many #P-complete problems which have a natural reduction to perfect matchings. The number of perfect matchings of a bipartite graph is equal to the permanent of its adjacency matrix. A perfect matching also corresponds to a bijection between the two vertex sets V_1 and V_2. This property is used in permutation tests in statistics, where it is necessary to generate random permutations from the uniform distribution.

There are algorithms for approximately counting, and for almost-uniformly sampling, perfect matchings of bipartite graphs in polynomial time [26]. These algorithms use a Markov chain that samples from perfect and near-perfect matchings. In Chapter 6 we will investigate an alternative chain due to Diaconis, Graham and Holmes that samples perfect matchings of bipartite graphs directly, and we will analyse this chain in special classes of bipartite graphs. There are currently no known polynomial-time algorithms for almost uniform sampling or approximately counting perfect matchings in general graphs.

When sampling from all matchings, we often want to select matchings with varying weights, depending on their size. Our goal is to sample from the stationary distribution given in Equation (2.7),

$$\pi(M) = \frac{\lambda^{|M|}}{\sum_M \lambda^{|M|}}.$$

The value $Z = \sum_M \lambda^{|M|}$ is known as the partition function. When $\lambda = 1$, Z is equal to the number of matchings of G. Determining Z in the case where $\lambda = 1$ is therefore equivalent to counting the total number of matchings of G.

Note that when $\lambda = 1$, the distribution is uniform over all matchings. It is possible to efficiently sample general matchings with arbitrary values of λ, using the Jerrum-Sinclair chain we saw in Chapter 2 [25]. However it is not feasible to efficiently sample perfect matchings using this chain, as we explain below.

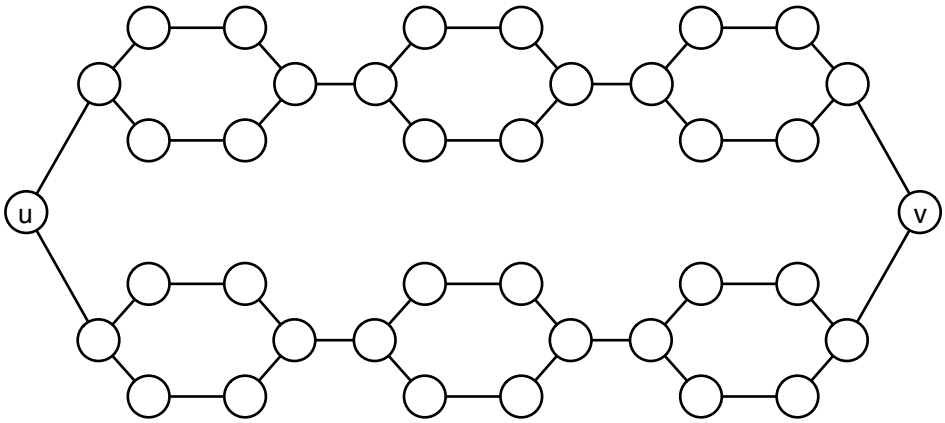

Figure 4.1: *A graph for which the Jerrum-Sinclair chain requires exponential time to sample perfect matchings (case $k = 3$).*

Theorem 4.1. *The time required for the Jerrum-Sinclair chain to sample a perfect matching from the uniform distribution is exponential in the worst case.*

Proof. We will show that there exists a family of graphs for which the mixing time of the Jerrum-Sinclair chain is exponential. Each graph in our family consists of a pair of chains of hexagons. Figure 4.1 shows the graph G, in which each chain contains three hexagons.

First, observe that the only moves available to the Jerrum-Sinclair chain from a perfect matching are delete moves, and that these will necessarily yield a near-perfect matching. Similarly, a perfect matching can only be reached from a near-perfect matching, using an insert move. It follows that for any given perfect matching M, there are exactly $|M|$ near-perfect matchings from which M can be reached in a single step.

Also note that, regardless of the choice of λ, every near-perfect matching has the same probability in the stationary distribution of the Jerrum-Sinclair chain.

Consider the graph G in Figure 4.1. This consists of a pair of chains of hexagons, each of length k, connected at each end by an additional vertex (Figure 4.1 shows the case where $k = 3$). Call these extra vertices u and v. Suppose we have a perfect matching of G that includes the edge leading upwards from u. Then the configuration of the entire upper chain of hexagons is uniquely determined. Each hexagon in the lower chain may take one of two configurations. G therefore has exactly 2^{k+1} perfect matchings. The additional factor of 2 comes from the initial choice of the edge covering

u.

Now consider the near-perfect matchings of G that leave the vertices u and v uncovered. All of the hexagons in both chains have two possible configurations. There are therefore at least 2^{2k} near-perfect matchings of G. There are also a large number of additional near-perfect matchings which have different uncovered vertices.

If we wish to sample perfect matchings in polynomial time, we certainly need the total weight of perfect matchings to be smaller than the total weight of near-perfect matchings by at most a polynomial factor. Were this not the case, the expected number of samples required to obtain a perfect matching would be super-polynomial. Therefore we require

$$\lambda \geq \frac{2^{k-1}}{p(k)},$$

where $p(k)$ is a polynomial in k.

Now let S be the set of perfect matchings of G. The conductance of $\mathcal{M}_{JS}$ on G is no greater than

$$\begin{aligned}\Phi_S &\leq \frac{\sum_{x\in S, y\notin S} \pi(x)P(x,y)}{\pi(S)} \\ &= (6k+1)P(x,y) \\ &\leq \frac{(6k+1)p(k)}{2^{k-1}}.\end{aligned}$$

Therefore by Theorem 2.22, the spectral gap is $1-|\lambda_1| \leq \frac{(6k+1)p(k)}{2^{k-1}}$, and so by Theorem 2.21 the mixing time is at least

$$\begin{aligned}\tau(\varepsilon) &\geq \frac{|\lambda_1|}{2(1-\max\{|\lambda_{\min}|,|\lambda_1|\})}\log\left(\frac{1}{2\varepsilon}\right) \\ &\geq \left(\frac{2^k}{(6k+1)p(k)}-2\right)\left(\frac{1}{2\varepsilon}\right).\end{aligned}$$

Therefore, the mixing time of $\mathcal{M}_{JS}$ on the graph G is exponential in the size of G, as required. □

The Markov chain proposed by Broder [5] uses the same transitions as the Jerrum-Sinclair chain, but with different transition probabilities and the state space restricted to only perfect and near-perfect matchings. The argument used in the proof of Theorem 4.1 can also be applied to Broder's chain.

Note that the proof of Theorem 4.1 relies on the fact that the Jerrum-Sinclair chain described in Chapter 2 weights all near-perfect matchings equally. Existing

polynomial-time algorithms for sampling perfect matchings in bipartite graphs sample from perfect and near-perfect matchings, and use simulated annealing to approximate weights for each possible pair of "holes" in near-perfect matchings [26]. Given these weights, it is possible to set transition probabilities such that perfect matchings are given a high enough weight to be sampled in polynomial time. The chain no longer samples near-perfect matchings uniformly, instead giving a higher probability of selecting near-perfect matchings that are close to perfect matchings. Markov chains for sampling perfect matchings using simulated annealing have only been shown to mix rapidly for bipartite graphs - it remains an open question whether a method exists for efficiently sampling perfect matchings in general graphs.

In Chapter 6, we will consider an alternative way of sampling perfect matchings in some special classes of bipartite graphs.

We can improve the bounds on the mixing time given by the canonical paths argument for the Jerrum-Sinclair chain, for restricted values of λ, by adapting existing Markov chains for independent sets. In Section 4.1, we show a correspondence between matchings and independent sets, and therefore show how to apply Dyer and Greenhill's chain for sampling independent sets [14] to the problem of sampling matchings. We then define a similar chain for sampling matchings directly, and show that the maximum λ for which this new chain is rapidly mixing can be improved slightly over that achieved by applying Dyer and Greenhill's chain directly.

In Section 4.2, we look at the problem of sampling matchings in regular lattice graphs. We use spatial properties of these graphs to show that a Markov chain using large but constant size block updates is rapidly mixing. We will follow the approach of Van den Berg and Brouwer, who showed a similar result [42], and noted that implementing these block moves is infeasible, even for relatively small blocks. They overcame this problem by defining a second chain using single-site updates, and using this chain to approximate the moves made by their block chain. They also noted that comparison techniques could be used to bound the mixing time of the single-site chain, but that such techniques would lead to a worsening of the hidden constants in the bound on the mixing time. We will follow this approach and use comparison techniques to show that the mixing time of the Jerrum-Sinclair chain is $O(nm)$ for lattice graphs.

4.1 Matchings in general graphs

In this section we will adapt the Dyer-Greenhill chain for sampling independent sets to sample matchings. We begin by defining the Dyer-Greenhill chain.

4.1.1 The Dyer-Greenhill chain for independent sets

As with matchings, we are often interested in sampling independent sets weighted by a parameter λ, distributed according to Equation (2.3). Luby and Vigoda defined a Markov chain for sampling independent sets, and used coupling to show that it is rapidly mixing when $\lambda \leq \frac{1}{\Delta-3}$, provided $\Delta \geq 4$. Dyer and Greenhill used path coupling to show an improved upper bound on the mixing time of the Luby-Vigoda chain, and also defined another Markov chain for sampling independent sets. They showed that their chain is rapidly mixing when $\lambda \leq \frac{2}{\Delta-2}$ using path coupling [14]. In Chapter 3, we introduced the Dyer-Greenhill chain, and showed that its mixing time when $\lambda = \frac{2}{\Delta-2}$ is $O(n^2)$.

We now show how to apply the Dyer-Greenhill chain to matchings, by considering the *line graph* of G, $\mathcal{L}(G)$. The line graph of a graph G is defined as follows:

Definition 4.2. *Let $G = (V,E)$ be a graph. Define a vertex set V' and an edge set E' on V' as follows:*

$$V' = \{v_e \mid e \in E\}$$
$$E' = \{(v_e, v_{e'}) \mid e \text{ and } e' \text{ are edges incident to a common vertex in } G\}.$$

The line graph *of G is $\mathcal{L}(G) = (V', E')$.*

Suppose I is an independent set in $\mathcal{L}(G)$. Each vertex in I corresponds to an edge of G. Since no pair of vertices of I are adjacent, no pair of the corresponding edges share a common endpoint. Therefore, I corresponds to a matching in G. Similarly, if M is a matching of G, then each edge in M corresponds to a vertex in $\mathcal{L}(G)$. No pair of edges are incident to a common vertex, so there is no edge connecting any pair of these vertices in $\mathcal{L}(G)$. There is therefore a bijection between matchings in G and independent sets in $\mathcal{L}(G)$.

If the maximum degree of the original graph G is Δ, then the maximum degree of $\mathcal{L}(G)$ is at most $2\Delta - 2$, since each endpoint of an edge can be shared with at most $\Delta - 1$ vertices. We can therefore sample matchings in G using the Dyer-Greenhill chain, provided $\lambda \leq \frac{1}{\Delta-2}$.

This bijection allows us to sample matchings of line graphs. However, line graphs also belong to a more general class of graphs called claw-free graphs. We will show in Chapter 5 that we can modify the Jerrum-Sinclair chain for sampling matchings to sample independent sets in general claw-free graphs.

4.1.2 A Markov chain for sampling matchings

In this section, we adapt the Dyer-Greenhill chain from Definition 3.4 to sample directly from weighted matchings, and hence obtain a proof of mixing for larger values of λ. While Dyer and Greenhill's analysis uses Hamming distance as a metric [14], we introduce a new metric which allows us to improve the maximum value of λ for which the chain is rapidly mixing. We will define a new chain $\mathcal{M}_{IDS}$, using the insert and delete moves from Dyer and Greenhill's chain, and a *slide* move based on their drag move. $\mathcal{M}_{IDS}$ is mostly analogous to the Dyer-Greenhill chain, but we assign a different probability to the slide move.

Given a matching M in a graph $G = (V,E)$, we call an edge $e \in E$ "blocked" if there is an edge $e' \in M$ such that e and e' share a common endpoint. e may be blocked by one or two other edges. We can now define the transitions of $\mathcal{M}_{IDS}$ as follows:

Definition 4.3. *Let X_t be the state of $\mathcal{M}_{IDS}$ at time t. The subsequent state X_{t+1} is determined by the following sequence of steps:*

(IDS1) Select an edge $e \in E$ uniformly at random.

*(IDS2) If $e \in X_t$, then let $X_{t+1} = X_t \setminus \{e\}$ with probability $\frac{1}{1+\lambda}$ (*delete*).*

*(IDS3) If $e \notin X_t$, and e is not blocked, then let $X_{t+1} = X_t \cup \{e\}$ with probability $\frac{\lambda}{1+\lambda}$ (*insert*).*

*(IDS4) If $e \notin X_t$, and e is blocked by exactly one edge $e' \in X_t$, then let $X_{t+1} = X_t \cup \{e\} \setminus \{e'\}$ with probability p (*slide*).*

(IDS5) Otherwise, let $X_{t+1} = X_t$.

The slide move is a direct analogue of Dyer and Greenhill's drag move. The probability p will be determined later, during the analysis of the mixing time of $\mathcal{M}_{IDS}$.

We will now show that $\mathcal{M}_{IDS}$ is ergodic, and that its unique stationary distribution is given by Equation (2.7).

Lemma 4.4. *The Markov chain $\mathcal{M}_{IDS}$ is ergodic, and has the unique stationary distribution*

$$\pi = \frac{\lambda^{|M|}}{\sum_M \lambda^{|M|}}.$$

Proof. Let $M \subseteq E$ be a matching of G. Suppose that some edge $e \in E$ is chosen in a transition of $\mathcal{M}_{IDS}$. If $e \in M$, then with probability $\frac{\lambda}{1+\lambda}$, e will remain in M. If $e \notin M$, then with probability $\frac{1}{1+\lambda}$, e will not be added to M. Therefore, for any M, there is a self-loop probability of at least $\min\{\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\}$, and therefore $\mathcal{M}_{IDS}$ is aperiodic.

To see that $\mathcal{M}_{IDS}$ is irreducible, observe that delete moves always have non-zero probability, and so the empty set can be reached from any matching by a sequence of delete moves. Similarly, if we wish to reach a matching M from the empty set, we can add each edge $e \in M$ in turn. None of these edges can be blocked (since M is a matching), and so these insert moves have non-zero probability. Therefore any two matchings M and M' are connected by a sequence of insert and delete moves, and so $\mathcal{M}_{IDS}$ is irreducible. It is therefore ergodic, and has a unique stationary distribution.

We now show that $\mathcal{M}_{IDS}$ has the stationary distribution $\pi = \frac{\lambda^{|M|}}{\sum_M \lambda^{|M|}}$, by showing that this distribution and the transition matrix P of $\mathcal{M}_{IDS}$ satisfy the detailed balance condition. We show that Condition (2.2) holds for every pair of matchings, M and M'. We must consider the following cases:

1. $|M| = |M'|$ and $P(M,M') \neq 0$. Since $|M| = |M'|$, we know that $\pi(M) = \pi(M')$. The only possible transition between two matchings of the same size is a slide move, which occurs with probability p. Therefore $P(M,M') = P(M',M)$, and Condition (2.2) holds in this case.

2. There is some edge e such that $M = M' \cup \{e\}$. From the definition of π, we know that $\pi(M) = \lambda\pi(M')$. M is reachable from M' by an insert move, and so $P(M',M) = \frac{\lambda}{n(1+\lambda)}$. M' is reachable from M by a delete move, and so $P(M,M') = \frac{1}{n(1+\lambda)}$. Therefore

$$\begin{aligned}\pi(M)P(M,M') &= \lambda\pi(M')\frac{1}{n(1+\lambda)} \\ &= \pi(M')\frac{\lambda}{n(1+\lambda)} \\ &= \pi(M')P(M',M).\end{aligned}$$

3. There is some edge e such that $M' = M \cup \{e\}$. This case is analogous to case 2.

4. In all other cases, $P(M,M') = P(M',M) = 0$, and so Condition (2.2) holds trivially.

Therefore the detailed balance condition holds for every pair of states M and M', and so $\mathcal{M}_{IDS}$ is time-reversible with respect to π. By Lemma 2.8, π is therefore the unique stationary distribution of $\mathcal{M}_{IDS}$. □

Theorem 4.5. *For any* $0 < \lambda \leq \frac{\Delta+\sqrt{3\Delta^2-5\Delta+2}}{2\Delta^2-5\Delta+2}$*, there is a value of* p *such that the mixing time of* $\mathcal{M}_{IDS}$ *is dominated by a binomial random variable* τ^* *with expectation at most*

$$\mathbb{E}[\tau^*(\varepsilon)] \leq \left\lceil \max\{en^2(n(1+\lambda)+1), en^2(2n(\Delta+\frac{1}{\lambda})+1)\} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil$$

Proof. We will prove Theorem 4.5 by path coupling. Instead of Hamming distance, we define a new metric. Let X_t and Y_t be two copies of $\mathcal{M}_{IDS}$ at time t. Consider the symmetric difference $X_t \oplus Y_t$. We will say that X_t and Y_t are adjacent if either $|X_t \oplus Y_t| = 1$, or $X_t \oplus Y_t$ is a path of length two (these correspond to the cases where there is a transition between X_t and Y_t). The adjacency relation S is defined as follows:

$$S = \{(M,M') : |M \oplus M'| = 1\} \cup \{(M,M') : \exists u,v,w \text{ s.t. } M \oplus M' = \{(u,v),(v,w)\}\}.$$

If $|X_t \oplus Y_t| = 1$, then the distance $d(X_t,Y_t) = 1$. If $X_t \oplus Y_t$ is a path of length two, then $d(X_t,Y_t) = 1+x$, where x is a positive constant to be chosen later, and $x < 1$. We can now extend d to a path metric on $\Omega \times \Omega$: $d(X_t,Y_t)$ is the minimum sum of distances of the transitions from X_t to Y_t, computed over all paths from X_t to Y_t.

We can also define d in terms of the paths and cycles making up $X_t \oplus Y_t$. For any X_t and Y_t, $X_t \oplus Y_t$ is composed of a set of paths and even-length cycles [25]. Each even-length path of length ℓ is given the value $\frac{\ell}{2}(1+x)$; each odd-length path is given the value $\frac{\ell-1}{2}(1+x)+1$; and each even-length cycle is given the value $\frac{l-2}{2}(1+x)+2$. These values correspond to the shortest sequences of transitions from X_t to Y_t for their respective components. The distance $d(X_t,Y_t)$ is simply the sum of the values assigned to each component of $X_t \oplus Y_t$.

Note that this metric does not necessarily take integer values, while Theorem 2.15 requires an integer-valued metric. We will instead use Corollary 3.2 to bound the mixing time of $\mathcal{M}_{IDS}$. Corollary 3.2 requires a bound on the variance of $d(X_{t+1},Y_{t+1})$ (conditioned on $d(X_t,Y_t)$), and avoids the need for an integer metric.

The parameters p and x will be determined later.

Assume that $\Delta \geq 3$, $\lambda > 0$, and $0 \leq p \leq \frac{\lambda}{1+\lambda}$.

We need to consider two cases to apply path coupling: when the distance $d(X_t, Y_t) = 1$, and when $d(X_t, Y_t) = 1 + x$. From these cases, we will obtain a set of bounds on the expected change in distance. First we consider $d(X_t, Y_t) = 1$.

Assume that there is a unique edge e_0 such that $e_0 \in X_t \setminus Y_t$ (the other case, where $e_0 \in Y_t \setminus X_t$, is equivalent). All other edges will be the same in both copies of the chain. We select an edge e uniformly at random.

1. If $e = e_0$, then with probability $\frac{1}{1+\lambda}$, we delete e from X_t and make no change to Y_t. With the remaining probability $\frac{\lambda}{1+\lambda}$, we insert e into Y_t and make no change to X_t. Therefore, with probability 1, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$.

2. If e is incident to an endpoint of e_0, and e is blocked by some edge $e' \in X_t \cap Y_t$, then there can be no change in X_t as e is blocked by both e_0 and e'. In Y_t, e is blocked by exactly one edge, and so with probability p, $Y_{t+1} = Y_t \cup \{e\} \setminus \{e'\}$. This yields a path of three edges in $X_{t+1} \oplus Y_{t+1}$, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1 + x$. Otherwise, with the remaining probability $1 - p$, there is no change.

3. If e is incident to an endpoint of e_0 and there is no other edge blocking e, then we make a slide move in X_t with probability p, and insert e into Y_t with probability $\frac{\lambda}{1+\lambda}$. We would like to do both of these together with maximum probability, as this will decrease the distance between X and Y. Since $p \leq \frac{\lambda}{1+\lambda}$, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$ with probability p. With the remaining probability $\frac{\lambda}{1+\lambda} - p$, we insert e into Y_t only, and $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + x$. Therefore, in this case the expected change in distance is $(\frac{\lambda}{1+\lambda} - p)x - p$.

4. In all other cases, e and all other edges incident to its endpoints agree in X_t and Y_t. We can make exactly the same move in X_t and Y_t, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t)$.

We will now construct upper bounds on the expected change in distance $\mathbb{E}[d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)]$, conditioned on $d(X_t, Y_t) \in \{1, 1 + x\}$. Note that there are at most $2\Delta - 2$ edges for which cases 2 or 3 can apply. If G is Δ-regular, then there are exactly $2\Delta - 2$ such edges, and the expected change is guaranteed to lie between the change when case 2 applies for every one of these edges, and the change when case 3 applies. If G is not Δ-regular, then there are fewer than $2\Delta - 2$ edges for which cases 2 or 3 can apply. Since case 2 always leads to an expected increase in distance when one of these edges is chosen, it follows that the expected increase when case 2 applies to every such edge is no greater than that for a Δ-regular graph. If the values of p and x are such that case 3 leads to an expected increase in distance when one of these edges

is chosen, then the increase when case 3 applies for all such edges is also no greater for a non-Δ-regular graph than a Δ-regular one. If case 3 leads to an expected decrease in distance, then the formula f_1 which we will construct as an upper bound is always negative, regardless of the number of edges for which cases 2 and 3 can apply.

