UNIVERSITY OF
LIVERPOOL

On Systematic Scan

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor in Philosophy by
Kasper Pedersen.

First Supervisor: Prof. Leslie Ann Goldberg
Second Supervisor: Dr. Paul W. Goldberg

Department of Computer Science
The University of Liverpool
January, 2008

Preface

This thesis is predominantly my own work and the sources from which material is drawn are identified within. This is a brief summary of these.

Chapters 1 and 2 contain introductory material and a literature survey drawing from the works of several different authors. Furthermore Chapter 2 contains definitions used throughout this thesis, some of which are taken from Weitz [55].

Chapter 3 is based on a paper [47] published in MFCS 2007. The bibliographical details of the paper are:

- Kasper Pedersen. Dobrushin conditions for systematic scan with block dynamics. In Luděk Kučera and Antonín Kučera, editors, *MFCS*, volume 4708 of *Lecture Notes in Computer Science*, pages 264–275. Springer, Berlin, 2007.

Chapter 3 furthermore contains two proofs of theorems by Weitz which are outlined in Weitz [55].

Chapter 4 is based on a paper [48] submitted for publication. The bibliographical details of the paper are:

- Kasper Pedersen. On systematic scan for sampling H-colourings of the path. arXiv:0706.3794 (submitted), 2007.

Chapter 5 is based on a paper [38] submitted for publication. The paper is joint work with Markus Jalseniuis and both authors made equal contributions to the preparation of that paper. The bibliographical details of the paper are:

- Markus Jalsenius and Kasper Pedersen. A systematic scan for 7-colourings of the grid. arXiv:0704.1625 (submitted), 2007.

Abstract

In this thesis we study the mixing time of systematic scan Markov chains on finite spin systems. A systematic scan Markov chain is a Markov chain which updates the sites in a deterministic order and this type of Markov chain is often seen as intuitively appealing in terms of implementation to scientists conducting experimental work. Until recently systematic scan Markov chains have largely resisted analysis and a gap in the parameters that imply rapid mixing has developed between systematic scan Markov chains and the more frequently studied random update Markov chains. We reduce this gap in this thesis by improving the parameters for which systematic scan mixes when applied to several well-known spin systems.

The main contribution of this thesis is the introduction of a new technique for proving rapid mixing of systematic scan Markov chains. It is known that, in a single-site setting, the mixing time of systematic scan can be bounded in terms of the influence that sites have on each other. We generalise this technique for bounding the mixing time of systematic scan to block dynamics, a setting in which a (constant size) set of sites are updated simultaneously. In particular we introduce a parameter corresponding to the maximum influence *on* any site and show that if this parameter is sufficiently small, then the corresponding systematic scan Markov chain mixes rapidly.

We present several applications of this new proof technique. In particular we show that systematic scan mixes rapidly on spin systems corresponding to proper q-colourings of (1) general graphs, (2) trees, and (3) the grid for improved parameters than were previously known. We also obtain rapid mixing of systematic scan Markov chains for sampling H-colourings of the n-vertex path for a restricted family of H using this technique. The H-colouring result is extended to general graphs H by placing more restrictions on the scan and using path coupling, a well-established technique for bounding mixing times of Markov chains. Path coupling is also used to prove rapid mixing of a single-site systematic scan for sampling proper q-colourings of bipartite graphs.

Acknowledgements

I would like to extend a particular debt of gratitude to my main supervisor Leslie Ann Goldberg who has been an excellent supervisor both as a graduate and undergraduate student. In my research I have befitted immensely from Leslie's technical insights, attention to detail and ability to explain ideas and concepts in a clear and understandable fashion. Leslie has provided me with countless detailed and useful suggestions for ways to improve drafts of the papers that form the basis of this thesis. Furthermore, her friendly and enthusiastic personality has helped to make my time as a PhD student highly enjoyable.

I would also like to thank my second supervisor whilst at Liverpool University, Paul Goldberg, for useful conversations and suggestions.

I am grateful to both of my examiners, Mark Jerrum and Russell Martin, for the time and effort they put into examining this thesis and for several very helpful comments and suggestions for improvement during my viva.

I have been fortunate to have been part of two very good research groups during my time as a PhD student. For that I would like to thank the members of the Algorithms and Complexity research group at Warwick University and the members of the Complexity Theory and Algorithmics Group at Liverpool University. Both of these groups have formed an excellent environment in which to conduct research. Special thanks go to my officemates and good friends: Markus Jalsenius (with whom I had the pleasure of coauthoring a paper), Nick Palmer and Pattarawit Polpinit (A).