We will now state two formulae providing upper bounds on the expected change in distance when $d(X_t, Y_t) = 1$. Let

$$f_1 = \frac{1}{n}\left((2\Delta-2)\left(\left(\frac{\lambda}{1+\lambda} - p\right)x - p\right) - 1\right), \qquad \text{and} \qquad (4.1)$$

$$f_2 = \frac{1}{n}((2\Delta-2)p(x+1) - 1). \qquad (4.2)$$

f_1 applies in the case where $d(X_t, Y_t) = 1$, G is Δ-regular, and case 3 applies to every one of the $2\Delta - 2$ edges incident to the endpoints of e_0. As we have discussed, if G is not Δ-regular, then either f_1 still provides an upper bound, or f_1 (and the expected change in this case, even for a non-Δ-regular graph) is negative. f_2 applies in the case where $d(X_t, Y_t) = 1$, G is Δ-regular, and case 2 applies to every edge incident to the endpoints of e_0. f_2 is always an upper bound on the change in distance, even for non-Δ-regular graphs.

Now suppose $d(X_t, Y_t) = 1 + x$. Assume that for some pair of edges e_0 and e_1, $e_0 \in X_t \setminus Y_t$, $e_1 \in Y_t \setminus X_t$, and that e_0 and e_1 are incident to some common vertex v. As before, we select an edge e uniformly at random, and consider the expected change in distance for each possible position of e with respect to e_0 and e_1. Instead of always updating the same edge in each copy of the chain, we update e in X_t and choose some edge e' that will be updated in Y_t. The choices of e' are such that Y_t is still a faithful copy of $\mathcal{M}_{IDS}$. In cases 1 and 2, the choice of e' is either e_0 or e_1, with appropriate probabilities; in cases 3 to 8, we always choose $e' = e$.

1. If $e = e_0$, then with probability p, let $e' = e_0$. Perform a slide move in Y_t and make no change in X_t. Both e_0 and e_1 now agree, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - (1+x)$. With the remaining probability $1 - p$, let $e' = e_1$. With probability $\frac{1}{1+\lambda}$, remove e from X_t and e' from Y_t. Again, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - (1+x)$.

2. If $e = e_1$, then with probability p, let $e' = e_1$. Perform a slide move in X_t and make no change in Y_t. As in case 1, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - (1+x)$. With the remaining probability $1 - p$, let $e' = e_0$, and make no change in either copy of the chain.

3. If e is not e_0 or e_1, e is incident to v, and e is blocked by some other edge in $X_t \cap Y_t$, then no change occurs in either chain, and $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t)$.

4. If e is not e_0 or e_1, e is incident to v, and there is no edge in $X_t \cap Y_t$ blocking e, then with probability p, make a slide move in both copies of the chain. e_0, e_1 and e now agree, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - (1+x)$.

5. If e shares an endpoint with exactly one of e_0 and e_1, and is blocked by some other edge, then with probability p, we perform a slide move in Y_t and make no change in X_t, and so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1 + x$. With the remaining probability $1-p$, we make no change. Note that the situation in this case is the same as case 2 when $d(X_t, Y_t) = 1$, since only one of e_0 and e_1 shares an endpoint with e.

6. If e shares an endpoint with exactly one of e_0 and e_1, and is not blocked by any other edge, then with probability p, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - x$, and with probability $\frac{\lambda}{1+\lambda} - p$, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1$. The coupling is equivalent to case 3 when $d(X_t, Y_t) = 1$, but the change in distance is different since we start with an even length path in $X_t \oplus Y_t$ instead of a single edge.

7. If e shares endpoints with both e_0 and e_1, but is not incident to v (that is, e forms a triangle with e_0 and e_1), then with probability p, make a slide move in both copies of the chain. e_0, e_1 and e now agree, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - (1+x)$.

8. In all other cases, make the same move in X_t and Y_t, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t)$.

The total number of instances of cases 3 and 4 is at most $\Delta - 2$, and the total number of instances of cases 5 and 6 is at most $2\Delta - 2$. We wish to construct upper bounds on the expected change in distance, as we did for the case when $d(X_t, Y_t) = 1$. Note that case 3 leads to no change in distance, and case 4 always leads to a decrease. Therefore if we assume that case 3 applies for all edges incident to v, then the formulae we construct will be upper bounds on the expected change in distance. Let

$$f_3 = \frac{1}{n}\left((2\Delta - 2)\left(\frac{\lambda}{1+\lambda} - p(x+1)\right) - 2p(x+1) - \frac{x+1}{1+\lambda}\right), \qquad \text{and} \tag{4.3}$$

$$f_4 = \frac{1}{n}\left((2\Delta - 4)p(1+x) - \frac{1+x}{1+\lambda}\right). \tag{4.4}$$

f_3 provides an upper bound for the situation where case 6 applies for all $2\Delta - 2$ edges for which cases 5 and 6 can apply; f_4 gives an upper bound for the situation where case 5 applies for all of these edges. There may be situations where case 7 can apply, depending on the structure of the graph. Since it can never lead to an increase in distance, however, we may safely assume that it does not occur.

We now have four formulae that provide upper bounds on the expected change in distance where $d(X_t,Y_t)=1$ or $d(X_t,Y_t)=1+x$. If $d(X_t,Y_t)=1$, then $\mathbb{E}[d(X_{t+1},Y_{t+1})-d(X_t,Y_t)]$ is bounded above by some weighted average of f_1 and f_2. Similarly, if $d(X_t,Y_t)=1+x$ then $\mathbb{E}[d(X_{t+1},Y_{t+1})-d(X_t,Y_t)]$ is bounded above by a weighted average of f_3 and f_4. Therefore, when $d(X_t,Y_t)=1$ or $d(X_t,Y_t)=1+x$,

$$\mathbb{E}[d(X_{t+1},Y_{t+1})-d(X_t,Y_t)] \leq \max\{f_1,f_2,f_3,f_4\}. \tag{4.5}$$

We now wish to find values for p, x and λ such that $\max\{f_1,f_2,f_3,f_4\}\leq 0$. Observe that $f_1+f_4-f_2-f_3=\frac{1}{n}(2\Delta-2)\frac{\lambda}{1+\lambda}(x-1)<0$ (recall our assumptions that $\Delta\geq 3$ and $x<1$). Therefore, if we are able to solve the system of equations $\{f_1=0,f_2=0,f_3=0\}$, it will follow that $f_4\leq 0$ and we will have an upper bound on $\mathbb{E}[d(X_{t+1},Y_{t+1})-d(X_t,Y_t)]$ of 0, conditioned on $d(X_t,Y_t)\in\{1,1+x\}$.

We begin by observing that if $f_2=0$ then $p(x+1)=\frac{1}{2\Delta-2}$. By substituting this equation into f_1 and solving $f_1=0$, we see that

$$\begin{aligned}
&(2\Delta-2)\left(\left(\frac{\lambda}{1+\lambda}-p\right)x-p\right)-1=0\\
\Rightarrow\quad &(2\Delta-2)\left(\frac{\lambda x}{1+\lambda}-p(x+1)\right)=1\\
\Rightarrow\quad &(2\Delta-2)\frac{\lambda x}{1+\lambda}=2\\
\Rightarrow\quad &x=\frac{1+\lambda}{\lambda(\Delta-1)}.
\end{aligned}$$

Now we may substitute this value for x back into the equation obtained from f_2 to obtain a value for p:

$$\begin{aligned}
&p\left(\frac{1+\lambda}{\lambda(\Delta-1)}+1\right)=\frac{1}{2\Delta-2}\\
\Rightarrow\quad &p=\frac{\lambda(\Delta-1)}{(2\Delta-2)(1+\lambda+\lambda(\Delta-1))}\\
&\quad=\frac{\lambda}{2(\lambda\Delta+1)}.
\end{aligned}$$

Finally, we substitute our equation from f_2 and our value for x into the equation $f_3=0$ and solve the resulting quadratic equation for λ:

$$\begin{aligned}
&(2\Delta-2)\left(\frac{\lambda}{1+\lambda}-p(x+1)\right)-2p(x+1)-\frac{x+1}{1+\lambda}=0\\
\Rightarrow\quad &(2\Delta-2)\frac{\lambda}{1+\lambda}-1-\frac{1}{\Delta-1}-\frac{1}{\lambda(\Delta-1)}-\frac{1}{1+\lambda}=0\\
\Rightarrow\quad &(\Delta-1)(2\Delta-2)\lambda^2-(\Delta-1)\lambda(1+\lambda)-\lambda(1+\lambda)-(1+\lambda)-\lambda(\Delta-1)=0\\
\Rightarrow\quad &(2\Delta^2-5\Delta+2)\lambda^2-2\Delta\lambda-1=0
\end{aligned}$$

$$\begin{aligned}\lambda &= \frac{2\Delta \pm \sqrt{12\Delta^2 - 20\Delta + 8}}{2(2\Delta^2 - 5\Delta + 2)} \\ &= \frac{\Delta \pm \sqrt{3\Delta^2 - 5\Delta + 2}}{2\Delta^2 - 5\Delta + 2}.\end{aligned}$$

Since we are interested in maximising the value of λ for which $\mathcal{M}_{IDS}$ is rapidly mixing, we take the larger of the two solutions. We can verify that $\max\{f_1, f_2, f_3, f_4\} \leq 0$ whenever λ is less than this larger value.

We therefore find that $\max\{f_1, f_2, f_3, f_4\} \leq 0$ when

$$\begin{aligned}\lambda &\leq \frac{\Delta + \sqrt{3\Delta^2 - 5\Delta + 2}}{2\Delta^2 - 5\Delta + 2} \\ x &= \frac{1+\lambda}{\lambda(\Delta - 1)} \\ p &= \frac{\lambda}{2(\lambda\Delta + 1)},\end{aligned}$$

if $\Delta \geq 3$.

We now apply Lemma 2.14 to show that there is a coupling such that $\mathbb{E}[d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)]$ is non-positive for all pairs of states X_t and Y_t. It follows from Lemma 2.14 that there exist values of x and p such that $d(X_{t+1}, Y_{t+1}) \leq d(X_t, Y_t)$ for all pairs of states X_t, Y_t, provided $\lambda \leq \frac{\Delta + \sqrt{3\Delta^2 - 5\Delta + 2}}{2\Delta^2 - 5\Delta + 2}$. Note that $p(x+1) = \frac{1}{2\Delta - 2}$, and therefore f_2 is independent of λ. It follows that for all λ, $f_2 = 0$, and so $\max\{f_1, f_2, f_3, f_4\} = 0$. The best bound we can obtain on β, such that $d(X_{t+1}, Y_{t+1}) \leq \beta d(X_t, Y_t)$, is therefore $\beta = 1$.

Since we have $\beta = 1$, we are unable to apply the first part of Theorem 2.15 (which requires $\beta < 1$) to bound the mixing time. Instead, we need to show that there is a sufficiently large probability that the distance changes in any step of the Markov chain. If the coupling given by the path coupling lemma always updates the same edge in each copy of the chain, then this is straightforward: if we select an edge $e \in X_t \setminus Y_t$, then there is a non-zero probability of removing e from X_t and making no change to Y_t. However, since our coupling does not necessarily update the same edge in each copy, we cannot generally guarantee that this occurs. We can easily see, however, that when X_t and Y_t are adjacent, there is an improvement of at least 1 with probability at least $\frac{2p}{n}$. This allows us to use Bordewich and Dyer's Theorem 2.16 to infer that a modified chain exists that is rapidly mixing.

In order to apply Theorem 2.16, we need to find a $\delta > 0$ such that the probability

$$p = p(\delta) = \min_{(v,w) \in S} \Pr(d(v, X_{t+1}) \geq \delta, d(v, X_{t+1}) + d(X_{t+1}, w) = d(v, w) \mid X_t = v)$$

is strictly greater than 0.

Suppose that $d(v,w) = 1$. Then there is a single edge e that differs between v and w. If $e \in X_t$, then with probability $\frac{1}{n(1+\lambda)}$, $X_{t+1} = w$; if $e \notin X_t$, then with probability $\frac{\lambda}{n(1+\lambda)}$, $X_{t+1} = w$.

Now suppose $d(v,w) = 1+x$. Then there are two adjacent edges e and e' that differ between v and w. Assume that $e \in X_t$. With probability $p = \frac{\lambda}{2n(\lambda\Delta+1)}$, a slide move is performed and $X_{t+1} = w$. We do not consider the case where e is deleted, because this fails to satisfy the condition that $d(v,X_{t+1}) + d(X_{t+1},w) = d(v,w)$.

We therefore see that $p(1) = \min\{\frac{1}{n(1+\lambda)}, \frac{\lambda}{2n(\lambda\Delta+1)}\}$. Applying Theorem 2.16, we see that there exists a modified chain $\mathcal{M}^*_{IDS}$ which has a coupling $\mathcal{C}$ such that $\beta(\mathcal{M}^*_{IDS}, \mathcal{C}) \leq 1$ and $\sigma^2(\mathcal{M}^*_{IDS}, \mathcal{C}) \geq \min\{\frac{1}{n(1+\lambda)+1}, \frac{\lambda}{2n(\lambda\Delta+1)+\lambda}\}$.

The transition probabilities of $\mathcal{M}^*_{IDS}$ are given by:

$$\Pr_{\mathcal{M}^*_{IDS}}(X_{t+1} = x' \mid X_t = x) = \begin{cases} \frac{\Pr_{\mathcal{M}_{IDS}}(X_{t+1}=x|X_t=x)+p}{1+p} & \text{if } x' = x \\ \frac{\Pr_{\mathcal{M}_{IDS}}(X_{t+1}=x'|X_t=x)}{1+p} & \text{otherwise.} \end{cases}$$

We can now apply Corollary 3.2 to bound the mixing time of $\mathcal{M}^*_{IDS}$:

$$\begin{aligned} \tau(\varepsilon) &\leq \left\lceil \frac{eD^2}{\sigma^2} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil \\ &= \left\lceil \max\{en^2(n(1+\lambda)+1), en^2(2n(\Delta+\frac{1}{\lambda})+1)\} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil. \end{aligned}$$

Therefore, the mixing time of $\mathcal{M}^*_{IDS}$ is polynomial, and the mixing time of $\mathcal{M}_{IDS}$ is dominated by a binomial random variable $\tau^*(\varepsilon) = \text{Bin}(\tau(\varepsilon), (1+p)^{-1})$ [4]. Since p is positive, $(1+p)^{-1} \leq 1$, and so the expected mixing time of $\mathcal{M}_{IDS}$ is bounded above by

$$\begin{aligned} \mathbb{E}[\tau^*(\varepsilon)] &\leq \left\lceil \max\{en^2(n(1+\lambda)+1), en^2(2n(\Delta+\frac{1}{\lambda})+1)\} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil \\ &= O(n^3). \end{aligned}$$

□

For certain Markov chains, it is possible to show that the probability of a change in distance at time $t+1$, α, is proportional to the distance at time t. We have not been able to show that this is the case for $\mathcal{M}_{IDS}$, but it is possible to do so for the unmodified Dyer-Greenhill chain. As we saw in Chapter 3, the bound on the mixing time of the Dyer-Greenhill chain can be improved to $O(n^2)$ when the probability that the distance changes is proportional to the distance, and therefore the Dyer-Greenhill chain mixes

in time $O(n^2)$ for $\lambda = \frac{2}{\Delta-2}$ - an improvement of $\log n$ over Dyer and Greenhill's result in the boundary case. It follows that if we use the Dyer-Greenhill chain to sample matchings, as we did in Section 4.1.1, then we can sample matchings in time $O(n^2)$ when $\lambda = \frac{1}{\Delta-2}$.

Unfortunately, we are not able to apply Theorem 3.3 to $\mathcal{M}_{IDS}$. The coupling we used to show rapid mixing does not have the property that the same edge is updated in each copy of the chain. It is therefore difficult to conclude that the probability of a change in distance is linear.

4.2 Matchings in lattice graphs

In this section, we consider the problem of sampling matchings in graphs with a lattice structure - in particular, the 2-dimensional torus. We will use a Markov chain $\mathcal{M}_\ell$ that updates large (but constant size) blocks of edges in a single step. While the Jerrum-Sinclair chain for sampling general matchings that we introduced in Section 2.3.5 allows us to sample matchings with arbitrary values of λ, in this section we will show mixing in $O(n \log n)$ time, albeit for a restricted class of graphs. We will also use comparison techniques to show that the Jerrum-Sinclair chain mixes in $O(nm)$ time for these graphs, where $n = |V|$ and $m = |E|$ of the graph $G = (V, E)$.

Our bound on the mixing time of $\mathcal{M}_\ell$ in Lemma 4.9 relies on spatial properties of the integer lattice, and is a duplication of a result by van den Berg and Brouwer [42]. The details of our proof differ from theirs, but we obtain the same asymptotic bound on the mixing time. We focus on the 2-dimensional torus, whereas van den Berg and Brouwer showed that their block chain mixes rapidly on the d-dimensional torus. Our proof technique could be applied to larger dimensions, with the replacement of some constants with appropriate functions of d. In Chapter 5, we will use a similar proof technique to show rapid mixing of a block chain for sampling independent sets in claw-free lattices. While the result of Lemma 4.9 is not new, the proof illustrates the techniques that we will use in Chapter 5 in the simpler setting of matchings. Our results for claw-free lattices in Chapter 5 will not be restricted to only 2-dimensional lattices.

It is infeasible to simulate a block chain such as $\mathcal{M}_\ell$ for large blocks, because we need some means of sampling random matchings of an $\ell \times \ell$ region of the underlying graph. Van den Berg and Brouwer used a second Markov chain to approximately sample matchings within the region selected by their block chain [42]. They also noted

that it is possible to use comparison techniques to infer the mixing time of a simpler chain, such as the Jerrum-Sinclair chain, but that comparison methods would worsen the mixing time by a factor of n. We will follow the comparison argument to bound the mixing time of the Jerrum-Sinclair chain.

Fix a constant block size ℓ. We will define a Markov chain $\mathcal{M}_\ell$ that samples matchings from a 2-dimensional toroidal graph $G = (V, E)$. Formally, the vertex and edge sets are defined as follows: for some integer N,

$$V = \{0, \ldots, \mathrm{N}-1\} \times \{0, \ldots, \mathrm{N}-1\}$$
$$E = \{(u,v) : u, v \in V;\ u \text{ and } v \text{ differ by 1 (modulo N} - 2) \text{ in exactly one dimension }\},$$

Let $|V| = n = \mathrm{N}^2$. Let M be the state of $\mathcal{M}_\ell$ at time t. The state M' at time $t+1$ is determined by the following sequence of steps:

(ℓ1) Select an $\ell \times \ell$ block of vertices R uniformly at random. Let ∂R be the edge boundary of R:

$$\partial R = \{(u,v) : u \in R, v \in V \setminus R, (u,v) \in E\}.$$

Let R' be the subgraph of G induced by the vertices $R \setminus V(\partial R \cap M)$, where $V(\partial R \cap M)$ is the set of endpoints of the edges of M in ∂R.

(ℓ2) Select a matching M_R of R' according to the distribution given in Equation (2.7) (this is the same weighted distribution that we used in Section 4.1).

(ℓ3) Let $M' = (M \setminus E(R')) \cup M_R$.

Figure 4.2 shows a section of the 2-dimensional lattice, with a 4×4 region R outlined. The edge boundary ∂R consists of the edges crossing the solid line around the region R, and the induced subgraph R' is the region inside the dashed line.

Note that our definition of ∂R here differs from that of Section 2.3.3, because we are concerned here with the configuration of edges instead of vertices.

We will now use path coupling to show that for a sufficiently large ℓ (determined by a function of Δ and λ), the chain $\mathcal{M}_\ell$ is rapidly mixing. Let X_t and Y_t be two copies of $\mathcal{M}_\ell$. We say that X_t and Y_t are adjacent if they differ at exactly one edge. The distance $d(X_t, Y_t)$ is the Hamming distance between the two states. Suppose that at time t, $d(X_t, Y_t) = 1$. Then we can assume without loss of generality that there is some edge $e \in X_t$, $e \notin Y_t$. We now show how to construct a path coupling for (X_t, Y_t), by considering three possible cases, depending on the position of e relative to R.

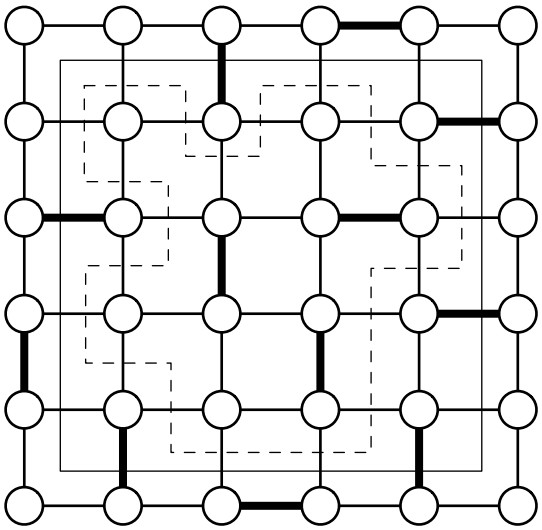

Figure 4.2: *An example of a region of the 2-dimensional lattice, showing the induced subgraph R'.*

1. If both endpoints of e are in R, then we can always choose the same M_R in each chain, thus making $X_{t+1} = Y_{t+1}$ with probability 1. When this occurs, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$.

2. If both endpoints of e are in $V \setminus R$, then we can choose the same M_R in each chain, and the distance is unchanged.

3. If e is in ∂R, then when we come to define R', we have $R'_X = R'_Y \setminus \{u\}$, where u is one of the endpoints of e. We must therefore choose M_R from different distributions in X_t and Y_t. We can couple these distributions so that the distance will either stay the same or increase (there is a non-zero probability of selecting the same M_R in each chain, since we are sampling from all matchings). The increase in distance is bounded above by a geometrically distributed random variable, as we will see in Lemma 4.7.

Observe that in the square lattice, there are 4ℓ choices of R such that e lies on the boundary of R, and there are $\ell^2 - \ell$ choices of R such that e lies within R. Let δ be the expected increase in distance in case 3. The overall expected change in distance is

$$\mathbb{E}[d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)] = \frac{4\ell\delta - \ell^2 + \ell}{n}. \tag{4.6}$$

If we choose $\ell \geq 4\delta + 1$, then the expected change is non-positive and the chain is rapidly mixing.

The requirement that our underlying graph is a torus comes from the method of choosing R: if we have a simple finite 2-dimensional grid, then for edges e close to the boundary of the underlying graph, the ratio of regions R containing e to regions R for which $e \in \partial R$ is constant, and so we cannot show contraction, regardless of the choice of ℓ. We could instead choose R by selecting a vertex v uniformly at random and letting R be the ball around v of radius ℓ. In this situation, however, $|R|$ and $|\partial R|$ are no longer necessarily uniform for all regions. More care is therefore required when we come to choose an appropriate value of ℓ. We will use this alternative method of choosing R in Chapter 5.

We now show that δ is stochastically dominated by a geometrically distributed random variable.

Definition 4.6. *Let X and Y be two random variables with cumulative distribution functions F_X and F_Y respectively. We say that Y is* stochastically dominated *by X if, for all x, $F_X(x) \leq F_Y(x)$.*

Lemma 4.7. *Let G be a 2-dimensional toroidal graph with maximum degree Δ, and X_t and Y_t two instances of $\mathcal{M}_\ell$ on G such that $|X_t \oplus Y_t| = 1$ and there is a unique edge $e \in X_t$, $e \notin Y_t$. There exists a coupling of (X_t, Y_t) such that the increase in distance when e lies on the edge boundary of a region R is stochastically dominated by a geometrically distributed random variable with mean at most $\delta \leq \max\{\lambda, 1\}(\Delta - 1)$.*

Proof. Let R'_X and R'_Y be the regions of X_t and Y_t to be updated. Note that there is a unique vertex v such that $R'_Y = R'_X \cup v$. We will choose two matchings M_X and M_Y, of R'_X and R'_Y respectively, so that we conform to the correct distribution and couple with some probability. We first consider the case where $\lambda = 1$ and all matchings are weighted equally.

Suppose that v is covered by M_Y. Then there is some vertex v', such that $(v, v') \in M_Y$, and the subgraph R''_Y induced by $V(R'_Y) \setminus \{v, v'\}$ is a subgraph of R'_X. Therefore $M_Y \setminus \{(v, v')\}$ is a matching of R'_X, so the number of matchings of R'_Y that cover v is at most $\Delta - 1$ times the number of matchings of R'_X. Now suppose v is not covered by M_Y. Since $V(R'_Y) \setminus \{v\} = V(R'_X)$, the matchings of R'_Y that do not cover v are exactly the matchings of R'_X.

Therefore, with probability at least $\frac{1}{\Delta}$, we can choose the same M_X and M_Y. With the remaining probability, we have to choose matchings of R'_X and R''_Y. Note that $R'_X = R''_Y \cup \{v'\}$. We are now in the same situation as before, with the roles of R'_X and R'_Y replaced with R''_Y and R'_X respectively. If we choose a matching of R'_X that does

not cover v' then we can choose the same matching for R''_Y. If we choose a matching of R'_X that does cover v', then there is some vertex v'' such that $(v',v'') \in M_X$. The subgraph R''_X induced by $V(R'_X) \setminus \{v',v''\}$ is a subgraph of R''_Y, and $R''_Y = R''_X \cup \{v''\}$. We now have two regions with the same properties as R'_X and R'_Y, both of which are smaller than our original regions. We can assign matchings to these smaller regions by recursively using the same process.

We can repeat this process until we reach a point where v is not covered and the process terminates. This gives us a pair of matchings that differ only along an alternating path of edges $(v,v'),(v',v''),\ldots$. The probability that the path terminates at each step is at least $\frac{1}{\Delta}$, and each trial is independent. The length of the path is bounded above by the number of failures before the path terminates. This is a geometrically distributed random variable with parameter $p = \frac{1}{\Delta}$, and so the expected length of the path is at most $\Delta - 1$.

If $\lambda \neq 1$, then different size matchings are given different weights. Let $W(R')$ be the total weight of all matchings of R'. The set of matchings M_Y that do not cover v is identical to the set of matchings M_X of R'_X, and so those matchings have the same weights in both sets. However, among the matchings M_Y that cover v, these may be larger than the matchings M_X of R'_X by 1. We therefore assign weights to the two cases: the total weight of matchings that cover v is at most $\max\{\lambda,1\}(\Delta-1)W(R'_X)$, while the total weight of matchings that do not cover v is exactly $W(R'_X)$. The probability of terminating the alternating path at each step is now at least $(\max\{\lambda,1\}(\Delta-1)+1)^{-1}$, and so the expected length of the path is at most $\max\{\lambda,1\}(\Delta-1)$, as required.

We use $\max\{\lambda,1\}$ in the upper bound on the weight of matchings that cover v because we do not know when the path will end, and so we cannot say whether M_Y will actually be larger than M_X. This is not a problem when $\lambda \geq 1$, because we are calculating an upper bound. However, if $\lambda < 1$ and it turns out that $|M_Y| = |M_X|$, then it is not valid to say that the total weight is bounded above by $\lambda(\Delta-1)W(R'_X)$. □

In principle, Lemma 4.7 may be applied to any graph with bounded degree. However, since the Markov chain $\mathcal{M}_\ell$ is defined specifically on the 2-dimensional torus, Lemma 4.7 is restricted to this type of graph.

Note that Lemma 4.7 does not give us a direct algorithm for obtaining M_X or M_Y from the appropriate distributions. During the proof of the lemma, we have a lower bound on the proportion of matchings that do not cover v at each step, but we do not know the proportion that includes each edge incident to v. We can therefore determine an upper bound on $|M_X \oplus M_Y|$ in our coupling, but we do not give an efficient algorithm

for finding either M_X or M_Y. Van den Berg and Brouwer used the Jerrum-Sinclair chain to efficiently obtain an almost-uniformly distributed matching to replace the contents of a region. We will assume that some method for implementing large block moves is available when we bound the mixing time of $\mathcal{M}_\ell$ in Lemma 4.9. In Theorem 4.10 we will use our bound on the mixing time of $\mathcal{M}_\ell$ to bound the mixing time of the Jerrum-Sinclair chain, which can be implemented efficiently.

Corollary 4.8. *The expected change in distance of $\mathcal{M}_\ell$ on the 2-dimensional torus is given by*

$$\mathbb{E}[d(X_{t+1},Y_{t+1}) - d(X_t,Y_t)] \leq \frac{12\ell\max\{\lambda,1\} - \ell^2 + \ell}{n}.$$

We therefore require $\ell \geq 12\max\{\lambda,1\}+1$ in order to ensure rapid mixing.

Proof. Combine Equation (4.6) and the result given in Lemma 4.7, noting that in the square lattice, $\Delta = 4$. □

We will now show that the mixing time of $\mathcal{M}_\ell$ is $O(n\log n)$ on the 2-dimensional torus in the case $\ell \geq 12\max\{\lambda,1\}+1$.

Lemma 4.9. *The mixing time of $\mathcal{M}_\ell$ satisfies*

$$\tau(\varepsilon) \leq \frac{n\log(n\varepsilon^{-1})}{\ell^2 - 12\ell\max\{\lambda,1\} - \ell}.$$

Proof. We will apply Lemma 2.14 and Theorem 2.15 to obtain a bound on the mixing time. This requires us to find a $\beta < 1$ such that $\mathbb{E}[d(X_{t+1},Y_{t+1})] \leq \beta d(X_t,Y_t)$. By Corollary 4.8, we know that we can achieve the contraction for any $0 < \beta < 1$, since we can always find a block size ℓ large enough to satisfy this requirement. We have $d(X_t,Y_t) = 1$, so

$$\begin{aligned}\mathbb{E}[d(X_{t+1},Y_{t+1})] &= 1 + \frac{12\ell\max\{\lambda,1\} - \ell^2 + \ell}{n} \\ &= d(X_t,Y_t)\left(1 + \frac{12\ell\max\{\lambda,1\} - \ell^2 + \ell}{n}\right).\end{aligned}$$

Therefore we satisfy the criterion that $\mathbb{E}[d(X_{t+1},Y_{t+1})] \leq \beta d(X_t,Y_t)$ for any $\beta > 0$ such that

$$\beta \leq 1 + \frac{12\ell\max\{\lambda,1\} - \ell^2 + \ell}{n},$$

which is equivalent to

$$1 - \beta \geq -\frac{12\ell\max\{\lambda,1\} - \ell^2 + \ell}{n}. \tag{4.7}$$

Applying Theorem 2.15, we obtain the following bound on the mixing time, for any β satisfying Equation (4.7):

$$\begin{aligned}\tau(\varepsilon) &\leq \frac{\log(D\varepsilon^{-1})}{1-\beta} \\ &\leq \frac{n\log(n\varepsilon^{-1})}{\ell^2 - 12\ell\max\{\lambda,1\} - \ell}.\end{aligned}$$

This is $O(n\log n)$ provided $\ell > 12\max\{\lambda,1\}+1$ (since we require $\beta < 1$).

□

We can now build on this result to obtain a mixing time for the Jerrum-Sinclair chain we defined in Section 2.3.5, by using the comparison technique.

Theorem 4.10. *The mixing time of the lazy form of the Jerrum-Sinclair chain on the 2-dimensional torus with n vertices and $m = 2n$ edges satisfies*

$$\tau(\varepsilon) = O(nm) = O(n^2).$$

Proof. Let $\mathcal{M}$ be the lazy form of the Jerrum-Sinclair chain, and let P be its transition matrix. Let $\mathcal{M}'$ our chain using block moves, with some ℓ such that $\mathcal{M}'$ is rapidly mixing, and let P' be its transition matrix.

Let $\mathcal{P}$ be the set of all paths γ using transitions of the Jerrum-Sinclair chain, between pairs of adjacent states of $\mathcal{M}'$. For any transition (z,w) in $\mathcal{M}$, and any path $\gamma \in \mathcal{P}$, let $r((z,w),\gamma)$ be the number of times that the transition (z,w) occurs on the path γ. Recall the definition of congestion for the comparison method from Section 2.3.6:

$$A_{z,w}(f) = \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} r((z,w),\gamma)|\gamma|f(\gamma).$$

We consider the same set of canonical paths $\Gamma = \{\gamma_{xy} \mid x,y \in \Omega\}$ originally used in Jerrum and Sinclair's argument for showing rapid mixing of their chain and presented in Theorem 2.26 of this thesis. We define an $(\mathcal{M},\mathcal{M}')$-flow f as follows:

$$\begin{aligned} f(\gamma) &= \pi(x)P'(x,y) && \text{if } \exists x,y\in\Omega \text{ such that } \gamma = \gamma_{xy} \\ f(\gamma) &= 0 && \text{if there is no } x,y\in\Omega \text{ for which } \gamma = \gamma_{xy}. \end{aligned}$$

Note that f satisfies Condition (2.8), and so is a valid $(\mathcal{M},\mathcal{M}')$-flow. f is not necessarily an odd flow, but since we use the lazy form of the chain, we know that all eigenvalues are positive, and so this is not a problem.

We now wish to compute the congestion $A_{z,w}(f)$. Each transition can appear at most once in any canonical path, so $r((z,w),\gamma) = 1$ if (z,w) appears on γ, or 0 otherwise. The length of each path is at most $2\ell(\ell-1)$ - that is, the number of edges contained within a block.

Our aim is to rewrite the bound on the congestion for a transition (z,w) such that it is a sum over regions containing $z \oplus w$. We can then use the existing bound on the congestion from Theorem 2.26 to obtain the congestion for each of these regions.

For a matching x, define $\pi_R(x)$ to be the probability of x conditioned on its configuration outside the region R, and $\pi_{V\setminus R}(x)$ the probability of selecting a state with the same configuration as x outside the region R. $\Pr(R)$ denotes the probability of selecting the region R, so $\Pr(R) = \frac{1}{n}$. Note that

$$\pi(x) = \pi_R(x)\pi_{V\setminus R}(x), \qquad \text{and} \tag{4.8}$$

$$P'(x,y) = \sum_{R: x\oplus y \subseteq R} \Pr(R)\pi_R(y). \tag{4.9}$$

Now we can compute the congestion:

$$\begin{aligned}
A_{z,w}(f) &= \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} r((z,w),\gamma)|\gamma|f(\gamma) \\
&= \frac{1}{\pi(z)P(z,w)} \sum_{x,y:(z,w)\in\gamma_{xy}} \pi(x)P'(x,y)|\gamma_{xy}| \\
&= \frac{1}{\pi(z)P(z,w)} \sum_{x,y:(z,w)\in\gamma_{xy}} \sum_{R:x\oplus y\subseteq R} \pi_R(x)\pi_{V\setminus R}(x)\Pr(R)\pi_R(y)|\gamma_{xy}| \\
&= \sum_{R:z\oplus w\subseteq R} \frac{1}{\pi_R(z)\pi_{V\setminus R}(z)P(z,w)}\Pr(R) \sum_{x,y:(z,w)\in\gamma_{xy},x\oplus y\subseteq R} \pi_{V\setminus R}(x)\pi_R(x)\pi_R(y)|\gamma_{xy}| \\
&= \sum_{R:z\oplus w\subseteq R} \Pr(R)\frac{1}{\pi_R(z)P(z,w)} \sum_{x,y:(z,w)\in\gamma_{xy},x\oplus y\subseteq R} \pi_R(x)\pi_R(y)|\gamma_{xy}|.
\end{aligned}$$

In the last line, we can cancel $\pi_{V\setminus R}(z)$ and $\pi_{V\setminus R}(x)$, based on the following observation: if $(z,w)\in\gamma_{xy}$, then z agrees with x on all edges outside of $x\oplus y$. If, additionally, $x\oplus y\subseteq R$, then z clearly agrees with x on all edges outside of R. Therefore, for every choice of z, w, x, y and R such that the conditions of the two sums are satisfied, $\pi_{V\setminus R}(z)$ = $\pi_{V\setminus R}(x)$.

This is a sum over regions containing z and w of the congestion defined in Definition 2.23, as required. We can therefore apply the result from Theorem 2.26, restricted

to the region R, to bound the congestion for each region:

$$\begin{aligned} A_{z,w}(f) &\leq \sum_{R: z \oplus w \subseteq R} \Pr(R) \left[n\bar{\lambda}^2 \sum_{x,y:(z,w) \in \gamma_{xy}, x \oplus y \subseteq R} \pi_R(\eta_t(x,y)) 2\ell(\ell-1) \right] \\ &\leq \sum_{R: z \oplus w \subseteq R} \Pr(R) 2n\ell(\ell-1)\bar{\lambda}^2 \\ &\leq 2\ell^2(\ell-1)^2\bar{\lambda}^2. \end{aligned}$$

This inequality applies uniformly for all choices of z and w, so the overall congestion is

$$A(f) \leq 2\ell^2(\ell-1)^2\bar{\lambda}^2.$$

By Theorem 2.28, the mixing time of the Jerrum-Sinclair chain satisfies

$$\begin{aligned} \tau_x(\mathcal{M}, \varepsilon) &\leq A(f) \left[\frac{\tau(\mathcal{M}', \delta)}{\log(1/2\delta)} + 1 \right] \log \frac{1}{\varepsilon \pi(x)} \\ &\leq 2\ell^2(\ell-1)^2\bar{\lambda}^2 \left[\frac{\tau(\mathcal{M}', \frac{1}{n})}{\log(n/2)} + 1 \right] \left(\log \frac{1}{\varepsilon} + m \right) \\ &\leq 2\ell^2(\ell-1)^2\bar{\lambda}^2 \left[\frac{n \log(n^2)}{(\ell^2 - 12\ell\bar{\lambda} - \ell)\log(n/2)} + 1 \right] \left(\log \frac{1}{\varepsilon} + m \right) \\ &= O(nm). \end{aligned}$$

□

While we focus here on the 2-dimensional torus, this technique can be applied to any graph where the number of edges contained within a ball of radius r is greater than the number of edges crossing its perimeter by at least a factor of r. In Chapter 5 we will use a similar argument for sampling independent sets in claw-free graphs.

Chapter 5

Sampling independent sets in claw-free graphs

In Chapter 4 we used the bijection between matchings in a graph G and independent sets in the line graph $\mathcal{L}(G)$ to adapt the Dyer-Greenhill chain for sampling independent sets to sample matchings. In fact, line graphs belong to a more general class of graphs known as claw-free graphs. Independent sets in claw-free graphs share some of the properties of matchings in general graphs, suggesting that we may be able to adapt Markov chains for sampling matchings to sample independent sets in claw-free graphs.

In Section 5.1 we look at the problem of sampling independent sets in claw-free graphs. We define a Markov chain $\mathcal{M}_{CF}$ for sampling independent sets. Our chain uses the same transitions as Dyer and Greenhill's chain for sampling independent sets [14], but with probabilities based on those of the Jerrum-Sinclair chain for sampling matchings which we introduced in Section 2.3.5. Lemma 5.2 will show that the symmetric difference of two independent sets in a claw-free graph forms a set of alternating paths and cycles. Using this fact, we use a canonical paths argument to show that $\mathcal{M}_{CF}$ is rapidly mixing.

In Section 5.2, we construct a family of chains for sampling independent sets using block updates, and show that these chains mix in $O(n \log n)$ time on claw-free lattice graphs. We will also show that $\mathcal{M}_{CF}$ mixes in $O(n^2)$ time on such lattices, using the same approach that we used in Chapter 4 to show that $\mathcal{M}_{JS}$ mixes in $O(nm)$ time on the square lattice. The square lattice we used in Chapter 4 is not claw-free. However, the set of claw-free graphs does include the triangular and kagome lattices.

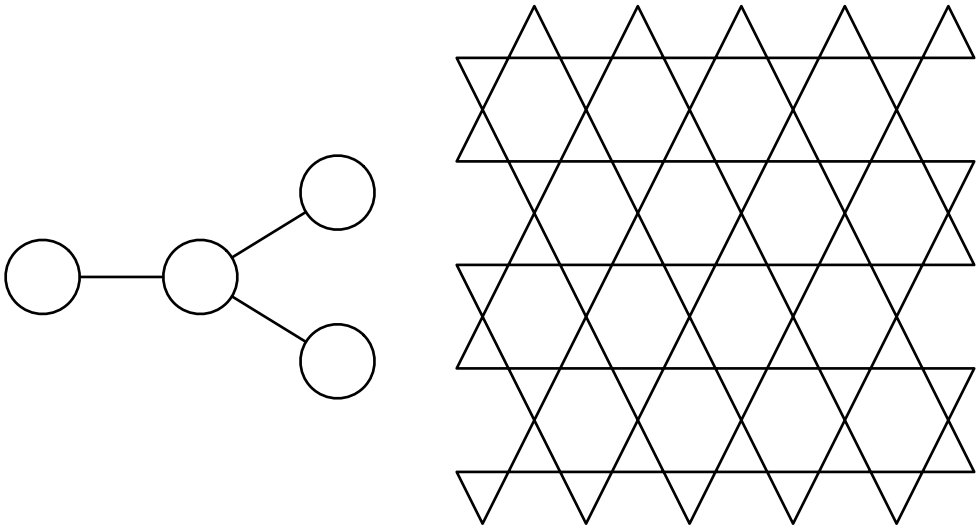

Figure 5.1: *Left, a claw; right, a section of the claw-free kagome lattice.*

5.1 Independent sets in claw-free graphs

Definition 5.1. *The structure G_C at the left of Figure 5.1 is referred to as a claw. A graph $G = (V,E)$ is called a* claw-free *graph if it does not contain G_C as a vertex-induced subgraph.*

Formally, G is claw-free if there is no set of four vertices $\{v, v_1, v_2, v_3\} \subseteq V$ such that

$$\{(v,v_1),(v,v_2),(v,v_3)\} \subseteq E, \qquad \text{and}$$
$$\{(v_1,v_2),(v_2,v_3),(v_3,v_1)\} \cap E = \emptyset.$$

We used the bijection between matchings in a graph G and independent sets in the line graph $\mathcal{L}(G)$ in Section 4.1.1. Line graphs are a special case of claw-free graphs: if a line graph $\mathcal{L}(G)$ contained a claw, this would imply that there was an edge incident to three other edges in G, no two of which share a common endpoint. Since each edge has only two endpoints, this is impossible. Other claw-free graphs include the triangular lattice and kagome lattice (shown in Figure 5.1) - in fact, the kagome lattice is the line graph of the hexagonal lattice. We now show that the symmetric difference of two independent sets in a claw-free graph is a collection of paths and cycles, as is the case for the symmetric difference of two matchings in general graphs.

Lemma 5.2. *For any pair of independent sets I and I' in a claw-free graph G, the*

subgraph of G induced by the vertices of $I \oplus I'$ is a collection of alternating paths and even-length cycles.

Proof. We first note that an odd-length cycle would have to contain two adjacent vertices either from I or from I', contradicting the fact that I and I' are independent sets. We can therefore immediately rule out the possibility that $I \oplus I'$ contains an odd-length cycle.

Suppose that we have a component that is neither a path nor a cycle. Then there must be some vertex $v \in I \oplus I'$ such that the degree of v in the subgraph of G induced by $I \oplus I'$ is at least 3. Without loss of generality, say that $v \in I$. Therefore, there are at least three vertices $v_1, v_2, v_3 \in I'$ adjacent to v. Since I' is an independent set, none of the edges (v_1, v_2), (v_2, v_3) and (v_1, v_3) can be present in G, and therefore $\{v, v_1, v_2, v_3\}$ is a claw. Hence, no such v can exist. □

We now define the Markov chain $\mathcal{M}_{CF}$. The state space Ω is the set of all independent sets in G. The stationary distribution of $\mathcal{M}_{CF}$ is that defined in Equation (2.3),

$$\pi(I) = \frac{\lambda^{|I|}}{\sum_{I' \in \Omega} \lambda^{|I'|}}.$$

As with matchings, the value $Z = \sum_{I' \in \Omega} \lambda^{|I'|}$ is known as the partition function.

Definition 5.3. *$\mathcal{M}_{CF}$ is a Markov chain for sampling independent sets. If the state of $\mathcal{M}_{CF}$ at time t is X_t, then the state at time $t+1$ is determined by the following sequence of steps:*

(CF1) Select a vertex $v \in V$ uniformly at random.