Last, but by no means least, I would like to thank my family and friends, especially my wife, my parents and my brother, for their support throughout my entire academic career.

Kasper Pedersen

Contents

List of Figures

List of Tables

Notation Glossary

Basic Notation

$\mathbb{Z}$	The set of integers.
$\mathbb{N}$	The set of positive integers including zero.
$\mathbb{R}$	The set of real numbers.
$\mathbb{R}_{\geq 0}$	The set of positive real numbers including zero.
$\mathbf{1}_{a=b}$	Indicator function taking value 1 if $a = b$ and value 0 otherwise.

Spin Systems and Markov Chains

V	The set of sites ($V = \{1, \ldots, n\}$).
C	The set of spins ($C = \{1, \ldots, q\}$).
Ω^+	The set of all configurations of a spin system ($\Omega^+ = C^V$).
π	The Boltzmann distribution of a spin system.
Ω	The set of legal configurations; configurations with positive measure in π.
x_i	The spin assigned to site i under some configuration $x \in \Omega^+$.
S_i	A pair of configurations differing only on the spin assigned to site i.
$\mathrm{d_{TV}}(\cdot, \cdot)$	The total variation distance between two probability distributions.
$\mathcal{M}_{\mathrm{RU}}$	A *random update Markov chain*. A random update Markov chain makes a transition by randomly selecting a subset of sites (from some specified set of subsets of V) and updating the spins assigned to the sites in the selected subset.
$\mathcal{M}_{\rightarrow}$	A *systematic scan Markov chain*. A systematic scan Markov chain makes a transition updating the subsets of sites (for some specified set of subsets of V) one at the time in a deterministic order.

Block Dynamics and Influence Parameters

Θ_k	A block with index k; $\Theta_k \subseteq V$.
Θ covers V	A set of blocks Θ covers the set of sites V if $\bigcup_k \Theta_k = V$.
$\partial\Theta_k$	The set of sites adjacent to Θ_k; $\partial\Theta_k$ is the *boundary* of Θ_k.
$x = y$ off Θ_k	The configurations x and y are assigned the same spin on all sites in $V \setminus \Theta_k$.
$x = y$ on Θ_k	The configurations x and y are assigned the same spin on all sites in Θ_k.
$P^{[k]}$	The transition matrix for updating block Θ_k.
$P^{[k]}(x, \cdot)$	The distribution on configurations resulting from applying $P^{[k]}$ to a configuration x.
$\Psi_k(x, y)$	A coupling of the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$.
$(x', y') \in \Psi_k(x, y)$	A pair of configurations (x', y') drawn from $\Psi_k(x, y)$.
$\rho^k_{i,j}$	The influence of site i on site j under the update of block Θ_k; $\rho^k_{i,j} = \max_{(x,y)\in S_i}\{\Pr_{(x',y')\in\Psi_k(x,y)} x'_j \neq y'_j\}$.
α	The maximum influence *on* any site in the graph; $\alpha = \max_k \max_{j\in\Theta_k} \sum_{i\in V} \rho^k_{i,j} w_i/w_j$ where w_i is a positive weight assigned to site i for each $i \in V$.

Chapter 1

Introduction

This thesis is concerned with the study of finite *spin systems*. A finite spin system is composed of a set of sites and a set of spins, both of which are finite. The sites are vertices of an underlying graph whose edges specify the interconnection between the sites. The underlying graph is assumed to be connected. A configuration of the spin system is an assignment of a spin to each site. If there are n sites and q available spins then this gives rise to q^n possible configurations, however some configurations may be illegal depending on the specification of the spin system. The specification of the system determines how spins interact with each other at a local level, such that different local configurations on a subset of the graph may have different relative likelihoods. In particular, for spin systems with so-called *hard-constraints* the specification states which pairs of spins are permitted to be assigned to adjacent sites and which pairs of spins are not. This interaction between sites specifies a well-defined probability distribution π (known as the Boltzmann distribution) on the set of all configurations of a spin system. Configurations with positive measure in π are said to be legal.