(CF2) If $v \in X_t$, then let $X_{t+1} = X_t \setminus \{v\}$ with probability $\min\{1, \lambda^{-1}\}$.

(CF3) If $v \notin X_t$, and there is no $v' \in X_t$ adjacent to v, then let $X_{t+1} = X_t \cup \{v\}$ with probability $\min\{1, \lambda\}$.

(CF4) If $v \notin X_t$, and there is exactly one $v' \in X_t$ adjacent to v, then let $X_{t+1} = (X_t \cup \{v\}) \setminus \{v'\}$ with probability 1. We call this a drag move.

(CF5) In all other cases, let $X_{t+1} = X_t$.

We will now show that the distribution in Equation (2.3) is a stationary distribution of $\mathcal{M}_{CF}$.

Lemma 5.4. *Let $G = (V,E)$ be a claw-free graph. The distribution*

$$\pi(I) = \frac{\lambda^{|I|}}{\sum_{I' \in \Omega} \lambda^{|I'|}}.$$

is a stationary distribution of $\mathcal{M}_{CF}$.

Proof. We will show that for all pairs of independent sets I_0 and I_1, the detailed balance condition, Condition (2.2), holds. We first consider the case where I_0 and I_1 differ at a single vertex.

Let I_0 and I_1 be two independent sets such that there exists a vertex $v \in V$ such that $I_1 = I_0 \cup \{v\}$. Observe that $\pi(I_1) = \lambda\pi(I_0)$.

$$\begin{aligned}
\pi(I_0)P(I_0,I_1) &= \pi(I_0)\min\{1,\lambda\} \\
&= \pi(I_0)\lambda\min\{1,\lambda^{-1}\} \\
&= \pi(I_1)\min\{1,\lambda^{-1}\} \\
&= \pi(I_1)P(I_1,I_0),
\end{aligned}$$

as required. The situation where, for some vertex v, $I_0 = I_1 \cup \{v\}$, is analogous.

Now suppose that there exists an edge $(v,v') \in E$ such that $I_1 = (I_0 \cup \{v\}) \setminus \{v'\}$. Then $\pi(I_1) = \pi(I_0)$, and $P(I_0,I_1) = P(I_1,I_0) = 1$, and so Condition (2.2) holds.

If $I_0 = I_1$, then $\pi(I_0) = \pi(I_1)$ and $P(I_0,I_1) = P(I_1,I_0)$, and so Condition (2.2) holds trivially.

In all other cases, there is no possible transition between I_0 and I_1, so $P(I_0,I_1) = P(I_1,I_0) = 0$, and so Condition (2.2) holds trivially. Therefore, $\pi(I)$ is a stationary distribution of $\mathcal{M}_{CF}$. □

We can see that $\mathcal{M}_{CF}$ is irreducible by observing that any independent set is reachable from the empty set by a series of insert moves, and that the empty set is reachable from any independent set by a series of delete moves. However, $\mathcal{M}_{CF}$ is not necessarily aperiodic, and so we cannot show that it is ergodic. Consider the graph consisting of a single vertex v, and set $\lambda = 1$. $\mathcal{M}_{CF}$ clearly oscillates between $\{v\}$ and the empty set with period 2. We will avoid the need for $\mathcal{M}_{CF}$ to be aperiodic by introducing a uniform self-loop probability, forming a lazy version of the chain, $\mathcal{M}'_{CF}$. The transition matrix of $\mathcal{M}'_{CF}$ is given by

$$P' = \frac{1}{2}P + \frac{1}{2}I,$$

where P is the transition matrix of $\mathcal{M}_{CF}$ and I is the identity matrix. This self-loop probability means that $\mathcal{M}'_{CF}$ is aperiodic, and therefore ergodic.

We will show that the mixing time of $\mathcal{M}'_{CF}$ is $O(\Delta n^3)$. Lemma 5.2 allows us to construct a set of canonical paths between independent sets of a claw-free graph, in a similar manner to those constructed in the proof of Theorem 2.26. We now state the main result of Section 5.1.

Theorem 5.5. *The mixing time of $\mathcal{M}'_{CF}$ on claw-free graphs is bounded by*

$$\tau(\varepsilon) \leq 2(\Delta-1)n^2\bar{\lambda}^2(2\log\varepsilon^{-1} + n(\log 2 + |\log\lambda|)),$$

where $\bar{\lambda} = \max\{1,\lambda\}$.

The proof of Theorem 5.5 will closely follow Jerrum and Sinclair's proof of an upper bound for sampling matchings, using canonical paths to bound congestion [24]. We gave an overview of this proof in Section 2.3.5. We first need to define a set of canonical paths Γ, and then show that the congestion of $\mathcal{M}'_{CF}$ with respect to Γ is small.

Given two states, I and F, we want to define a canonical path γ_{IF}. Consider the symmetric difference $I \oplus F$. We know from Lemma 5.2 that $I \oplus F$ induces a set of alternating paths and cycles. If we have some arbitrary fixed ordering on V, then we can impose an order on the components of $I \oplus F$, based on the smallest vertex in each component. We will construct a canonical path from I to F by processing the components in this order. For each component C, there are two possible cases:

- If C is an alternating path, then we begin at the smaller endpoint of C, v. If $v \in I$, then we remove v. We then perform a series of drag moves along the length of the path. Finally, if the larger endpoint of C, v', is in F, then we insert v'.
- If C is an even-length alternating cycle, we begin at the smallest vertex that is present in I, v_0 (this is not necessarily the smallest vertex in the cycle). We remove this vertex, and proceed around the cycle with a series of drag moves, starting in the direction of the larger neighbour of v_0. Finally, we insert the smaller neighbour of v_0.

We can now define an encoding $\eta_t(I,F)$ for each transition $t = J \to J'$. If we know t and $\eta_t(I,F)$ (and one additional piece of information which we will describe later), we will be able to recover I and F. This will form the basis of our bound on the congestion of $\mathcal{M}'_{CF}$. If t is a transition on the canonical path from I to F, then we consider the component C that the vertices affected by t belong to.

- If C is a path, then let $\eta_t(I,F) = I \oplus F \oplus (J \cup J')$.
- If C is a cycle, then let $\eta_t(I,F) = (I \oplus F \oplus (J \cup J')) \setminus \{v_0\}$.

Figure 5.2 shows an example of the encoding $\eta_t(I,F)$ for a transition made while processing an 8-cycle in $I \oplus F$. v_0 is marked on Figure 5.2, and the vertices of the cycle are processed in clockwise order from this vertex. J is the intermediate point in the path after v_0 has been removed and two drag moves have been performed. Observe that $J \cup J'$ agrees with F on those vertices that have already been processed, with I on those that have not yet been processed, and includes both vertices that are affected by the transition $J \to J'$. Since the first transition while processing a cycle is to remove v_0, we can see that $J \cup J'$ does not include v_0 or its smaller neighbour (which has not yet been processed and does not belong to I).

$\eta_t(I,F)$ therefore agrees with I on those vertices that have already been processed, with F on those that have not yet been processed, and does not include the vertices that are affected by $J \to J'$. It includes the smaller neighbour of v_0, but not v_0 itself because this is explicitly excluded from η_t in the case of a cycle.

We now show that the sum of $\pi(\eta_t)$ over all paths using any transition t is not too large.

Lemma 5.6. *Using the canonical paths defined above, let* $\mathrm{cp}(t)$ *be the set of paths using transition* t*. Then*

$$\sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \leq \Delta - 1.$$

Proof. There are two parts to the proof. First, we show that $\eta_t(I,F)$ is an independent set. Secondly, we show that if we are given t, $\eta_t(I,F)$ and v_0 (if it exists), then it is possible to determine the states I and F.

Assume that $\eta_t(I,F)$ is not an independent set. Then there are two vertices v and v' in $\eta_t(I,F)$ that are adjacent. Therefore one of v and v' must belong to I and the other to F, and so both belong to $I \oplus F$. Therefore, neither v nor v' belongs to $J \cup J'$. If we consider $J \cup J'$ for each transition $J \to J'$ that occurs during the processing of a component C, we see that the only time there can be two adjacent vertices in $C \setminus (J \cup J')$ is when C is a cycle, and one of the vertices is v_0. However, by definition it is not possible for v_0 to belong to $\eta_t(I,F)$, and so no such v, v' can exist. Hence, $\eta_t(I,F)$ is an independent set.

Clearly from $\eta_t(I,F)$, J and J', it is possible to reconstruct $I \oplus F$ if C is a path, or $(I \oplus F) \setminus \{v_0\}$ if C is a cycle. If C is a cycle, then we also need to identify v_0 to obtain

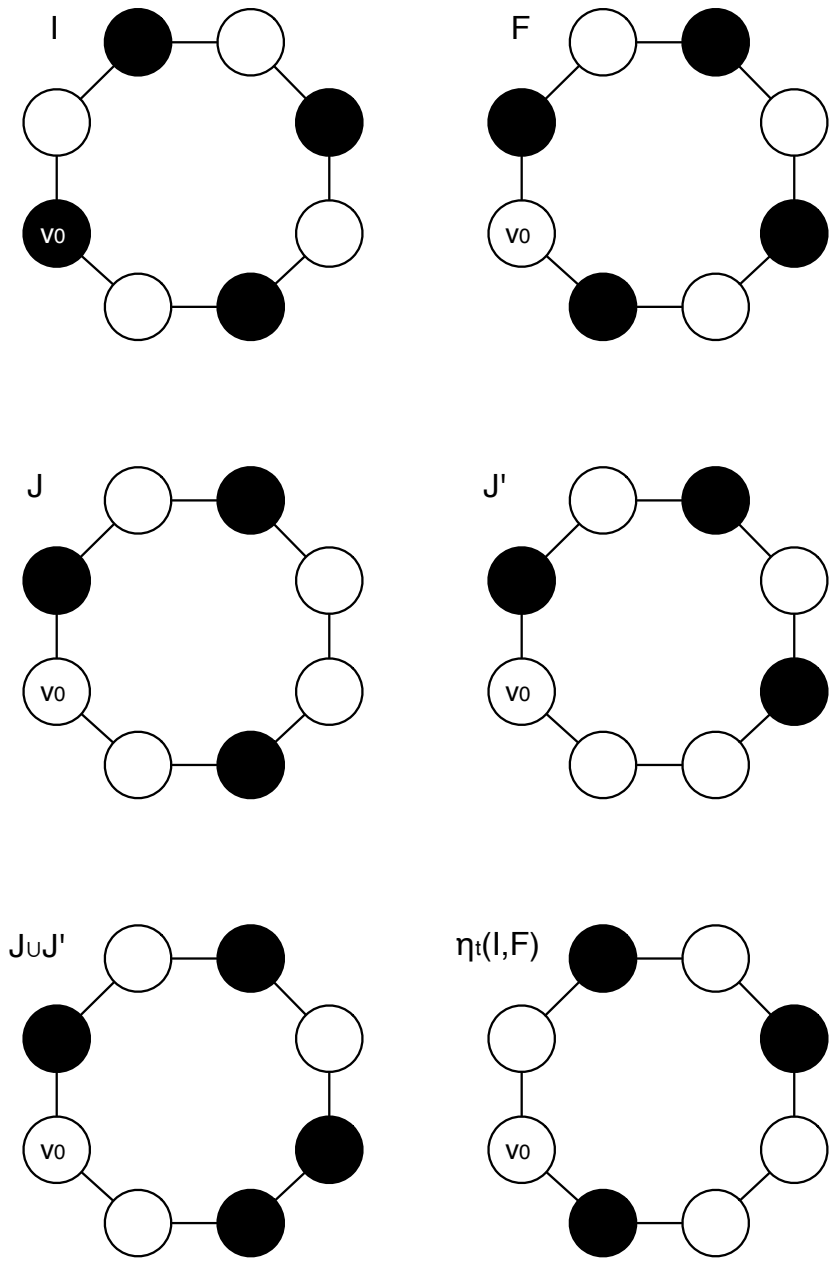

Figure 5.2: *The encoding* $\eta_t(I, F)$ *for a transition made while processing an 8-cycle.*

$I \oplus F$. When sampling matchings, there is a similar problem: we can recover all but one *edge* of a cycle. In the case of matchings, there is a unique edge that can complete the cycle [25]. However, for independent sets, there are up to $\Delta - 1$ possible choices of v_0.

To determine whether C is a path or a cycle, we look at $\eta_t(I,F) \oplus (J \cup J')$. This will provide an alternating path on which t is a transition. If t is consistent with a canonical path moving away from the smaller endpoint, then C is a path. If t is consistent with a path moving towards the smaller endpoint, then C is a cycle and some v_0 exists to complete C.

Therefore, for each independent set J^*, there are at most $\Delta - 1$ pairs of states (I,F) such that $\eta_t(I,F) = J^*$. Since $\sum_{J^* \in \Omega} \pi(J^*) = 1$, the required bound follows. □

We now have a set of canonical paths and a suitable encoding η_t, and can proceed with the proof of Theorem 5.5.

Proof of Theorem 5.5. To bound the mixing time using Theorem 2.24, we also need to obtain a bound on the value of $\pi(I)\pi(F)$ for the pairs of states (I,F) passing through each transition $t = (J,J')$. Consider $\lambda^{|I|}\lambda^{|F|}$ and $\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}$. Each vertex $v \in V$ contributes a factor of 1, λ, or λ^2 to each of these. There are three mutually exclusive cases:

- $v \in I \cap F$. v contributes λ^2 to $\lambda^{|I|}\lambda^{|F|}$. Since v is never removed during the processing of γ_{IF}, it must be in both J and J'. Furthermore, $v \notin I \oplus F$, and so $v \in \eta_t(I,F)$. Hence, the contribution to $\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}$ is also λ^2.

- $v \notin I \cup F$. In this case, we can easily see that $v \notin I$, $v \notin F$, $v \notin J \cup J'$, and so $v \notin \eta_t(I,F)$. Therefore the contribution to both expressions is 1.

- $v \in I \oplus F$. v is in exactly one of I and F, and so the contribution to $\lambda^{|I|}\lambda^{|F|}$ is λ. Assuming $v \neq v_0$, we can see from the definition of $\eta_t(I,F)$ that $v \in \eta_t(I,F)$ if and only if $v \notin J \cup J'$. Therefore the contribution to $\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}$ is λ. If $v = v_0$, then $v \notin \eta_t(I,F)$, even if $v \notin J \cup J'$. Therefore the contribution to $\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}$ may be 1 in this case.

The contribution to each expression is the same for every vertex, except for v_0 (where it exists). $\lambda^{|I|}\lambda^{|F|}$ may therefore be greater than $\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}$ by at most a factor of $\bar{\lambda} = \max\{1,\lambda\}$. Hence, for any pair of independent sets (I,F) passing through $t = (J,J')$, we have

$$\lambda^{|I|}\lambda^{|F|} \leq \bar{\lambda}\lambda^{|J \cup J'|}\lambda^{|\eta_t(I,F)|}. \tag{5.1}$$

Observing that $|J|, |J'| \geq |J \cup J'| - 1$ and dividing by Z^2, we get

$$\begin{aligned}\pi(I)\pi(F) &\leq \bar{\lambda}^2 \min\{\pi(J), \pi(J')\}\pi(\eta_t(I,F)) \\ &= n\bar{\lambda}^2 \pi(J)P(J,J')\pi(\eta_t(I,F)). \end{aligned} \tag{5.2}$$

We are now in a position to compute the congestion, and therefore the mixing time. The congestion is

$$\begin{aligned}\rho &= \max_{t=(J,J')} \left\{ \frac{1}{\pi(J)P(J,J')} \sum_{(I,F)\in \mathrm{cp}(t)} \pi(I)\pi(F)|\gamma_{IF}| \right\} \\ &\leq n\bar{\lambda}^2 \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F))|\gamma_{IF}| && \text{by Equation (5.2)} \\ &\leq n^2\bar{\lambda}^2 \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \\ &\leq (\Delta-1)n^2\bar{\lambda}^2 && \text{by Lemma 5.6.}\end{aligned}$$

Applying Theorem 2.24, we obtain the bound on the mixing time

$$\tau_x(\varepsilon) \leq 2(\Delta-1)n^2\bar{\lambda}^2(2\log\varepsilon^{-1} + \log\pi(x)^{-1}),$$

for a given start state x. Finally, we need to bound $\log\pi(x)^{-1}$ for any $x \in \Omega$. Observe that there are at most 2^n possible independent sets, and suppose that $\lambda > 1$. Then

$$\begin{aligned}\log\pi(x)^{-1} &= \log\left(\frac{\sum_S \lambda^{|S|}}{\lambda^{|x|}}\right) \\ &\leq \log(2^n\lambda^n) \\ &= n(\log 2 + |\log\lambda|).\end{aligned}$$

Now suppose $\lambda < 1$.

$$\begin{aligned}\log\pi(x)^{-1} &\leq \log\left(\frac{2^n}{\lambda^n}\right) \\ &= n(\log 2 - \log\lambda) \\ &= n(\log 2 + |\log\lambda|).\end{aligned}$$

We therefore obtain the bound on the mixing time

$$\tau(\varepsilon) \leq 2(\Delta-1)n^2\bar{\lambda}^2(2\log\varepsilon^{-1} + n(\log 2 + |\log\lambda|)),$$

as required. □

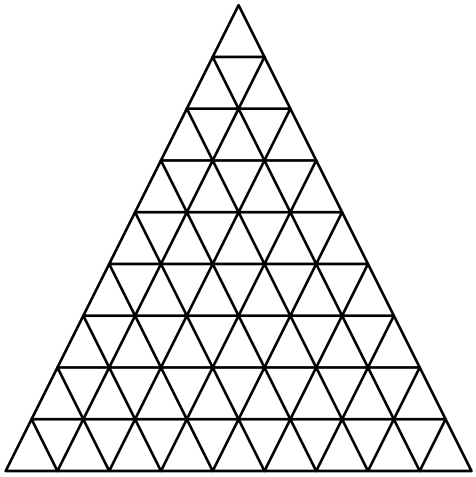

Figure 5.3: *A section of the 2-dimensional triangular lattice.*

Note that we can improve on the bound on the mixing time given by Theorem 5.5 if the graph G contains no 4-cycles. In the proof of Lemma 5.6, we show that at most $\Delta - 1$ pairs of states (I, F) exist that can map to a given value of η_t. This bound arises because, given a cycle with one vertex removed, there are up to $\Delta - 1$ possible ways to add a vertex v_0 that completes a cycle. If G contains no 4-cycles, then there is a unique choice of v_0. It follows that our bound on the mixing time of $\mathcal{M}'_{CF}$ can be improved by a factor of $\Delta - 1$ in this case. The kagome lattice is an example of a claw-free graph that contains no 4-cycles.

5.2 Claw-free lattices

We now consider the problem of sampling independent sets in claw-free graphs with a lattice structure. While the square lattice is not claw-free, the 2-dimensional triangular lattice (Figure 5.3) and the kagome lattice are. We will use a Markov chain that updates large (but constant size) blocks of vertices in a single step. We will show that for any Δ and λ, there exists a chain using block moves that mixes in time $O(n \log n)$. We will use comparison techniques to show that the mixing time of $\mathcal{M}_{CF}$ is $O(n^2)$ on claw-free lattices, an improvement on the bound of $O(n^2 \log(n))$ for $\mathcal{M}_{CF}$ that we obtained in Section 5.1.

As before, we sample from the set of all independent sets of a graph G, according

to the probability distribution $\pi(I)$:

$$\pi(I) = \frac{\lambda^{|I|}}{\sum_{I' \in \Omega} \lambda^{|I'|}}.$$

The critical property of lattice graphs that we require is as follows: for any vertex $v \in V$, the number of vertices at distance ℓ from v (as a function of ℓ) is asymptotically smaller than the number of vertices at distance less than ℓ from v. Formally, for any $v \in V$ and $\ell \in \mathbb{N}$, let R be the ball of radius ℓ around v. The *vertex boundary* of R is defined as

$$\partial R = \{u \in R \mid \exists u' \in V \setminus R \text{ s.t. } (u,u') \in E\}.$$

In the case of a lattice of fixed size, note that if we choose v close to the boundary of the lattice, the points at distance ℓ from v will not necessarily belong to ∂R if they lie on the boundary of the lattice.

Definition 5.7. *Let $G = (V,E)$ be a graph. For any vertex v, let R be the ball of radius ℓ around v, and let ∂R be the vertex boundary of R. If, for all vertices v,*

$$\frac{|R|}{|\partial R|} \to \infty \text{ as } \ell \to \infty, \tag{5.3}$$

then we say that G satisfies the boundary condition *that is required to show rapid mixing of G using Lemma 5.9 and Theorem 5.10.*

Condition (5.3) applies for all lattice graphs. Our results are applicable to non-lattice graphs provided the boundary condition holds, but in practice it is easier to demonstrate that this is the case for lattices. The boundary condition will not be required in the proof of Lemma 5.9 and Theorem 5.10, but it will be necessary when we come to apply these lemmas to specific graphs.

Definition 5.8. *Set a constant radius ℓ. We define a Markov chain $\mathcal{M}'_\ell$ for sampling independent sets. Let X_t be the state of $\mathcal{M}'_\ell$ at time t. The state at time $t+1$, X_{t+1} is determined by the following sequence of steps:*

($\ell 1'$) Select a vertex v uniformly at random. Let R be the ball of radius ℓ centred on v, and ∂R be the vertex boundary of R. Let R' be the subgraph of G induced by the vertices $R \setminus (\partial R \cup \{u \mid \exists u' \in \partial R \cap X_t \text{ s.t. } (u,u') \in E\})$, where X_t is the current state of the chain. That is, R' contains the vertices of R that are not on the vertex boundary, and are not adjacent to any member of X_t that is on the vertex boundary.

($\ell 2'$) Let $\Omega_{R'}$ be the set of independent sets of R', and $\pi_{R'}(I)$ the weighted distribution on independent sets on the graph induced by R'. Select an independent set $I_{R'}$ of R' according to the distribution $\pi_{R'}$.

($\ell 3'$) Let $X_{t+1} = (X_t \setminus R') \cup I_{R'}$.

Note that unlike the chain we defined in Section 4.2 for sampling matchings in lattice graphs, $\mathcal{M}'_\ell$ selects a vertex at random and replaces the ball around it, instead of selecting a ball uniformly at random. This means that for the probability of selecting any given vertex is exactly $\frac{1}{n}$.