Many models, often originating from the field of statistical physics, fall under the general category of spin systems. As a simple, but important, example consider a spin system in which no two adjacent sites are permitted to be assigned the same spin. This spin system corresponds to the q-state anti-ferromagnetic Potts model at zero temperature, a frequently studied model in statistical mechanics. This spin system is also well-known in the field of theoretical computer science where a legal configuration of the system is commonly known as a *proper q-colouring* of the underlying graph. Several of the results presented in this thesis will be for this spin system, and when discussing proper q-colourings it is natural to refer to the spins as *colours*.

Another well-known example of a spin system is the *independent sets* model.

In the independent sets model each site is either "occupied" or "unoccupied" and in a legal configuration no two adjacent sites are allowed to be occupied. It is usual to assign a positive weight λ to each occupied site, and in this weighted setting the spin system is known as the *hard-core lattice gas model*. This spin system has been used as a model of gas in the field of statistical physics (Georgii [30] cited in Weitz [54]) and has also been used in the modeling of communication networks by Kelly [42].

A natural formalisation of spin systems with hard constraints is the H-colouring model. An H-colouring of a graph G is a *homomorphism* from G to some fixed graph H. The vertices of H correspond to spins and the edges of H specify which spins are allowed to be adjacent in an H-colouring of G. The H-colouring model is a natural generalisation of the proper colouring model since if H is the q-clique then an H-colouring of a graph is a proper colouring. H-colouring problems have attracted much interest from computer scientists and combinatorialists alike and much progress has been made. In fact, Hell and Nešetřil [37] gave a complete characterisation of graphs H for which the decision problem of determining whether a given graph has an H-colouring for a specific H is NP-complete. They showed that if H has a loop or is bipartite then the problem is in P, and that the problem is NP-complete for any other fixed H. A complete dichotomy is also known for the problem of counting the number of H-colourings of a given graph. This counting problem is of natural interest to combinatorialists, and we will be interested in studying problems closely related to counting in this thesis. This dichotomy is due to Dyer and Greenhill [24] who showed that if H has at least one *nontrivial component* then the counting problem is complete for the complexity class #P. Otherwise it is in P. A trivial component is a connected component which is either a complete graph with all loops present, or a complete bipartite graph with no loops present. The complexity class #P was introduced by Valiant [52] in 1979 and it contains enumeration problems. For a more detailed description of this complexity class see Jerrum [40]. Dyer and Greenhill furthermore showed that the same dichotomy holds even when the underlying graph is of bounded degree. This is an interesting observation since in many physical applications the underlying graph tends to be of low degree. Interestingly the above characterisation for the decision problem does not hold for bounded degree graphs as was shown by Galluccio, Hell and Nešetřil [29]. Despite the hardness of exactly counting the number of H-colourings of a graph, it remains possible to *approximately* count the number of H-colourings as we will discuss subsequently.

For a given spin system it is of interest to sample from the probability distri-

bution π, especially when π is uniform over the set of legal configurations Ω of the spin system. In statistical physics this interest is due to the connection that π has with various equilibrium properties of a spin system. In theoretical computer science much of the reason for interest in the sampling problem is the, now well-established, connection between (nearly) uniform sampling and approximate counting established by Jerrum, Valiant and Vazirani [41]. They showed that the (nearly) uniform sampling problem and the approximate counting problems are equally hard for a subclass of counting problems which satisfy a property called self-reducibility. This subclass contains many interesting instances of counting problems, notably proper q-colourings. Specifically, the problem of uniform sampling reduces to the problem of approximately counting the number of elements in Ω and vice versa for all self-reducible counting problems. For an exposition account of these developments see for example the book by Jerrum [40] or the survey paper by Dyer and Greenhill [23]. Both of these publications focus on some of the most well-studied models in computer science, such as proper q-colourings and independent sets, and many papers concerned with studying techniques for sampling proper colourings or independent sets have been motivated by this explicit connection between sampling and counting. The first counting-to-sampling reduction applicable to general H-colourings was due to Dyer, Goldberg and Jerrum [17] although currently no completely general sampling-to-counting reduction is known. Hence, if there exists a polynomial time (in the number of sites of the underlying graph) algorithm for sampling from the (near) uniform distribution of H-colourings of a graph then there also exists a polynomial time algorithm for approximately counting the number of H-colourings of that graph. With this result in mind we will focus on the problem of sampling from π for a given spin system.