We can use path coupling to show that for a sufficiently large ℓ (determined by a function of Δ and λ), the chain $\mathcal{M}'_\ell$ is rapidly mixing. Let X_t and Y_t be two copies of $\mathcal{M}'_\ell$. The distance $d(X_t, Y_t)$ is the Hamming distance between the two states, and the adjacency relation is the set of pairs (I, I') of states such that $d(I, I') = 1$. Suppose that at time t, $d(X_t, Y_t) = 1$, with some vertex $u \in X_t$, $u \notin Y_t$. There are three possible cases, depending on the position of u relative to R.

1. If $u \in R \setminus \partial R$, then our coupling will always choose the same $I_{R'}$ in each chain. Therefore we are guaranteed that $d(X_{t+1}, Y_{t+1}) = 0$.

2. If $u \notin R$, then our coupling will always choose the same $I_{R'}$ in each chain, and the distance is unchanged.

3. If $u \in \partial R$, then the regions that we update in step $\ell 2'$ may differ between X_t and Y_t, and in these cases we must choose $I_{R'}$ from different distributions in X_t and Y_t. The distance in these cases will usually increase.

Let R^* be the ball of radius ℓ around u. Observe that there are $|\partial R^*|$ choices of v such that u lies on the boundary of R (so the distance may increase), and there are $|R^* \setminus \partial R^*|$ choices of v such that u lies within R (so the distance will decrease). If δ is a uniform upper bound on the expected increase when u lies on the boundary of R, and δ is independent of $|R|$, then the expected change in distance is

$$\mathbb{E}[d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)] = \frac{\delta|\partial R| - |R \setminus \partial R|}{n}. \tag{5.4}$$

Hence, if $|R^*|$ is sufficiently large relative to its boundary, then the expected change is non-positive and the chain is rapidly mixing. In particular, if $|\partial R^*| = o(|R^* \setminus \partial R^*|)$, then for some sufficiently large ℓ the chain $\mathcal{M}'_\ell$ is rapidly mixing.

For lattice graphs, $|R^*|$ and $|R^* \setminus \partial R^*|$ are fixed values (in terms of ℓ) for all but an asymptotically small number of choices of u. The vertices for which this is not the case are those lying close to the edge of the lattice itself, and we will need to consider these special cases in order to show that the ratio $\frac{|R^*|}{|\partial R^*|}$ has a uniform lower bound for all vertices u.

We will now use path coupling to show that there exists a coupling where the increase in distance in case 3 is bounded above by a geometrically distributed random variable.

Lemma 5.9. *Let G be a claw-free graph with maximum degree Δ, and R be a ball of radius ℓ centred on some vertex v. Let X and Y be two independent sets of G, and let u be a unique vertex on the boundary ∂R such that $u \in X$, $u \notin Y$. There exists a coupling such that the increase in Hamming distance between X and Y when R is updated according to steps $(\ell 1')$ and $(\ell 2')$ of $\mathcal{M}'_\ell$ is stochastically dominated by a geometrically distributed random variable with mean at most $(3\Delta - 1)\bar{\lambda}$, where $\bar{\lambda} = \max\{\lambda, 1\}$.*

Proof. Let $R'_{X,0}$ and $R'_{Y,0}$ be the regions of X and Y to be updated. We will show that it is possible to choose two independent sets I_X and I_Y of $R'_{X,0}$ and $R'_{Y,0}$ respectively, such that the Hamming distance between I_X and I_Y is not too large. Let $V_0 = R'_{Y,0} \setminus R'_{X,0}$ be the set of vertices that may be in I_Y but not I_X (that is, the neighbours of v that are in R). We will select a number of independent vertices in V_0 that will belong to I_Y, and recursively select the remainder of I_Y conditioned on the choice of these vertices. For the first part of this proof, the only assumption we make about the distributions we choose from is that if we choose no vertices from V_0, then the induced distribution on the remainder of $R'_{Y,0}$ is identical to the induced distribution on $R'_{X,0}$. We will consider the weights given to different size independent sets in the distribution $\pi(I)$ when we come to bound the number of levels of recursion required to choose I_X and I_Y.

If we have chosen $I_Y \cap V_0 = \emptyset$, then the distributions of I_X and I_Y, given that $I_Y \cap V_0 = \emptyset$, are identical. Therefore, in this case we can choose the same independent set for I_X and I_Y and the Hamming distance does not increase. We now consider the coupling for sets where $I_Y \cap V_0 \neq \emptyset$, considering I_Y first.

If one or more vertices exist in $I_Y \cap V_0$, then any independent set of $R'_{X,0}$ must have Hamming distance at least $|I_Y \cap V_0|$ from the independent set I_Y. Let $V_1 = R'_{X,0} \cap \Gamma(I_Y \cap V_0)$, where $\Gamma(I_Y \cap V_0)$ is the set of vertices adjacent to those in $I_Y \cap V_0$. Let $R'_{X,1} = R'_{X,0}$ and $R'_{Y,1} = R'_{X,1} \setminus V_1$. Since G is claw-free, $|I_Y \cap V_0| \leq 2$, and therefore $|V_1| \leq 2\Delta - 2$.

We now select the vertices in $I_X \cap V_1$. If no such vertices exist, then the distributions of $I_X \setminus V_0$ and $I_Y \setminus V_0$, conditional on the fact that $I_X \cap V_1 = \emptyset$, are identical, as before. The Hamming distance therefore increases by $|I_Y \cap V_0|$, which is at most 2.

If $I_X \cap V_1$ is non-empty, then we set $V_2 = R'_{Y,1} \cap \Gamma(I_Y \cap V_1)$, $R'_{X,2} = R'_{X,1} \setminus V_2$ and $R'_{Y,2} = R'_{Y,1}$. While there may be two vertices in $I_Y \cap V_0$, we know that each of these already has one neighbour in X, and therefore can have only one neighbour in $V_1 \setminus V_0$. Hence, $|I_X \cap V_1| \leq |I_Y \cap V_0| \leq 2$. Iterating these steps, we can see that if we can bound the expected number of steps t until $(I_X \cup I_Y) \cap V_t = \emptyset$, then the expected increase in Hamming distance is at most $2t$. We will now obtain a bound on the number of steps required, as a function of Δ and λ.

We wish to obtain a lower bound on the probability that $(I_X \cap I_Y) \cap (V_t \setminus V_{t-1}) = \emptyset$. Without loss of generality, assume that t is even. We assign a weight $w(I) = \lambda^{|I|}$ to each independent set. By the definition of the stationary distribution, the probability of any given set I being selected is proportional to $w(I)$. However, the proportion so assigned to independent sets of both $R'_{X,0}$ and $R'_{Y,0}$ will be different in the two distributions. Let $Z(I_X \setminus V_t)$ be the total weight of all independent sets of $I_X \setminus V_t$ (Z is very much like the partition function, but applied only to a subset of the vertices of G). If $I_Y \cap (V_t \setminus V_{t-1}) = \emptyset$, then $Z(I_Y \setminus V_t) = Z(I_X \setminus V_t)$. Otherwise, note that $R'_{Y,t+1} \subseteq R'_{X,t}$, and therefore $Z(I_Y \setminus V_t) \leq \max\{\lambda, 1\}^2 Z(I_X \setminus V_t)$. There are at most $(\Delta - 1)^2$ ways of choosing two vertices from V_t that satisfy the conditions of an independent set. There are also at most $2\Delta - 2$ ways of choosing a single vertex, and one way of choosing none. The total weight of independent sets where $|I_Y \cap V_t| = 2$ is at most

$$(\Delta - 1)^2 \max\{\lambda, 1\}^2 Z(I_X \setminus V_t),$$

the total weight of those where $|S_Y \cap V_t| = 1$ is at most

$$\begin{aligned} &(2\Delta - 2) \max\{\lambda, 1\} Z(I_X \setminus V_t) \\ \leq\ &(2\Delta - 2) \max\{\lambda, 1\}^2 Z(I_X \setminus V_t), \end{aligned}$$

and the total weight of those where $|S_Y \cap V_t| = 0$ is

$$\begin{aligned} &Z(I_X \setminus V_t) \\ \leq\ &\max\{\lambda, 1\}^2 Z(I_X \setminus V_t). \end{aligned}$$

There are two ways we can reach a time t such that $|I_Y \cap V_t| = 0$:

1. $|I_X \cap V_{t-1}| = 2$. The probability that $|I_Y \cap V_t| = 0$, given that $|I_X \cap V_{t-1}| = 2$, is at least
$$\frac{Z(I_X \setminus V_t)}{((\Delta-1)^2 + 2\Delta - 2 + 1)\max\{\lambda,1\}^2 Z(I_X \setminus V_t)} = \frac{1}{\Delta^2 \max\{\lambda,1\}^2}.$$

2. $|I_X \cap V_{t-1}| = 1$. In this case, there is some time s, for which $|I_Y \cap V_s| = 1$ and $|I_Y \cap V_{s-1}| = 2$ (assuming again that s is even - if not then the roles of I_X and I_Y are reversed). The probability that $|I_Y \cap V_s| = 1$, given that $|I_X \cap V_{s-1}| = 2$, is at least
$$\begin{aligned} &\frac{(2\Delta-2)\max\{\lambda,1\}Z(I_X \setminus V_s)}{\Delta^2 \max\{\lambda,1\}^2} \\ &= \frac{2\Delta-2}{\Delta^2 \max\{\lambda,1\}} \\ &\geq \frac{1}{\Delta \max\{\lambda,1\}} \qquad \text{(provided } \Delta \geq 2\text{)}, \end{aligned}$$
and the probability that $|I_Y \cap V_t| = 0$, given that $|I_Y \cap V_{t-1}| = 1$, is at least
$$\frac{Z(I_X \setminus V_t)}{(2\Delta-1)\max\{\lambda,1\}Z(I_X \setminus V_t)} = \frac{1}{(2\Delta-1)\max\{\lambda,1\}}.$$

The time t is therefore stochastically dominated by either a geometrically distributed random variable with parameter at least $\frac{1}{\Delta^2 \max\{\lambda,1\}^2}$, or by the sum of two geometrically distributed random variables with parameters at least $\frac{2}{\Delta \max\{\lambda,1\}}$ and $\frac{1}{(2\Delta-1)\max\{\lambda,1\}}$ respectively, whichever occurs first.

$$\begin{aligned} \mathbb{E}[t] &\leq \min\{\Delta^2 \max\{\lambda,1\}^2, (3\Delta-1)\max\{\lambda,1\}\} \\ &\leq (3\Delta-1)\max\{\lambda,1\} \qquad \text{(again assuming } \Delta \geq 2\text{)}. \end{aligned}$$

□

Now we can bound the ratio β of $d(X_t, Y_t)$ and $\mathbb{E}[d(X_{t_1}, Y_{t+1})]$, and thus apply Theorem 2.15 to bound the mixing time.

Theorem 5.10. *The mixing time of $\mathcal{M}'_\ell$ on claw-free graphs satisfying the boundary condition satisfies*
$$\tau(\varepsilon) \leq \frac{n \log(n\varepsilon^{-1})}{|R \setminus \partial R| - (3\Delta-1)\bar{\lambda}|\partial R|},$$
provided $\Delta \geq 2$ and $|R \setminus \partial R| > (3\Delta-1)\bar{\lambda}|\partial R|$.

Proof. Combining Equation (5.4) and Lemma 5.9, we obtain the following bounds on the expected change in distance and the contraction ratio β:

$$\begin{aligned}\mathbb{E}[d(X_{t+1},Y_{t+1}) - d(X_t,Y_t)] &= \frac{\delta|\partial R| - |R\setminus\partial R|}{n} \\ &\leq \frac{(3\Delta-1)\bar{\lambda}|\partial R| - |R\setminus\partial R|}{n} \\ \Rightarrow \qquad \mathbb{E}[d(X_{t+1},Y_{t+1})] &\leq 1 + \frac{(3\Delta-1)\bar{\lambda}|\partial R| - |R\setminus\partial R|}{n} \quad \text{since } d(X_t,Y_t)=1.\end{aligned}$$

Therefore, in the context of the path coupling lemma, Lemma 2.14, we have the following bound on the contraction ratio β:

$$\begin{aligned}\beta &\leq 1 + \frac{(3\Delta-1)\bar{\lambda}|\partial R| - |R\setminus\partial R|}{n} \\ \Rightarrow \qquad 1-\beta &\geq \frac{|R\setminus\partial R| - (3\Delta-1)\bar{\lambda}|\partial R|}{n}.\end{aligned}$$

Therefore, by Theorem 2.15, the mixing time of $\mathcal{M}'_\ell$ satisfies

$$\begin{aligned}\tau(\varepsilon) &\leq \frac{\log(D\varepsilon^{-1})}{1-\beta} \\ &\leq \frac{\log(n\varepsilon^{-1})(n-|R|)}{|R\setminus\partial R| - (3\Delta-1)\bar{\lambda}|\partial R|}.\end{aligned}$$

If we have $|R\setminus\partial R| > (3\Delta-1)\bar{\lambda}|\partial R|$, then this ensures that the expected distance decreases, and hence that $\beta < 1$ and we can apply Theorem 2.15. □

The bound given by Theorem 5.10 applies for the Markov chain $\mathcal{M}'_\ell$ which performs block moves. We now use the comparison theorem, Theorem 2.28, to show that the mixing time of the single-site chain $\mathcal{M}_{CF}$ is $O(n^2)$ on lattice graphs.

Lemma 5.11. *The mixing time of the lazy form of $\mathcal{M}_{CF}$ on claw-free graphs satisfying the boundary condition satisfies*

$$\tau(\varepsilon) \leq |R\setminus\partial R|^2\bar{\lambda}^2(\Delta-1)\left[\frac{2(n-|R|)\log n}{(|R\setminus\partial R| - (3\Delta-1)\bar{\lambda}|\partial R|)(\log n - \log 2)}\right](\log(1/\varepsilon)+n).$$

Proof. Let $\mathcal{M}$ be the lazy version of $\mathcal{M}_{CF}$, and let $\mathcal{M}' = \mathcal{M}'_\ell$. Let P' be the transition matrix of $\mathcal{M}'$. We use the canonical paths of Section 5.1 to construct a canonical path for every pair of independent sets x, y which are connected by a single transition of $\mathcal{M}'_\ell$. Then we define the $\{\mathcal{M},\mathcal{M}'\}$-flow f as follows:

$$\begin{aligned}f(\gamma) &= \pi(x)P'(x,y) && \text{if } \exists x,y\in\Omega \text{ such that } \gamma = \gamma_{xy} \\ f(\gamma) &= 0 && \text{if there is no } x,y\in\Omega \text{ for which } \gamma = \gamma_{xy}.\end{aligned}$$

Now we can compute the congestion of f. By construction, $r((z,w),\gamma) \leq 1$ for every canonical path γ, and $|\gamma| \leq |R|$. Let $\pi_R(x)$ be the probability of state x conditioned on its configuration outside the region R, and $\pi_{V\setminus R}(x)$ be the probability of *any* state that agrees with x outside of the region R. $\Pr(R)$ denotes the probability that we select the region R in the first step of $\mathcal{M}'_\ell$ (note that since R is selected by choosing a vertex uniformly at random, $\Pr(R) = \frac{1}{n}$). Then the following hold:

$$\pi(x) = \pi_R(x)\pi_{V\setminus R}(x) \tag{5.5}$$

$$P'(x,y) = \sum_{R: x\oplus y \subseteq R} \Pr(R)\pi_R(y) \tag{5.6}$$

Now, the congestion $A_{z,w}(f)$ satisfies

$$\begin{aligned}
A_{z,w}(f) &= \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} r((z,w),\gamma)|\gamma|f(\gamma) \\
&= \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} \pi(x)P'(x,y)|\gamma| \\
&= \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma} \sum_{R: x\oplus y\subseteq R} \pi_R(x)\pi_{V\setminus R}(x)\Pr(R)\pi_R(y)|\gamma| \\
&= \sum_{R:z,w\in R} \frac{1}{\pi_R(z)\pi_{V\setminus R}(z)P(z,w)} \pi_{V\setminus R}(x)\Pr(R) \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma, x\oplus y\subseteq R} \pi_R(x)\pi_R(y)|\gamma| \\
&= \sum_{R:z,w\in R} \Pr(R)\frac{1}{\pi_R(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma, x\oplus y\subseteq R} \pi_R(x)\pi_R(y)|\gamma|.
\end{aligned}$$

Now let us return to the proof of Theorem 5.5. From Equation (5.1), we have

$$\lambda^{|x|}\lambda^{|y|} \leq \bar{\lambda}\lambda^{|z\cup w|}\lambda^{|\eta_t(x,y)|}.$$

If we assume that for a region R, $z,w \in R$, and divide through by $Z\pi_{V\setminus R}(z)$, we get

$$\begin{aligned}
\pi_R(x)\pi_R(y) &\leq \bar{\lambda}^2 \min\{\pi_R(z),\pi_R(w)\}\pi_R(\eta_t(x,y)) \\
&= n\bar{\lambda}^2\pi_R(z)P(z,w)\pi_R(\eta_t(x,y)).
\end{aligned}$$

Therefore

$$\begin{aligned}
A_{z,w}(f) &\leq \sum_{R:z,w\in R} \Pr(R)\frac{1}{\pi_R(z)P(z,w)} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma, x\oplus y\subseteq R} n\bar{\lambda}^2\pi_R(z)P(z,w)\pi_R(\eta_t(x,y))|\gamma| \\
&= \sum_{R:z,w\in R} \Pr(R) \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma, x\oplus y\subseteq R} n\bar{\lambda}^2\pi_R(\eta_t(x,y))|\gamma| \\
&= \sum_{R:z,w\in R} \sum_{\gamma\in\mathcal{P}:(z,w)\in\gamma, x\oplus y\subseteq R} \bar{\lambda}^2\pi_R(\eta_t(x,y))|\gamma| \\
&\leq |R\setminus\partial R|^2\bar{\lambda}^2(\Delta-1).
\end{aligned}$$

This inequality holds for all choices of z and w, so the overall congestion satisfies

$$A(f) \leq |R \setminus \partial R|^2 \bar{\lambda}^2 (\Delta - 1).$$

The mixing time of $\mathcal{M}_{CF}$ therefore satisfies

$$\begin{aligned}
\tau_x(\mathcal{M}_{CF}, \varepsilon) &\leq A(f) \left[\frac{\tau(\mathcal{M}'_\ell, \delta)}{\log(1/2\delta)} + 1 \right] \log\left(\frac{1}{\varepsilon \pi(x)} \right) \\
&\leq |R \setminus \partial R|^2 \bar{\lambda}^2 (\Delta - 1) \left[\frac{\tau(\mathcal{M}_{CF}, 1/n)}{\log(n/2)} + 1 \right] (\log(1/\varepsilon) + n) \\
&\leq |R \setminus \partial R|^2 \bar{\lambda}^2 (\Delta - 1) \left[\frac{2n \log n (\log(1/\varepsilon) + n)}{(|R \setminus \partial R| - (3\Delta - 1)\bar{\lambda} |\partial R|)(\log n - \log 2)} \right] \\
&= O(n^2),
\end{aligned}$$

subject to the restrictions that $|R \setminus \partial R| > (3\Delta - 1)\bar{\lambda}|\partial R|$ and $\Delta \geq 2$. □

5.2.1 Application to the triangular and kagome lattices

We now show that suitable values of ℓ exist for the triangular and kagome lattices.

Proposition 5.12. *$\mathcal{M}'_\ell$ mixes in time $O(n \log n)$ and $\mathcal{M}_{CF}$ mixes in time $O(n^2)$ on the triangular and kagome lattices, provided*

$$\ell \geq \frac{136\bar{\lambda} + 1}{3} \qquad \textit{for the triangular lattice, and}$$

$$\ell \geq \frac{55}{2}\bar{\lambda} + 1 \qquad \textit{for the kagome lattice.}$$

Proof. We begin by showing some properties of the triangular lattice that allow us to find a uniform lower bound on $\frac{|R \setminus \partial R|}{|\partial R|}$.

The triangular lattice is such that for any fixed ℓ, and for all but an asymptotically small number of vertices v, the perimeter and area of the ball of radius ℓ around a vertex are uniform. We can easily bound $\frac{|R \setminus \partial R|}{|\partial R|}$ for these vertices: $|R \setminus \partial R| = 3\ell(\ell - 1) + 1$, and $|\partial R| = 6\ell$. Some difficulty arises when we consider those balls for which the area and perimeter are not uniform. These are the balls around vertices that are at distance less than $\ell + 1$ from the boundary of the graph G.

Suppose we begin with a vertex v at distance $\ell + 1$ from the boundary of G, and consider the change in $|\partial R|$ and $|R \setminus \partial R|$ as we move v towards the boundary. When we move from distance $\ell + 1$ to distance ℓ, $|\partial R|$ decreases, and $|R \setminus \partial R|$ increases. When we move from distance ℓ to distance $\ell - 1$, we remove two vertices from ∂R, and we remove $\ell - 1$ vertices from $R \setminus \partial R$. As we move closer to the boundary, at each step we

remove two vertices from ∂R, and an increasing number of vertices from $R \setminus \partial R$. These changes in ∂R and $R \setminus \partial R$ apply regardless of whether we move v towards a corner of G or towards the centre of one side of the boundary of G.

Now, if we consider the ratio $\frac{|R \setminus \partial R|}{|\partial R|}$ as we move v towards the boundary, we see that the ratio increases when we move from distance $\ell+1$ to distance ℓ. As we move v further towards the boundary, the ratio of the vertices we remove from $R \setminus \partial R$ to the vertices we remove from ∂R increases, and so $\frac{|R \setminus \partial R|}{|\partial R|}$ either never falls below its initial value, or is at its lowest when the distance is 0 and v lies on the boundary of G. In this situation, we have $|R \setminus \partial R| = \frac{3\ell^2-\ell}{2}$ and $|\partial R| = 3\ell+1$. We therefore have a uniform lower bound on the ratio of the smaller of $\frac{3\ell(\ell-1)}{6\ell}$ and $\frac{3\ell^2-\ell}{6\ell+2}$. We can easily see that the former is $\frac{\ell-1}{2}$. For the latter, we obtain a bound of

$$\begin{aligned} \frac{3\ell^2-\ell}{6\ell+2} &\geq \frac{3\ell^2-\ell}{8\ell} && \text{(if } \ell \geq 1\text{)} \\ &= \frac{3\ell-1}{8}, \end{aligned}$$

and this is a lower bound on $\frac{\ell-1}{2}$ if $\ell \geq 3$.

We require that $\frac{|R \setminus \partial R|}{|\partial R|} > (3\Delta-1)\bar{\lambda}$ to show rapid mixing of $\mathcal{M}'_\ell$. For the triangular lattice, this is satisfied if

$$\begin{aligned} & \frac{3\ell-1}{8} > (3\Delta-1)\bar{\lambda} \\ \Rightarrow \quad & \ell > \frac{8(3\Delta-1)\bar{\lambda}+1}{3} \\ & = \frac{136\bar{\lambda}+1}{3} && \text{since } \Delta = 6 \text{ for the triangular lattice.} \end{aligned}$$

For the kagome lattice, $\Delta = 4$, and the size of the boundary depends on whether ℓ is odd or even. If ℓ is odd, then $|\partial R| = 4\ell$. If ℓ is even, then $|\partial R| = 5\ell$. We know then that the volume of $R \setminus \partial R$ is at least $\sum_{i=1}^{\ell-1} 4i$, so $|R \setminus \partial R| \geq 2\ell(\ell-1)$. We therefore have

$$\frac{|R \setminus \partial R|}{|\partial R|} \geq \frac{2\ell(\ell-1)}{5\ell},$$

and so $\mathcal{M}'_\ell$ and $\mathcal{M}_{CF}$ are rapidly mixing on the kagome lattice if $\ell > \frac{55}{2}\bar{\lambda}+1$. □

Given the difficulty of determining the perimeter and volume of a ball in the kagome lattice (we only give upper and lower bounds, respectively), we do not consider the behaviour of the ratio $\frac{|R \setminus \partial R|}{|\partial R|}$ near the boundary of G. This bound on ℓ therefore only applies if G is defined on a torus, so that the lower bound on the ratio is uniform for all vertices. The argument we used to show a uniform lower bound for the triangular lattice should also apply to the kagome lattice if we are able to find formulae for the perimeter and volume of balls near the boundary of G.

Chapter 6

Sampling perfect matchings of bipartite graphs

We have looked at the problem of sampling from all matchings of a general graph, weighted by a parameter λ, in Chapter 4. However, we are also interested in the problem of sampling from only the perfect matchings of a bipartite graph. If we can find an algorithm to uniformly sample perfect matchings, then we can approximate the permanent of a 0-1 matrix by considering the matrix as the adjacency matrix of a bipartite graph and approximately counting the number of perfect matchings of that graph. In addition, permutation tests in the field of statistics require us to be able to sample randomly from permutations [9]. This, too, is equivalent to sampling perfect matchings of a bipartite graph.

While we have randomised algorithms for sampling weighted matchings of graphs, it is not possible to sample perfect matchings using these algorithms, as we saw in Theorem 4.1.

Recall from Definition 2.1 that a near-perfect matching of a graph G is a matching that covers all but two vertices of G. In 2003, after decades of research in sampling perfect matchings, Jerrum, Sinclair and Vigoda showed that it is possible to sample perfect matchings of bipartite graphs in polynomial time using a Markov chain on perfect and near-perfect matchings [26]. Rather than weighting near-perfect matchings by their size, weights are determined by the positions of the two uncovered vertices. For an appropriate choice of weights, it is possible to sample perfect matchings in a polynomial number of steps. Simulated annealing is used to approximate the required weights. With an appropriate cooling schedule, it is possible to approximately count perfect matchings in time $O(n^7 \log^4 n)$ [2], and given sufficiently close approximations

to the required weights, the time required to sample a perfect matching is $O(n^4 \log n)$.

In this chapter, we examine a natural Markov chain for sampling perfect matchings directly. We analyse the mixing time of this chain (the Diaconis chain) in a restricted class of bipartite graphs that has applications in statistical testing. We show that this chain is *not* rapidly mixing. However, we will see in Section 6.5 that the mixing time of this chain is sub-exponential for graphs that satisfy additional requirements, and that it is polynomial for a very special case in Section 6.6.

6.1 Truncated data and permutation testing

In this section, we give an overview of permutation testing using non-parametric test statistics, in order to provide some motivation for investigating the specific classes of graphs that we will consider in the remainder of this chapter.

Suppose we have two random variables X and Y with some joint distribution (X,Y), and we wish to determine whether X and Y are independent. We can do this by selecting a uniformly random sample of n pairs from the joint distribution. Let $S = \{(X_i, Y_i) \mid 1 \leq i \leq n\}$ be our sample. To determine independence, we compute a test statistic $f(S)$, and compare $f(S)$ against a threshold value (either given by a formula or obtained from a table of known values). This allows us to obtain a bound on the probability that X and Y are independent. For the purposes of this section, we consider only non-parametric test statistics - that is, those in which no assumptions are made on distributions X, Y. Non-parametric tests are concerned with the relative order of the X_i and Y_i values and not their absolute values. We will also make the simplifying assumption that the X_i and Y_i values are distinct. This assumption is reasonable when we consider real-valued data.

We may now assume without loss of generality that S is in fact an ordered sequence of pairs (X_i, Y_i), such that the X_i values are in ascending order. Let S_Y be the sequence of the ranks of the Y_i values in S. S_Y is therefore some permutation of the integers $1, \cdots, n$, and our non-parametric test statistic is some function f of S_Y.

Suitable non-parametric test statistics (suitable f functions) include Spearman's rank correlation coefficient and the Kendall tau coefficient.

Definition 6.1. *Let $S = \{(X_i, Y_i)\}$ be a sequence of paired data such that for each $1 \leq i \leq n$, $X_i = i$, and the Y_is are distinct integers from 1 to n. Spearman's rank correlation*

coefficient, ρ*, is given by*

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(X_i - Y_i)^2}{n(n^2-1)}.$$

The Kendall tau coefficient, τ*, is given by*

$$\tau = \frac{4\sum_{i=1}^{n}|\{(X_j,Y_j) \in S \text{ s.t. } X_j > i, Y_j > i\}|}{n(n-1)} - 1.$$

When we come to interpret $f(S_Y)$ to determine whether X and Y are independent, we compare $f(S_Y)$ against the distribution π_{XY} obtained by evaluating f for *all possible* sequences S_Y, that is, all permutations of $1, \cdots, n$. If $f(S_Y)$ is sufficiently far from the expected value of this distribution, then we conclude that X and Y are not independent. The level of confidence we place in our conclusion is determined by tail inequalities of the distribution π_{XY}.

The assumption that is made in the above paragraph is that our sample is selected uniformly. If this is the case, then when X and Y are independent, every permutation S_Y is equally likely to be selected. If we are unable to sample uniformly, we can still test for independence, provided we compute the distribution π_{XY} weighted according to the probability that each S_Y would be selected, if X and Y were independent.

A particular situation of interest is that of *restricted positions*. In this situation, for each X_i in our sample set, there is some set $J_i \subseteq \{1, \cdots, n\}$, such that, were X and Y independent, our sampling methodology could have obtained every pair $(X_i, Y_j) : j \in J_i$, and could *not* have obtained any pair $(X_i, Y_j) : j \notin J_i$. We call the pairs that our methodology could have detected *observable*. Furthermore, the probability of selecting each observable pair, assuming X and Y are independent, is $\frac{1}{|J_i|}$. We call a permutation $Y_1, \cdots, Y_n$ an *observable permutation* if every pair (X_i, Y_i) is observable. Under these conditions, every observable permutation occurs with equal probability, and so to compute suitable threshold values for our test statistic f, we wish to determine the distribution of f over all observable permutations.

We can now see a connection between observable permutations and perfect matchings of bipartite graphs. Let $G = (V_1, V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = n$, and let

$$E = \{(u_i, v_j) \mid u_i \in V_1, v_j \in V_2, j \in J_i\}.$$

That is, the edge (u_i, v_j) exists in G if and only if the pair (X_i, Y_j) is observable. If every pair (X_i, Y_j) in a permutation is observable, then the set of the corresponding edges $M = \{(u_i, v_j)\}$ is a matching of G, and since each X and Y value is used exactly once, M is a perfect matching. Conversely, if we have a perfect matching M of G, then

every edge $(u_i, v_j) \in M$ corresponds to an observable pair, each vertex in V_1 and V_2 is used exactly once, and so the sequence $[Y_1, \cdots, Y_n]$ is an observable permutation.

The problem of finding observable permutations is therefore equivalent to the problem of sampling perfect matchings in a bipartite graph, and the number of observable permutations is equal to the permanent of the adjacency matrix of the corresponding bipartite graph. This means that the problem of counting observable permutations is #P-complete.

For large datasets, the number of observable permutations can be very high (indeed, if every permutation is observable, there are $n!$ permutations), and it is infeasible to enumerate every observable permutation. It is not necessary, however, to consider every observable permutation: if we can estimate the value of f over a sufficiently large number of almost-uniformly sampled observable permutations, we can still obtain useable threshold values. This form of permutation test is also known as a random permutation test [34].

In this chapter, we consider two specific types of data with restricted positions: doubly truncated data and singly truncated data.

Definition 6.2. *A sample space is referred to as* doubly truncated *if, for every* $1 \leq i \leq n$*, the range of* Y_i *for observable samples* (X_i, Y_i) *is*

$$J_i = \{a_i, \cdots, b_i\},$$

for some pair $a_i, b_i \in \mathbb{N}$*, with* $1 \leq a_i \leq b_i \leq n$*.*

A sample space is referred to as singly truncated *if, for every* $1 \leq i \leq n$*, the range of observable samples* (X_i, Y_i) *is*

$$J_i = \{1, \cdots, b_i\},$$

for some $b_i \in \mathbb{N}$*, with* $1 \leq b_i \leq n$*.*

The reference to *sample spaces* in Definition 6.2 is a reflection of the fact that truncation is not a property of individual samples, nor necessarily of the underlying joint distribution. Rather, it is a property of the set of samples that our experimental methodology allows. Truncation may occur as a result of limitations in equipment (in which case the sample space is truncated but the underlying distribution is not) or, as in the example concerning number of children and birth order given by Diaconis, Graham and Holmes, as a result of some pairs (X_i, Y_i) being meaningless in the context of the experimental setup (in which case the underlying distribution itself is truncated) [9].

Let $G = (V_1, V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = n$, and let $\leq$ be an ordering on V_2. G corresponds to a doubly-truncated sample space if and only if the following condition is satisfied:

$$\forall u \forall v, v_1, v_2 \text{ s.t. } v_1 \leq v \leq v_2, \{(u, v_1), (u, v_2)\} \subseteq E \Rightarrow (u, v) \in E. \tag{6.1}$$

This condition is equivalent to saying that each row of the adjacency matrix of G contains a single contiguous block of 1s, and that all other entries are 0. Graphs that satisfy this property are known as *convex* bipartite graphs [29]. We will consider the special case of singly truncated data, for which results are already known [9], in Section 6.6.

6.2 A Markov chain for perfect matchings

We now define a Markov chain $\mathcal{M}_{PM}$, due to Diaconis, Graham and Holmes [9], whose state space is the set of perfect matchings of a graph G. Current algorithms using simulated annealing use a version of the Jerrum-Sinclair chain that samples from perfect and near-perfect matchings; $\mathcal{M}_{PM}$ will sample only perfect matchings. We will show that the state space of $\mathcal{M}_{PM}$ is not connected for all bipartite graphs; however, it does connect the state space in the case of the special bipartite graphs with which we are concerned.

Definition 6.3. *Let $G = (V_1, V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = n$ and $|E| = m$. Define the transitions of the Markov chain $\mathcal{M}_{PM}$ on G as follows: Suppose that at time t, we have a perfect matching X_t of G.*

1. *Choose two vertices u_1 and u_2 of V_1 uniformly at random. Let (u_1, v_1) and (u_2, v_2) be the edges in X_t incident to these vertices.*

2. *Let M be*
$$M = X_t \cup \{(u_1, v_2), (u_2, v_1)\} \setminus \{(u_1, v_1), (u_2, v_2)\}.$$

3. *If $M \subseteq E$, then M' is a perfect matching, and set $X_{t+1} = M$. Otherwise, set $X_{t+1} = X_t$.*

We call this a transposition *move, and will refer to transposing pairs of edges. Such moves are also referred to as Diaconis moves.*

Note that the transition probability is non-zero exactly when $\{(u_1,v_1),(u_2,v_2)\} \subseteq M$, and $\{(u_1,v_1), (u_1,v_2), (u_2,v_2), (u_2,v_1)\}$ is a 4-cycle in G. Every such valid transition has the same probability: $P(M,M') = \frac{2}{n^2}$. u_1 and u_2 are chosen with replacement, which means there is a minimum self-loop probability of $\frac{1}{n}$ for all states. We will use this property in Section 6.6.

Diaconis, Graham and Holmes showed that if the adjacency matrix of G satisfies Condition 6.1, then the set of all perfect matchings in G is connected by this type of transition [9]. Attempts to show rapid mixing using canonical paths also naturally lead to a proof that the state space is connected, as we will see in Section 6.3.

The state space of perfect matchings of general bipartite graphs is *not* connected by $\mathcal{M}_{PM}$. Consider the graph consisting of a single 6-cycle. This graph is bipartite and has two perfect matchings, but as it contains no 4-cycles, no transitions are possible and so the perfect matchings cannot be connected by transposition moves.

6.3 Adjacency matrices with contiguous rows only

Let G be a convex bipartite graph (that is, a graph satisfying Condition (6.1)). Given two perfect matchings of G, M and M', the symmetric difference $M \oplus M'$ consists of a set of even-length alternating cycles. This suggests that a canonical paths argument could be used to bound the mixing time of a Markov chain on perfect matchings. Strictly $M \oplus M'$ is a set of edges. However, it will be useful to consider a cycle in terms of its set of vertices. Let $V(M \oplus M') = \bigcup_{(u,v)\in M\oplus M'} \{u,v\}$ be the set of all endpoints of the edges in $M \oplus M'$.

Lemma 6.4 will show that, given an alternating cycle, it is always possible to find a pair of edges that can be transposed, yielding a smaller cycle. This immediately implies that the state space of $\mathcal{M}_{PM}$ is connected.

Lemma 6.4. *Given a two perfect matchings M and M' of a convex bipartite graph $G = (V_1,V_2,E)$, and a cycle C of length $\ell \geq 6$ in $M \oplus M'$, there exists a 4-cycle C' such that three edges of C' are edges of C, and $C \oplus C'$ is a cycle of length $l-2$.*

Proof. Let v_0 be the smallest vertex of V_2 in $V(C)$, and a and b the two vertices of V_2 such that, for some vertices $c,d \in V_1$, (c,v_0), (c,a), (d,v_0) and (d,b) are edges of C.

If $a < b$, then by Condition (6.1), $(d,a) \in E$, and the edges (c,v_0), (d,v_0), (d,a) and (c,a) form the 4-cycle C'. Likewise, if $a > b$, then $(c,b) \in E$, and the edges (c,v_0),

(d, v_0), (d, b) and (c, b) make up C'. The case $a = b$ cannot occur, as this would imply C is already of length 4. □

Corollary 6.5. *The state space of perfect matchings in any convex bipartite graph G is connected by transposition moves.*

Proof. Given two perfect matchings M and M' in G, and assuming $M \oplus M'$ contains an alternating cycle of length at least 6, Lemma 6.4 allows us to find a 4-cycle C' such that $|M \oplus M' \oplus C'| = |M \oplus M'| - 2$. If there is no such cycle, then let C' be any 4-cycle in $M \oplus M'$. C' corresponds to a valid transition either from M or from M'. Repeating this process gives a sequence of transitions connecting M and M'. □

Corollary 6.5 provides an obvious set of canonical paths connecting states of $\mathcal{M}_{PM}$. However, for any cycle C of length at least 6, the cycle C' that we obtain by applying Lemma 6.4 contains only three edges of C, and the remaining edge is not necessarily a member of $M \oplus M'$. We need to modify at least one edge that is not in $M \oplus M'$ in order to move from M to M', unless $M \oplus M'$ consists entirely of 4-cycles. This extra edge would require additional information to be supplied when we try to recover M and M' from the encoding η_t later, and such additional information would increase the bound on the mixing time.

In fact, we can easily see that there are cycles that require the use of at least three extra edges to process in a path from M to M', regardless of the choice of canonical paths. Figure 6.1 shows the steps required to process a 12-cycle. Edges shown as dashes are those that are implied by the ordering of the vertices, but which are not in $M \oplus M'$, and there are no additional edges between the vertices of the cycle that are not shown. From the initial state (the bold edges), we may transpose either the 4-cycle BbCc, or DdEe. Either choice adds two additional edges. Suppose we transpose BbCc. In the second step, we transpose AaBb, which leaves only one additional edge remaining (we could instead transpose DdEe, which would add two more additional edges, giving four in total). We are now forced to transpose DdEe, and have three additional edges. It is possible that cycles could be constructed that require arbitrary numbers of additional edges, and repeatedly applying Lemma 6.4 does not necessarily lead to the most efficient path. The paths given by Lemma 6.4 are therefore not a good candidate for bounding congestion. Furthermore, as we will see in Theorem 6.6, it is possible to construct graphs that satisfy Condition (6.1) that can be shown to have high congestion.

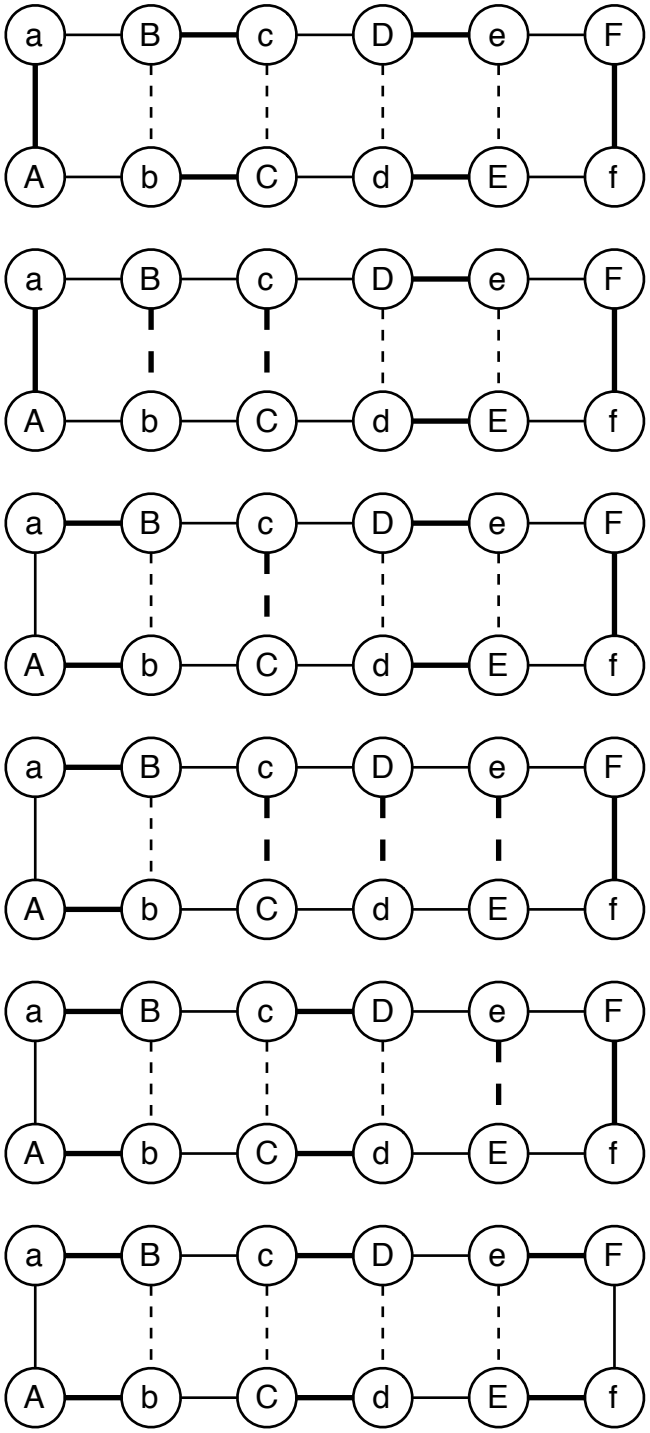

Figure 6.1: *Processing a 12-cycle.*

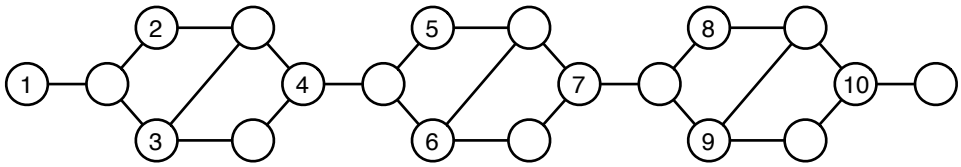

Figure 6.2: *The graph* $(V_1, V_2, E_{H1} \cup E_{J1})$ *for* $k = 3$.

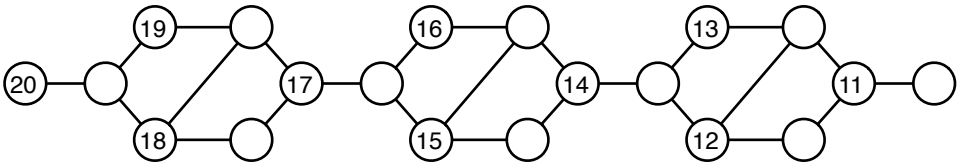

Figure 6.3: *The graph* $(V_1, V_2, E_{H2} \cup E_{J2})$ *for* $k = 3$.

In Section 6.5, we will consider a further restricted class of graphs, and give a method for finding canonical paths that require at most a logarithmic number of additional edges, yielding a sub-exponential bound on the mixing time for this more restricted class of graphs.

We now show that there exist graphs whose adjacency matrices have contiguous rows of 1s, for which the mixing time of $\mathcal{M}_{PM}$ is exponential. We do this by constructing a family of such graphs and showing that their conductance is exponentially small.