Given a spin system, the problem of sampling from π is a challenging task. Goldberg, Kelk and Paterson [32] studied the complexity of this sampling problem for H-colourings in the case when π is uniform over Ω and showed that if H has no nontrivial components then the sampling problem is intractable in a complexity-theoretic sense. That is, they prove that there is unlikely to be any algorithm that can efficiently obtain a sample from π (this is known as a Polynomial Almost Uniform Sampler) by reducing the problem of approximately counting independent sets in bipartite graphs, which in turn is complete with respect to approximation preserving reductions for a logically-defined subclass of #P (see Dyer, Goldberg, Greenhill and Jerrum [15] for results about this complexity class), to the problem of sampling from the (near) uniform distribution of

H-colourings. This does, however, not rule out the possibility of sampling from the uniform distribution of general H-colourings of more restricted graphs G.

As the task of sampling from π is computationally difficult it is often the case that the only feasible method of carrying out this task is by simulating some suitable random dynamics converging to π. Ensuring that such a dynamics converges to π is generally straightforward, but obtaining good upper bounds on the number of steps required for the dynamics to become sufficiently close to π is a much more difficult problem. One of the most common type of dynamics used is a Markov chain. A Markov chain is a stochastic process whose states (in our case) are the set of configurations of the given spin system with positive measure in π. By construction of the Markov chain it is generally straightforward to ensure that it converges to π, however providing good upper bounds on the rate of convergence, known as the *mixing time* of the Markov chain, is a much more difficult task. For this sampling method to be feasible we need to ensure that the Markov chain converges to π in a polynomial number of steps. Due to a lack of theoretical convergence results, scientists conducting experiments by simulating such dynamics are at times forced to "guess" (using some heuristic methods) the number of steps required for their dynamics to be sufficiently close to the desired distribution. Cowles and Carlin [9] give a comprehensive review of some diagnostic tools used to empirically determine these convergence rates and include some examples from applications in the field of bio-statistics. One immediate problem, which is pointed out by Cowles and Carlin, with many convergence diagnostics is that they might prematurely claim convergence of the dynamics and another is that by continuously monitoring the dynamics one may implicitly introduce a conditioning that can in turn create a bias in the sampling procedure (see Cowles, Roberts and Rosenthal [10]). The negative effect these and other issues have on the effectiveness of practical applications can be greatly reduced using more sophisticated diagnostic tools, however the existence of good analytical bounds on the convergence rates would eliminate the need for such techniques to be employed in the first place. By establishing rigorous bounds on the mixing time of these Markov chains, computer scientists can provide underpinnings for this type of experimental work and also allow a more structured approach to be taken.

Analysing the mixing time of Markov chains for sampling from π for various spin systems is a well-studied area in theoretical computer science and as a result of this interest there is a substantial body of literature concerned with inventing Markov chains for sampling from π and providing upper bounds on their mixing

times. We now briefly survey some of the contributions made. When the spin system corresponds to proper q-colourings of a graph with maximum vertex-degree Δ and π is uniform over the set of proper colourings then Jerrum [39], and independently Salas and Sokal [50], showed that a simple Markov chain mixes in $O(n \log n)$ updates when $q > 2\Delta$. This Markov chain makes transitions by selecting a site v and a colour[1] c uniformly at random, and then recolouring site v to c if doing so results in a proper q-colouring of the graph. By considering a more complicated Markov chain Vigoda [53] was able to weaken the restriction on q to $q > (11/6)\Delta$ being sufficient for proving mixing in $O(n \log n)$ updates. This remains the least number of colours required for rapid mixing of a Markov chain for uniformly sampling q-colourings of general graphs, however the number of colours can be further reduced for restricted families of graphs. For example, in the important case when the underlying graph is the grid then Goldberg, Martin and Paterson [33] gave a hand-proof that $q = 7$ colours are sufficient for mixing in $O(n \log n)$ updates by establishing a condition called "strong spatial mixing" which in turn implies rapid mixing (see Dyer, Sinclair, Vigoda and Weitz [26]). Achlioptas, Molloy, Moore and van Bussel [1] further showed that $q = 6$ colours are sufficient for a Markov chain for proper colourings of the grid to mix in $O(n \log n)$ updates using a computer-assisted proof. As a final example for proper q-colourings Martinelli, Sinclair and Weitz [46] showed that $q = \Delta + 2$ colours are sufficient for $O(n \log n)$ mixing when the underlying graph is a tree, improving a related result by Kenyon, Mossel and Peres [43].