Theorem 6.6. *The mixing time of* $\mathcal{M}_{PM}$ *is exponential in the worst case, for the class of graphs satisfying Condition* (6.1).

Proof. We will construct a family of graphs, and show that their conductance is exponentially small, and therefore the mixing time is exponential.

For any integer $k \geq 1$, let $V_1 = \{u_1, \ldots, u_{6k+2}\}$, and $V_2 = \{v_1, \ldots, v_{6k+2}\}$. Now

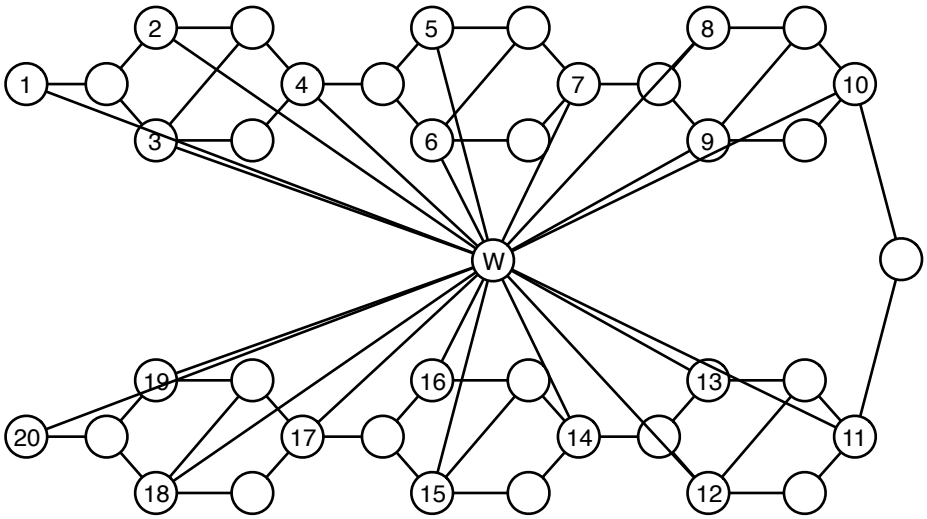

Figure 6.4: *A graph with contiguous rows and high congestion.*

define the following edge sets:

$$
\begin{aligned}
E_{H1} = \bigcup_{0\le i<k} \{&(u_{3i+1}, v_{3i+2}), (u_{3i+1}, v_{3i+3}), (u_{3i+2}, v_{3i+2}),\\
&(u_{3i+2}, v_{3i+3}), (u_{3i+2}, v_{3i+4}), (u_{3i+3}, v_{3i+3}), (u_{3i+3}, v_{3i+4})\}\\
E_{H2} = \bigcup_{k\le i<2k} \{&(u_{3i+2}, v_{3i+2}), (u_{3i+2}, v_{3i+3}), (u_{3i+3}, v_{3i+2}),\\
&(u_{3i+3}, v_{3i+3}), (u_{3i+3}, v_{3i+4}), (u_{3i+4}, v_{3i+3}), (u_{3i+4}, v_{3i+4})\}\\
E_{J1} = \{&(u_{3k+1}, v_{3k+1})\} \cup \bigcup_{0\le i<k} \{(u_{3i+1}, v_{3i+1})\}\\
E_{J2} = \{&(u_{3k+1}, v_{3k+2})\} \cup \bigcup_{k\le i<2k} \{(u_{3i+4}, v_{3i+5})\}\\
E_W = &\bigcup_{1\le i<6k+3} \{(u_{6k+2}, v_i)\}\\
E = &E_{H1} \cup E_{H2} \cup E_{J1} \cup E_{J2} \cup E_W .
\end{aligned}
$$

E_{H1} and E_{H2} define two sets of hexagons, each with an extra edge to satisfy Condition (6.1). Each hexagon has exactly three perfect matchings, so the graph (V_1, V_2, E_{H1}) has 3^k perfect matchings, as does the graph (V_1, V_2, E_{H2}). E_{J1} and E_{J2} define edges connecting the hexagons of E_{H1} and E_{H2}. Figure 6.2 shows the graph $(V_1, V_2, E_{H1} \cup E_{J1})$ for $k = 3$, with the v_i vertices labelled. Note that there is exactly one perfect matching of this graph. Figure 6.3 shows the graph $(V_1, V_2, E_{H2} \cup E_{J2})$ for $k = 3$, also with the v_i vertices labelled. These two graphs differ only in the labelling of the v_i vertices, and

each has a unique perfect matching.

E_W adds a set of "whisker" edges connecting u_{6k+2} to every vertex in V_2. E is the union of these sets. The graph $G = (V_1, V_2, E)$ (again, for $k = 3$) is shown in Figure 6.4, also with the v_i vertices labelled. Vertex u_{6k+2} is labelled W. It is straightforward to verify that Condition (6.1) is satisfied for each of the u_i vertices in this graph.

Any perfect matching must contain exactly one of the whisker edges connected to W. Consider the case when (W, v_x) exists in a perfect matching of G and $x \leq 3k+1$. All the edges in E_{J2} must be present, and so there is exactly one configuration for the edges $E_{H2} \cup E_{J2}$. Similarly, if $x \geq 3k+2$ there is exactly one configuration for the edges $E_{H1} \cup E_{J1}$.

When (W, v_{3k+1}) or (W, v_{3k+2}) is present in a perfect matching, there is exactly one configuration for the remaining edges. Call these two states a and b respectively. Note also that there is a unique transition, namely $a \to b$, between a state where $x \leq 3k+1$ and one where $x \geq 3k+2$.

If (W, v_1) is present in a matching, then the edges in E_{J1} cannot be present, and so there are 3^k possible configurations of E_{H1}. Likewise, if (W, v_{6k+2}) is present, then there are 3^k possible configurations of E_{H2}. This gives a lower bound on the total number of perfect matchings: $|\Omega| \geq 2(3^k)$. Therefore, for any state s, $\pi(s) \leq \frac{1}{2(3^k)}$.

We will now show that the conductance is exponentially small. Let S be the set of perfect matchings where $x \leq 3k+1$. By symmetry, note that $|S| = |\Omega \setminus S|$, and so $\pi(S) = \pi(\Omega \setminus S) = \frac{1}{2}$. Thus, the conductance is no greater than

$$\begin{aligned}
\Phi &\leq \frac{\sum_{x \in S, y \notin S} \pi(x) P(x,y)}{\pi(S)} \\
&\leq \frac{4\pi(a)}{(2(6k+2))^2} \\
&\leq \frac{1}{2(3^k)(6k+2)^2}.
\end{aligned}$$

Therefore by Theorem 2.22, the spectral gap is $1 - |\lambda_1| \leq \frac{1}{2(3^k)(6k+2)^2}$, and so by Theorem 2.21 the mixing time is at least

$$\begin{aligned}
\tau(\varepsilon) &\geq \frac{|\lambda_1|}{2(1-|\lambda_1|)} \log\left(\frac{1}{2\varepsilon}\right) \\
&\geq \frac{2(3^k)(6k+2)^2 - 1}{2} \log\left(\frac{1}{2\varepsilon}\right).
\end{aligned}$$

□

$$\begin{bmatrix}
0 & 0 & 0 & 0 & \mathbf{1} & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & \mathbf{1} & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{1} & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & \mathbf{1} & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{1} & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & \mathbf{1} \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \mathbf{1} & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \mathbf{1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & \mathbf{1} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & \mathbf{1} & 0 & 0 & 0 & 0
\end{bmatrix}$$

Figure 6.5: *The adjacency matrix of* (V_1, V_2, E) *for* $k = 6$. *The entries in bold face correspond to the edges of one perfect matching.*

6.4 Contiguous rows and columns

We will now attempt to find a further restricted class of graphs for which $\mathcal{M}_{PM}$ is rapidly mixing. Suppose that as well as an ordering on V_2, the vertices in V_1 are also ordered.

Definition 6.7. *We say that the adjacency matrix of a bipartite graph* $G = (V_1, V_2, E)$ *has* contiguous rows and columns *if we have an ordering on both* V_1 *and* V_2, *Condition* (6.1) *holds, and additionally,*

$$\forall v \forall u, u_1, u_2 \text{ s.t. } u_1 \leq u \leq u_2, \{(u_1, v), (u_2, v)\} \subseteq E \Rightarrow (u, v) \in E. \tag{6.2}$$

Informally, Definition 6.7 means that any row or column of the adjacency matrix of G contains a single contiguous block of 1s, and all other entries are 0. Graphs satisfying both Condition (6.1) and Condition (6.2) are also referred to as *doubly convex* [29].

It is clear that the counterexample from Section 6.3 does not fit into this more restricted class of graphs. However, we can still find a family of graphs for which the mixing time of $\mathcal{M}_{PM}$ is exponential.

Theorem 6.8. *The worst-case mixing time of* $\mathcal{M}_{PM}$ *is exponential for graphs whose adjacency matrices have contiguous rows and columns.*

Proof. As before, we define a family of graphs and show that the conductance is exponentially small. For $k > 2$, let $V_1 = \{u_1, \ldots, u_{2k-1}\}$ and $V_2 = \{v_1, \ldots, v_{2k-1}\}$. The edge

set is defined as:

$$E_1 = \bigcup_{\substack{1 \le i \le k \\ \max(k-i,1) \le j < k}} \{u_i, v_j\}$$

$$E_2 = \bigcup_{\substack{0 \le i < k \\ k < j \le \min(2k-i, 2k-1)}} \{u_{k+i}, v_j\}$$

$$E = E_1 \cup E_2.$$

Due to the number of edges, it is easier to see the structure of G in its adjacency matrix. Figure 6.5 shows the adjacency matrix of G for $k = 6$, with the entries corresponding to the edges of one perfect matching shown in bold face.

For a given matching M, consider the edge incident to u_k. This edge corresponds to a 1-entry in row k of the adjacency matrix. Suppose $(u_k, v_x) \in M$ for some x.

If $x > 1$, then (u_{k-1}, v_1) must be in M. Generally, for every $1 \le y < \min(x,k)$, $(u_{k-y}, v_y) \in M$. Similarly, for every $\max(x,k) < y < 2k$, $(u_{3k-y}, v_y) \in M$. It follows that there is exactly one perfect matching M such that $(u_k, v_k) \in M$. There are $2(k-1)$ transitions that can be made from this state: $k-1$ to states where $x < k$, and $k-1$ to states where $x > k$.

Observe that any transition involves four 1-entries in the adjacency matrix, and that these entries must form the corners of a rectangle. It is not possible to move directly from a state where $x < k$ to one where $x > k$, because there is no choice of four 1-entries such that at least one is in a column left of column k, at least one is in a column right of column k, and the four entries are the corners of a rectangle. There is therefore no path between any state where $x < k$ to any state where $x > k$ that does not pass through the unique state where $x = k$.

For each y such that $x < y \le k$, given the edges incident to the vertices $v_1, \ldots, v_{y-1}$ (that is, the entries in columns 1 to $y-1$), there are exactly two possible edges incident to vertex v_y. This means that there are 2^{k-x} perfect matchings for any $1 \le x \le k$, and therefore $2^k - 2$ perfect matchings where $x < k$. Due to the symmetry of the adjacency matrix, it immediately follows that there are $2^k - 2$ perfect matchings where $x > k$.

We are now in a position to compute the conductance. Let S be the set of perfect matchings with $x \le k$. $|S| = 2^k - 1$, $|\Omega| = 2^{k+1} - 3$, and $|\Omega \setminus S| = 2^k - 2$. The

conductance is

$$\begin{aligned}\Phi &\leq \frac{\sum_{x\in S, y\notin S} \pi(x)P(x,y)}{\pi(S)} \\ &= \frac{2(2^{k+1}-3)(k-1)}{(2k-1)^2(2^k-1)(2^{k+1}-3)} \\ &= \frac{2(k-1)}{(2k-1)^2(2^k-1)},\end{aligned}$$

where we are using the fact that $n = (2k-1)^2$ in our substitution for $\pi(x)$.

Then by Theorem 2.22 the spectral gap is $1-|\lambda_1| \leq \frac{4(k-1)}{(2k-1)^2(2^k-1)}$, and by Theorem 2.21 the mixing time satisfies

$$\begin{aligned}\tau(\varepsilon) &\geq \frac{|\lambda_1|}{2(1-|\lambda_1|)} \log\left(\frac{1}{2\varepsilon}\right) \\ &\geq \frac{(2k-1)^2(2^k-1)-4(k-1)}{(2k-1)^2(2^k-1)} \frac{(2k-1)^2(2^k-1)}{8(k-1)} \log\left(\frac{1}{2\varepsilon}\right) \\ &= \frac{(2k-1)^2(2^k-1)-4k+4}{8k-8} \log\left(\frac{1}{2\varepsilon}\right).\end{aligned}$$

□

6.5 Monotonic matrices

We now restrict the class of graphs we wish to sample from even further. This new class of graphs still encompasses many of those associated with doubly truncated data, such as the quasar data considered by Efron and Petrosian [16]. As in Section 6.4, we require that there is an ordering on both V_1 and V_2. In the adjacency matrix of G, we now require that for all $i < j \leq n$, the leftmost 1 in row j is no further left than the leftmost 1 in row i, and that the rightmost 1 in row j is no further left than the rightmost 1 in row i. Formally, in addition to Condition (6.1), we require that for all $1 \leq i \leq j < n$:

$$\min_{(u_i,v_k)\in E} v_k \leq \min_{(u_j,v_k)\in E} v_k \tag{6.3}$$

$$\max_{(u_i,v_k)\in E} v_k \leq \max_{(u_j,v_k)\in E} v_k. \tag{6.4}$$

Note that these properties automatically hold for columns as well as rows. We refer to adjacency matrices satisfying Equations (6.1), (6.3) and (6.4) as *monotonic* because the function $f(i) = \min_{(u_i,v_j)\in E} v_j$ is monotone, as are $\max_{(u_i,v_j)\in E} v_j$, $\min_{(u_j,v_i)\in E} u_j$ and $\max_{(u_j,v_i)\in E} u_j$.

Diaconis, Graham and Holmes also considered matrices of this form, referring to them as *monotone*. They showed that perfect matchings of graphs with monotone adjacency matrices are connected under transpositions of adjacent rows and columns in the matrix [9].

This is a further restriction to the case of contiguous rows and columns that we considered in Section 6.4. While we have been unable to show that $\mathcal{M}_{PM}$ is rapidly mixing for this class of graphs, we can at least show that its mixing time is sub-exponential.

Theorem 6.9. *The mixing time of $\mathcal{M}_{PM}$ on graphs with monotonic adjacency matrices satisfies*

$$\tau(\varepsilon) \leq 2n^3 m^{2\lg n}(2\log\varepsilon^{-1} + n\log n).$$

We will prove Theorem 6.9 by using canonical paths to bound the congestion. For each pair of states I and F, we will define a canonical path γ_{IF} from I to F using the transitions of the Markov chain. The symmetric difference $I \oplus F$ consists of a set of even length alternating cycles. Given I and F, we will process each cycle C in turn, in some deterministic order (for example, increasing order of smallest vertex). We require that the number of paths using each transition of the chain is not too large. Before we proceed with the proof of Theorem 6.9, however, we will prove some additional properties of graphs with monotonic adjacency matrices.

Lemma 6.10 will prove a property of these graphs that is useful for constructing canonical paths.

Lemma 6.10. *Let $G = (V_1, V_2, E)$ be a bipartite graph with a monotonic adjacency matrix A. Suppose that (a,b) and (c,d) are edges of G. If $a < c$ and $b > d$, then (a,d) and (c,b) are also edges of G.*

Proof. Consider row a of A. From Condition (6.3), we know that $\min_{(a,v_i)\in E} v_i \leq d$, and since $(a,b) \in E$, that $\max_{(a,v_i)\in E} v_i \geq b$. Therefore

$$\min_{(a,v_i)\in E} v_i \leq d \leq \max_{(a,v_i)\in E} v_i,$$

and so $(a,d) \in E$.

Applying Condition (6.4) in the same way to row c shows that $(b,c) \in E$. □

For a cycle C, define a pair of *parallel edges* as two edges (a,b) and (c,d) of G, such that (a,d) and (b,c) are edges present in C. Furthermore, we require that there is a path from a to b along the cycle C that contains neither c nor d, and that there is a

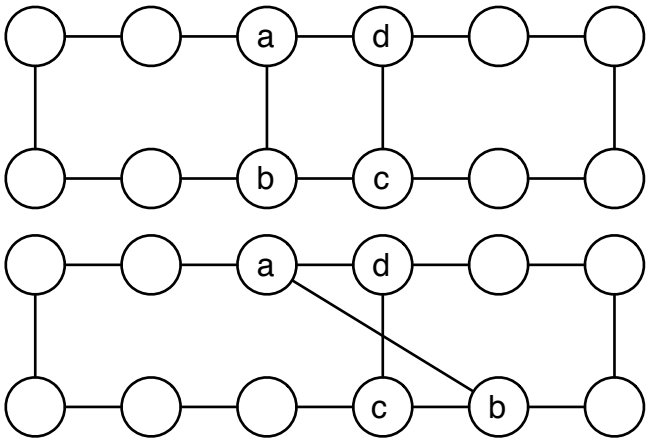

Figure 6.6: *Above, a pair of parallel edges bisecting a cycle C. Below, a pair of crossing edges - note that either (a,d) or (b,c) exists in any path connecting a to b through the edges of C.*

path from c to d along the cycle C that contains neither a nor b. This ensures that the edges are "parallel" rather than crossing, as shown in Figure 6.6.

We will now show that a pair of parallel edges exists that bisects any cycle, i.e. by removing (a,d) and (b,c), and adding (a,b) and (c,d), two cycles are formed whose lengths differ by no more than 2. Since G is bipartite, it is not possible to exactly bisect a cycle whose length is not a multiple of 4, so we must allow the new cycles to differ slightly in length.

Lemma 6.11. *Let C be a cycle of length $\ell \geq 6$ in a graph with a monotonic adjacency matrix. There exists a pair of parallel edges (a,b) and (c,d), such that $C \cup \{(a,b),(c,d)\} \setminus \{(a,d),(b,c)\}$ contains the edges of a pair of vertex-distinct cycles, each of length no greater than $\frac{\ell}{2}+1$. (We also permit the case where the cycle is of length 6 and there is only one remaining cycle - in this case the result is a special case of Lemma 6.4.)*

Proof. Let the edges of C be (u_0,v_0), (u_0,v_1), (u_1,v_1), $(u_1,v_2),\ldots,$ $(u_{\frac{\ell}{2}-1},v_{\frac{\ell}{2}-1})$, $(u_{\frac{\ell}{2}-1},v_0)$, such that u_0 is the smallest vertex of V_1 in $V(C)$, and some u_α is the largest. For convenience, indices will be computed modulo $\frac{\ell}{2}$ throughout this proof.

Define $\delta = \lfloor \frac{\ell}{4} \rfloor$.

Consider a pair of edges (u_i,v_i) and $(u_{i+\delta},v_{i+\delta})$, and examine the signs of $u_i - u_{i+\delta}$ and $v_i - v_{i+\delta}$. If exactly one of these is positive, then Lemma 6.10 implies that the parallel edges $(u_i,v_{i+\delta})$ and $(u_{i+\delta},v_i)$ exist and bisect C.

If this is not the case, then consider the next pair of edges (u_i, v_{i+1}) and $(u_{i+\delta}, v_{i+\delta+1})$ and examine the signs of $u_i - u_{i+\delta}$ and $v_{i+1} - v_{i+\delta+1}$. Only one of the signs (that of $v_i - v_{i+\delta}$) can have changed, if any. The next pair of edges is (u_{i+1}, v_{i+1}) and $(u_{i+\delta+1}, v_{i+\delta+1})$. Again, only one of the signs (this time, that of $u_i - u_{i+\delta}$) can have changed.

Now, beginning with (u_0, v_0) and (u_δ, v_δ), we follow a pair of edges around the cycle. At some point, we reach the edges (u_α, v_α) and $(u_{\alpha+\delta}, v_{\alpha+\delta})$. Since u_0 is the smallest vertex of $V_1 \cap V(C)$ and u_α is the largest, the sign of $u_i - u_{i+\delta}$ must have changed at some point during this process, and since only one sign can change at any step, there must be a point at which exactly one sign is positive.

Therefore there is some pair of edges (u_i, v_j) and $(u_{i+\delta}, v_{j+\delta})$ in C such that the parallel edges $(u_i, v_{j+\delta})$ and $(u_{i+\delta}, v_j)$ are present in G and bisect C. □

The following properties of matchings and cycles will also be useful in proving Theorem 6.9:

Lemma 6.12. *In any graph $G = (V, E)$, let $C \subseteq E$ be a set of disjoint cycles.*

1. *If C' is another set of disjoint cycles in G, and for any vertex $v \in V$*

$$\deg_C(v) = \deg_{C'}(v) = 2 \Rightarrow \exists u \mid (u, v) \in C \cap C' \tag{6.5}$$

   then $C \oplus C'$ is a set of disjoint cycles.

2. *If M is a perfect matching of G, and for any $v \in V$*

$$\deg_C(v) = 2 \Rightarrow \exists u \mid (u, v) \in C \cap M \tag{6.6}$$

   then $C \oplus M$ is a perfect matching.

Proof. First, note that for every vertex v in V, $\deg_C(v) \in \{0, 2\}$, and $\deg_M(v) = 1$.

1. For each vertex $v \in V$, consider $\deg_C(v)$ and $\deg_{C'}(v)$. If both are zero, then $\deg_{C \oplus C'}(v) = 0$. If exactly one is 2, then the two incident edges are present in $C \oplus C'$, and so $\deg_{C \oplus C'}(v) = 2$. If both are 2, then there are two possibilities: there is a single common edge, in which case that edge is not present in $C \oplus C'$ and $\deg_{C \oplus C'}(v) = 2$; or there are two common edges, in which case both cancel out, and $\deg_{C \oplus C'}(v) = 0$. Note that Condition (6.5) guarantees there is at least one common edge.

   Since the degree of every vertex of $C \oplus C'$ is even and at most 2, each connected component is Eulerian and has a simple Eulerian cycle. Therefore, $C \oplus C'$ is a set of simple disjoint cycles.

2. For any vertex $v \in V$, assume $\deg_C(v) = 0$. There is exactly one edge $(u,v) \in M$ incident to v, so $\deg_{C \oplus M} = 1$. Now assume $\deg_C(v) = 2$. By Condition (6.6), the unique incident edge (u,v) is also in C, so it is not in $C \oplus M$. There is a second edge $(u',v) \in C$ that is incident to v. This edge is in $C \oplus M$, and so $\deg_{C \oplus M}(v) = 1$.

   Since the degree of every vertex in $C \oplus M$ is 1, $C \oplus M$ is a perfect matching. □

We now complete the proof of Theorem 6.9.

Proof of Theorem 6.9. For any pair of perfect matchings I and F, we define a canonical path from I to F. To do this, process each cycle in $I \oplus F$ in turn, in order of smallest vertex in V_1. To process each cycle C, we perform the following procedure:

1. Find a pair of parallel edges (a,b) and (c,d) bisecting C. Let $P = \{(a,b), (c,d), (a,d), (b,c)\}$ be the 4-cycle formed by these edges. If C is of length 4 then let $P = C$.

2. If (a,d) and (b,c) are present in I, then transpose the quadrilateral formed by P. This forms up to two smaller cycles C_1 and C_2, each including one of the parallel edges. Recursively process cycles C_1 and C_2, beginning with whichever contains the smallest vertex in V_1.

3. If (a,d) and (b,c) are present in F, then process C_1 and C_2 beginning with whichever does *not* have the smallest vertex in V_1, and transpose the quadrilateral P *after* processing C_1 and C_2.

Observe that the order in which a given cycle C is processed in the path from I to F is exactly the reverse of the order in which C is processed in the path from F to I. This will be important for recovering I and F from the encoding η_t later.

At any point in processing C, we will have a "stack" of 4-cycles that have been transposed. Let P^* be the symmetric difference of these for the cycle (and sub-cycles) currently being processed, but not the one corresponding to the current transition. If the cycle is of length ℓ, then there are at most $\lfloor \lg \ell \rfloor - 1$ such sets, one for each level of recursion.

We now need to bound the number of paths using each transition $t = M \to M'$. For each pair of states I and F, we can construct a function

$$\eta_t(I,F,P^*) = I \oplus F \oplus P^* \oplus ((M \cap M') \cup \{(a,d),(b,c)\}).$$

At each level, P is a 4-cycle that has two edges in common with $I \oplus F$. Therefore, we can apply the first part of Lemma 6.12, and so $I \oplus F \oplus P^*$ is a set of disjoint cycles. Every vertex of $M \cap M'$ has degree 1, except for the four vertices a, b, c, and d, which have degree 0. Thus, $((M \cap M') \cup \{(a,d),(b,c)\})$ is a perfect matching. Also note that (a,d) and (b,c) are edges in $I \oplus F \oplus P^*$. Therefore, $((M \cap M') \cup \{(a,d),(b,c)\})$ satisfies Condition (6.6), and so by the second part of Lemma 6.12, η_t is a perfect matching.

Now, given η_t, P^*, and the edges (a,d) and (b,c), it is easy to reconstruct $I \oplus F$, and therefore the order in which cycles are processed. We can find out which cycle C is currently being processed by looking at the vertices of the edges that t changes (we introduce extra edges while processing C, but never extra vertices). η_t agrees with I for those cycles that have been fully processed, and with F for those that have not yet been processed. Any edges in η_t that do not share any vertices with the cycles in $I \oplus F$ are present in both I and F.

The preceding observation about η_t allows us to reconstruct I and F for all edges except those of the current cycle C. In order to reconstruct I and F for these edges, we need to consider the transition t. There are two possible perfect matchings of the cycle C - call these α and β. One of α and β is a subset of I, and the other is a subset of F. To determine which is a subset of I, and therefore recover I and F, assume $\alpha \subseteq I$, and decompose the cycle as before. This gives a sequence of transitions, and if t belongs to this sequence then $\alpha \subseteq I$ and $\beta \subseteq F$. Otherwise, the reverse of t - that is, $t' = M' \to M$ - belongs to the sequence of moves, so $\beta \subseteq I$ and $\alpha \subseteq F$. While it is not strictly necessary, the fact that the path from α to β is exactly the reverse of that from β to α makes it easier to see that the two paths do not share any transitions.

Thus, we can recover I and F from t, η_t, P^*, and the two edges (a,d) and (b,c). Each 4-cycle in P^* can be uniquely identified by two of its edges (since G is bipartite), so there are at most $|\Omega| m^{2\lg n}$ paths using each transition. The length of each path is at most n transitions.

This allows us to bound the congestion of $\mathcal{M}_{PM}$:

$$\begin{aligned}\rho(\Gamma) &= \max_{t=(u,v)} \left\{ \frac{1}{\pi(u)P(u,v)} \sum_{x,y\,:\,\gamma_{xy} \text{ uses } t} \pi(x)\pi(y)|\gamma_{xy}| \right\} \\ &\leq \frac{1}{2} n^3 m^{2\lg n}.\end{aligned}$$

Therefore, by Lemma 2.24, the mixing time satisfies

$$\begin{aligned}\tau_x(\varepsilon) &\leq 2\rho(2\log\varepsilon^{-1} + \log\pi(x)^{-1}) \\ &\leq n^3 m^{2\lg n}(2\log\varepsilon^{-1} + n\log n)\end{aligned}$$

as required. □

It is important that we can find a pair of parallel edges. If we can find only a single edge bisecting C, then it is possible to process the cycle in a similar way, but not to find the encoding required for the canonical paths argument. Indeed, it is possible to find such an edge even in the case with contiguous rows only. However, we have already seen that $\mathcal{M}_{PM}$ has exponential mixing time for graphs whose matrices have only contiguous rows.

6.6 Left-aligned matrices

Finally, we consider the case where the 1-entries of each row of the adjacency matrix are aligned to the left. This corresponds to the problem of sampling permutations where every sample y_i lies in the range $[1, b_i]$ - that is, where the data are singly truncated. We can use coupling to show that $\mathcal{M}_{PM}$ is rapidly mixing for these graphs.

Note that there are already known algorithms for sampling permutations of singly truncated data exactly, and that there is also an exact formula for computing the permanent of such matrices [9]. We carry out the analysis of $\mathcal{M}_{PM}$ here to show that there is a sufficiently restricted class of graphs for which it is rapidly mixing.

Lemma 6.13. *The mixing time of $\mathcal{M}_{PM}$ on graphs with left-aligned adjacency matrices satisfies*

$$\tau(\varepsilon) \leq \frac{n^2(n-1)}{2}\lceil\lg(\varepsilon^{-1})\rceil.$$

Proof. We will prove Lemma 6.13 by coupling. Consider two copies of $\mathcal{M}_{PM}$, X_t and Y_t. Define the distance between X_t and Y_t:

$$d(X_t, Y_t) = \{\max i \mid \nexists j \text{ s.t. } (u_j, v_i) \in X_t, (u_j, v_i) \in Y_t\}.$$

Informally, d is the index of the rightmost column in the adjacency matrix where X_t and Y_t disagree.

For one step of the coupling, we select the edges (u_1, v_1) and (u_2, v_2) of X_t as usual. Assume without loss of generality that $v_2 \geq v_1$. Find u_3 and v_3 such that (u_1, v_3) and

(u_3, v_2) are edges in Y_t. Attempt to transpose (u_1, v_1) and (u_2, v_2) in X_t, and (u_1, v_3) and (u_3, v_2) in Y_t. Note that u_3 and v_3 are unique, and that it is possible to recover v_1 and u_2 given (u_1, v_3) and (u_3, v_2). Therefore there is a bijection between the transitions in X_t and Y_t. Since all transitions in X_t have the same probability, it follows that Y_t is a faithful copy of $\mathcal{M}_{PM}$.

In analysing this coupling, we wish to show that there is no move that can result in an increase in distance, and that there is always at least one move that can result in a decrease in distance. v_2 corresponds to a column in the adjacency matrix of G. We will refer to the index of this column as $c(v_2)$. The aim of this coupling is to encourage column $c(v_2)$ to agree between X_t and Y_t. Suppose first that $c(v_2) > d(X_t, Y_t)$. Then either $c(v_1) \leq d(X_t, Y_t)$ and $c(v_3) \leq d(X_t, Y_t)$, or $c(v_1) > d(X_t, Y_t)$ and $c(v_3) > d(X_t, Y_t)$. In either case, either both moves succeed or both moves fail, and column $c(v_2)$ continues to agree. In the former case, columns $c(v_1)$ and $c(v_3)$ both lie to the left of column $d(X_t, Y_t)$ and so the distance does not increase (one may lie on column $d(X_t, Y_t)$, which leads to a possible decrease in distance); in the latter case, $v_1 = v_3$, and so the distance does not change.

Now suppose $c(v_2) \leq d(X_t, Y_t)$ and $c(v_3) > d(X_t, Y_t)$. This would mean that column $c(v_3)$ agreed between X_t and Y_t, and therefore that $v_1 = v_3$. This contradicts our assumption that $v_2 \geq v_1$. We have covered all possible situations in which any of $c(v_1)$, $c(v_2)$ and $c(v_3)$ lie to the right of column $d(X_t, Y_t)$, and hence shown that it is impossible for the distance to increase.

We now need to show that it is possible for the distance to decrease. This can occur if $c(v_2) = d(X_t, Y_t)$. We know from the previous paragraph that $c(v_1) \leq c(v_2)$ and $c(v_3) \leq c(v_2)$. If (u_1, v_2) is an edge in G, then both moves will succeed and column $c(v_2)$ will agree. If (u_1, v_2) is not an edge, then both moves will fail and the distance will be unchanged. If $d(X_t, Y_t) \geq 2$, there is always at least one possible choice of (u_1, v_1) that will cause these moves to succeed, and therefore that there is at least one possible case for which the distance decreases.

Thus, if $d(X_t, Y_t) \geq 1$, then at time t, with probability at least $\frac{2}{n^2}$, $d(X_{t+1}, Y_{t+1}) \leq d(X_t, Y_t) - 1$, and with the remaining probability, $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t)$.

At any given time, there is at least one move, chosen with probability $\frac{2}{n^2}$, that can cause the distance $d(X_t, Y_t)$ to decrease. Assume that there is only ever one such move, and that the distance never decreases by more than 1 (except when $d(X_t, Y_t) = 2$, where the distance must decrease by 2) - if either of these assumptions does not hold, then the chains will converge more rapidly, so the mixing time will be lower. Since the distance

can never increase, the chains must have converged by the time $n-1$ such events have occurred.

Let X be a binomially distributed random variable with t trials and $p = \frac{2}{n^2}$. The number of times the distance has contracted by time t is given by $\max\{X, d(X_0, Y_0) - 1\}$. Since the maximum possible value for $d(X_0, Y_0)$ is n, by time $\frac{n^2(n-1)}{2}$, the chains have converged with probability at least $\frac{1}{2}$. If we run $\lceil \lg(\varepsilon^{-1}) \rceil$ independent trials, then the probability that the chains have not coupled by time $\frac{n^2(n-1)}{2}\lceil \lg(\varepsilon^{-1}) \rceil$ is at most ε. Therefore, by Lemma 2.13,

$$\tau(\varepsilon) \leq \frac{n^2(n-1)}{2}\lceil \lg(\varepsilon^{-1}) \rceil,$$

as required. □

In fact, in this situation it is straightforward to sample exactly from the uniform distribution without the requirement of a Markov chain algorithm. Diaconis, Graham and Holmes give an algorithm for sampling from the set of permutations in this case, and also a formula for the number of permutations [9, Lemma 3.1]. Scanning the columns of the adjacency matrix from right to left, we select a valid edge (that is, one that shares no endpoints with the edges already selected) uniformly at random from each column in turn. No matter which edge is selected, there are exactly the same number of configurations of the remainder of the graph. Continuing until one edge has been selected from each column yields a random sample from perfect matchings.

1. Let the initial matching $M = \emptyset$.

2. For each i from n to 1, in reverse order: select an edge adjacent to v_i uniformly at random that does not share an endpoint with any existing edge in M, and add it to M.

This method is sufficient to allow sampling of permutations of singly truncated data, but this approach does not appear to extend to the more general cases.

Chapter 7

Concluding remarks

In the preceding chapters, we have considered a number of Markov chains for sampling matchings, independent sets, and proper graph colourings.

The Markov chain $\mathcal{M}_{IDS}$ we defined for sampling matchings can be shown to be rapidly mixing for larger values of λ than we can achieve by applying the Dyer-Greenhill chain for independent sets directly to the line graph $\mathcal{L}(G)$. However, the bound we obtain on the the mixing time of our chain is $O(n^3)$. The Jerrum-Sinclair chain for sampling matchings mixes in time $O(nm\log n)$. Our bound is better than that for the Jerrum-Sinclair chain only if $m = \Omega(n^2/\log n)$. Since our chain requires that Δ is bounded, however, we have $m = O(n)$. As it stands, our chain performs no better than the Jerrum-Sinclair chain.

It may be possible to give a linear bound on the probability of change for $\mathcal{M}_{IDS}$. Such a bound would allow us to use Theorem 3.3 to show that $\mathcal{M}_{IDS}$ mixes in time $O(n^2)$. We may also be able to use a more complex analysis of $\mathcal{M}_{IDS}$, considering longer paths than we did in Chapter 4. In practice, attempts to do this required further restrictions on the graph, and became too complicated for what would be only a small improvement in the value of λ for which the chain mixes. A more plausible way to approach this chain would be variable-length path coupling [21], which we have not generally considered.

We have adapted the Jerrum-Sinclair chain for sampling matchings to sample independent sets in claw-free graphs. We have shown that our chain $\mathcal{M}_{CF}$ mixes in time $O(\Delta n^2 \log n)$. We could easily drop the requirement that Δ is bounded, in which case the mixing time is $O(n^3 \log n)$, given the trivial observation that $\Delta \leq n$ (assuming there are no self-loops in G). This is comparable to the mixing time of the Jerrum-Sinclair chain for dense graphs where $m = \Theta(n^2)$. Also, as we noted in Chapter 5, we can

eliminate the factor of Δ from the bound on the mixing time of $\mathcal{M}_{CF}$ if G contains no 4-cycles.

We have shown that the Jerrum-Sinclair chain for sampling matchings and our chain $\mathcal{M}_{CF}$ for sampling independent sets in claw-free graphs are rapidly mixing on lattice graphs, in time $O(nm)$ and $O(n^2)$ respectively. This is an improvement of $\log n$ over the bounds achieved for general graphs (and general claw-free graphs) using canonical paths. The crucial fact that we use is that the volume of the ball of radius ℓ around any vertex grows faster than the size of its boundary, by a factor of ℓ. In principle, our results apply to any graph for which this is true. When we considered the Jerrum-Sinclair chain, we considered only the two-dimensional square lattice. Our proof of rapid mixing of $\mathcal{M}_{CF}$ is more general, and can be applied to any claw-free graph where we have appropriate bounds on the volume and boundary of the ball.

Diaconis, Graham and Holmes claimed that the Diaconis chain $\mathcal{M}_{PM}$ mixes in time $O(n^2 \log n)$ for graphs corresponding to doubly truncated data [9]. However, we have shown that there are counterexamples for which the mixing time is exponential. We have managed to show that the mixing time is sub-exponential in certain cases, which do encompass many instances of doubly truncated data. It is unusual, although not unknown, for a natural problem to have a complexity of $\Theta(n^{\log n})$. It is therefore quite possible that $\mathcal{M}_{PM}$ has polynomial mixing time for these cases, although we have been unable to find a proof. On the other hand, if our class of graphs is genuinely a case with super-polynomial mixing time, there may be another class between ours and the trivial singly-truncated case, for which the mixing time is polynomial.

Diaconis, Graham and Holmes also showed that a variation of $\mathcal{M}_{PM}$ which allows only transitions between adjacent rows and columns of the adjacency matrix is connected on the class of graphs with monotonic adjacency matrices [9, Lemma 3.4]. This may offer a starting point for an alternative proof of mixing. They observed that in some cases, such as the quasar data presented by Efron and Petrosian [16], the number of transitions required is equal to Kendall's tau distance. However, this is not the case in general, and those cases where the number of transitions is larger may present difficulties. We considered this subset of transitions on more general graphs, but were unable to show connectedness.

The method we use to recursively decompose cycles in the proof of Theorem 6.9 is unusual. We can actually find parallel edges that divide a cycle into two parts of any length we choose (provided both parts are even length). Our general strategy therefore seems to cover most plausible ways of processing cycles. In order to recover

the start and end states from our encoding, we need to know the set of edges P^*. Any canonical paths argument that shows polynomial mixing time will either have to show that the number of extra edges in P^* is bounded above by a constant, or use a different encoding. In the singly truncated case, for example, we could show that there are canonical paths such that successive sets of parallel edges always share an edge in common, and so P^* only ever contains a single edge that is not in either I or F.

In order to show rapid mixing of $\mathcal{M}_{IDS}$, we required a slightly more general version of Theorem 2.15. Our more general version is Theorem 3.1, which we proved in Chapter 3. We also showed that when we have a coupling for which we can find a linear bound on the probability of a change in distance, we can obtain a bound on the mixing time of $O(n^2)$. We were unable to apply this to $\mathcal{M}_{IDS}$, but we can apply it to Dyer and Greenhill's original chain for sampling independent sets. As such, we have a quadratic bound on the mixing time of $\mathcal{M}_{DG}$ for $\lambda = \frac{2}{\Delta-2}$. Dyer and Greenhill's own argument in the boundary case gave a mixing time of $O(n^2 \log n)$.

We also applied Lemma 3.3 to the problem of sampling 2Δ-colourings of graphs using a simple single-site heat-bath chain, and obtained a bound of $O(n^2)$ on the mixing time of this chain. There are other chains for sampling 2Δ-colourings for which the mixing time is $O(n \log n)$ or $O(n^2)$. However, these chains are more complicated to implement. While it is possible to use comparison methods to bound the mixing time of the single-site chain from these more complex chains, the corresponding worsening of the bound means that comparison tends to give a mixing time worse than n^2.

Bibliography

[1] D. J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Lecture Notes in Mathematics*, 986:243–297, 1983.

[2] I. Bezáková, D. Štefankovič, V. V. Vazirani, and E. Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 900–907, New York, NY, USA, 2006. ACM Press.

[3] J. A. Bondy and U. S. R. Murty. *Graph Theory with Applications*. Elsevier Science Publishers Ltd., Essex, UK, 1976.

[4] M. Bordewich and M. Dyer. Path coupling without contraction. *Journal of Discrete Algorithms*, 5(2):280–292, 2007.

[5] A. Z. Broder. How hard is it to marry at random? (on the approximation of the permanent). In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 50–58, New York, NY, USA, 1986. ACM Press.

[6] A. Z. Broder. Errata to "how hard is it to marry at random? (on the approximation of the permanent)". In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, page 551, New York, NY, USA, 1988. ACM Press.

[7] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 223–231, Washington, DC, USA, 1997. IEEE Computer Society.

[8] A. Czumaj and M. Kutylowski. Delayed path coupling and generating random permutations via distributed stochastic process. *Random Structures and Algorithms*, 17(3-4):238–259, 2000.

[9] P. Diaconis, R. Graham, and S. P. Holmes. Statistical problems involving permutations with restricted positions. In *State of the Art in Probability and Statistics*, pages 195–222, 2001.

[10] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, 21(4):2131–2156, 1993.

[11] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[12] M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin. Markov chain comparison. *Probability Surveys*, 3:89–111, 2006.

[13] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colourings. *Random Structures and Algorithms*, 13(3–4):285–317, 1998.

[14] M. Dyer and C. Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.

[15] J. Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

[16] B. Efron and V. Petrosian. Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association*, 94(447):824–834, 1999.

[17] A. Frieze and E. Vigoda. A survey on the use of Markov chains to randomly sample colorings. In G. Grimmett and C. McDiarmid, editors, *Combinatorics, Complexity and Chance*. Oxford University Press, Oxford, UK, 2007.

[18] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, NY, USA, 1979.

[19] L. A. Goldberg, R. Martin, and M. Paterson. Strong spatial mixing for lattice graphs with fewer colours. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 562–571, Washington, DC, USA, 2004. IEEE Computer Society.

[20] D. Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31:95–106, 1975.

[21] T. P. Hayes and E. Vigoda. Variable length path coupling. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 103–110, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

[22] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

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

[24] M. Jerrum. *Counting, Sampling and Integrating: Algorithms and Complexity*. Lectures in Mathematics. ETH Zürich. Birkhäuser, Basel, Switzerland, 2003.

[25] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.

[26] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51(4):671–697, 2004.

[27] M. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43(2-3):169–188, 1986.

[28] V. S. A. Kumar and H. Ramesh. Coupling vs conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.

[29] W. Lipski and F. P. Preparata. Efficient algorithms for finding maximum matchings in convex bipartite graphs and related problems. *Acta Informatica*, 15(4):329–346, 1981.

[30] M. Luby and E. Vigoda. Approximately counting up to four. In *Twenty-ninth Annual ACM Symposium on Theory of Computing*, pages 682–687, New York, NY, USA, 1997. ACM Press.

[31] M. Luby and E. Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures and Algorithms*, 15(3–4):229–241, 1999.

[32] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, pages 148–188, 1989.

[33] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, UK, 2005.

[34] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15:1–25, 2002.

[35] C. H. Papadimitriou. *Computational Complexity*. Addison Wesley, Reading, MA, USA, 1994.

[36] V. V. Petrov. On lower bounds for tail probabilities. *Journal of Statistical Planning and Inference*, 137(8):2703–2705, 2007.

[37] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. In *Proceedings of the Third Latin American Symposium on Theoretical Informatics*, pages 292–304, London, UK, 1998. Springer-Verlag.

[38] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[39] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.

[40] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979.

[41] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.

[42] J. van den Berg and R. Brouwer. Random sampling for the monomer-dimer model on a lattice. *Journal of Mathematical Physics*, 41(3):1585–1597, 2000.

[43] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.

[44] D. Weitz. *Mixing in Time and Space for Discrete Spin Systems*. PhD thesis, UC Berkeley, 2004.

[45] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, UK, 1991.