When the spin system corresponds to independent set configurations with parameter λ then the condition $\lambda < \frac{2}{\Delta - 2}$ is sufficient for $O(n \log n)$ mixing as shown by Dyer and Greenhill [25] and independently Luby and Vigoda [45] (although the latter result is restricted to triangle-free graphs). When $\Delta \leq 4$ these results include the $\lambda = 1$ case which is of special interest to computer scientists since it corresponds to sampling from the uniform distribution on independent sets of the graph. Weitz [56] has recently given a completely different algorithm, namely a deterministic algorithm with polynomial running time, which improves the condition on λ to $\lambda < (\Delta - 1)^{\Delta - 1}/(\Delta - 2)^{\Delta}$. This notably includes the $\lambda = 1$ case for $\Delta = 5$. An interesting aspect of work carried out on the independent sets model is that, as well as the aforementioned positive results regarding the mixing times of various Markov chains, a number of *negative* results are known as we will now discuss. When $\Delta \geq 6$ and $\lambda = 1$ then Dyer, Frieze and Jerrum [14] have shown

[1]Recall that we use the term *colour* rather than *spin* when discussing spin systems corresponding to proper colourings.

that there exists a bipartite graph G_0 such that any so-called *cautious* Markov chain on independent set configurations of G_0 has (at least) exponential mixing time (in the number of sites of G_0). A Markov chain is said to be cautious if it is only allowed to change the state of a constant fraction of sites at the time. This negative result was generalised to H-colourings by Cooper, Dyer and Frieze [8]. Their result applies to graphs H that are either bipartite or have at least one loop present, and is not a complete graph with all loops present (observe that for such an H the decision problem is in P and the counting problem is in #P as discussed above). In particular this result guarantees the existence of a Δ-regular graph G_0 (with Δ depending on H) such that any cautious Markov chain on the set of H-colourings of G_0, and with uniform stationary distribution, has a mixing time that is at least exponential in the number of sites of G_0.

While much is understood about the mixing times of Markov chains for sampling from π, the types of Markov chains frequently studied by computer scientists do not always correspond to the types of dynamics used in experimental work. Most of the Markov chains previously studied make transitions by randomly selecting a set of sites (often just a single site) and updating the spins assigned to those sites according to some well-defined distribution induced by π. We call this type of chain a *random update Markov chain* and point out that all the positive results described above are for random update Markov chains. The mixing time of a random update Markov chain is measured in the number of *updates* required in order for the Markov chain to mix. An alternative to random update Markov chains is to construct a Markov chain that cycles through and updates the sites (or subsets of sites) in a deterministic order. We call this a *systematic scan Markov chain* (or *systematic scan* for short). The mixing time of a systematic scan Markov chain is measured in the number of *scans* of the graph required to mix and throughout this thesis it holds that one scan of the graph takes $O(n)$ updates. It is important to note that systematic scan remains a random process since the method used to update the colour assigned to the selected set of sites is a randomised procedure drawing from some well-defined distribution induced by π. Systematic scan may be more intuitively appealing that random update Markov chains in terms of implementation, however until recently this type of dynamics has largely resisted analysis when applied to spins systems with hard constraints. Dynamics that make deterministic choices about about the order in which sites are updated have however been used in practical applications. In a study of the effect the rules for selecting sites for update has on the convergence rates Fishman [27] outlined five plans for selecting the update order, three of

which were deterministic rules, as well as giving some practical comparisons. A practical comparison is also given by Roberts and Sahu [49] for the problem of sampling from a Gaussian distribution with applications in image analysis. They showed that for two classes of sampling problems a deterministic strategy is better than a random update strategy. However they also gave examples of instances from outside those classes where random update performs better. An example that is more combinatorial in nature and as such is closer to the applications we will consider in this thesis is Diaconis and Ram [11] who studied systematic scan in the context of generating random elements of a finite group and successfully bounded the number of scans required to mix. This thesis is concerned with studying the problem of sampling from π for any given spin system by simulating systematic scan Markov chains, and especially with bounding the mixing times of these chains.

Only few results providing bounds on the mixing time of systematic scan Markov chains for sampling from π exist in the literature and almost all of them focus on proper q-colourings of bounded degree graphs. For general graphs, systematic scan is known to mix in $O(\log n)$ scans whenever $q > 2\Delta$ where Δ is the maximum vertex-degree of the graph. This result is obtained by studying the influences that the sites have on each other and is due to Dyer, Goldberg and Jerrum [18]. This approach also gives a mixing time of $O(n^2)$ scans in the $q = 2\Delta$ case. In Chapter 3 we improve the mixing time of systematic scan for general graphs in the $q = 2\Delta$ case to $O(\log n)$ scans. If the underlying graph is bipartite then a systematic scan mixes in $O(\log n)$ scans whenever $q > f(\Delta)$ where $f(\Delta) \to \beta\Delta$ as $\Delta \to \infty$ and $\beta \approx 1.76$. This result is obtained by a careful construction of the metric used in the path coupling construction and is due to Bordewich, Dyer and Karpinski [4]. When considering tree graphs, it is known that systematic scan mixes in $O(\log n)$ scans whenever $q > \Delta + 2\sqrt{\Delta - 1}$ and in $O(n^2 \log n)$ scans whenever $q = \Delta + 2\sqrt{\Delta - 1}$ is an integer; see e.g. Hayes [36] or Dyer, Goldberg and Jerrum [19]. In Chapter 3 we will further reduce the number of colours required to prove rapid mixing for systematic scan on trees. Furthermore, Dyer, Goldberg and Jerrum [20] have shown that a systematic scan for proper 3-colourings of the n-vertex path mixes in $\Theta(n^2 \log n)$ scans when considering a systematic scan that updates one site at the time using the *Metropolis* update rule. In the same paper it is also proved that systematic scan for general H-colourings of the n-vertex path mixes in $O(n^5)$ scans for any fixed H and that a random update Markov chain for H-colourings of the n-vertex path mixes in $O(n^5)$ updates. The authors suggest, however, that both of these bounds are un-

likely to be tight and we will improve them to $O(\log n)$ and $O(n \log n)$ respectively in Chapter 4.

A comparison between the known results for systematic scan and random update Markov chains clearly reveals a gap between the parameters that imply mixing in the two cases. When analysing the mixing time of random update Markov chains one often only needs to study the effect of updating one randomly selected site starting from two configurations that are identical except on the spin assigned to a single site. This relatively simple situation is in contrast to the task faced when analysing a systematic scan Markov chain in which case one needs to study the effect of one entire scan of the graph and hence keep track of all intermediate configurations of the chain. Analytically this is clearly a much more difficult task. It is worth observing at this point that there is one spin system for which systematic scan is known to mix faster than any random update Markov chain. This is the relatively uninteresting case when considering q-colourings of a graph with no edges. In this case it is known (see Dyer, Goldberg, Greenhill, Jerrum and Mitzenmacher [16] for a simple proof of this fact) that $\Omega(n \log n)$ is a lower bound on the number of updates any random update Markov chain needs to make before mixing, whereas a systematic scan clearly mixes in just one scan which corresponds to n updates. In this thesis we reduce the gap between the parameters that imply mixing of systematic scan and random update Markov chains by weakening the conditions required for mixing of systematic scan for several spin systems. We achieve this by introducing a new technique, based on Dobrushin uniqueness, for proving rapid mixing of systematic scan for general spin systems and applying this technique to specific spin systems such as proper colourings of general graphs. We will also use path coupling on some restricted families of graphs to improve the conditions for rapid mixing of systematic scan.

When analysing the mixing time of Markov chains it can be useful to consider chains that make use of *block dynamics*. A block dynamics Markov chain is permitted to change the spin at more than one site during each step of the process, provided that the number of sites that are being updated at each step is not "too large" in an appropriate sense. One reason for studying block dynamics rather than single-site dynamics is that in some cases single-site chains do not yield to analysis whilst block dynamics do, as we shall see. Block dynamics is not a new concept and it was used in the mid 1980s by Dobrushin and Shlosman [13] in their study of conditions that imply uniqueness of the Gibbs measure of a spin system, a topic closely related to studying the mixing time of Markov chains (see for example Weitz's PhD thesis [54]). Roberts and Sahu [49] also considered

the concept of block updates in their (more practical) comparisons of various update strategies for sampling from Gaussian distributions and concluded that making use of block updates could often increase the convergence rate of such an algorithms, however they also gave examples of block dynamics that converged slower than their single-site counterparts. More recently, block dynamics has been used by Weitz [55] when, in a generalisation of the work of Dobrushin and Shlosman, studying the relationship between various *influence parameters* (also in the context of Gibbs measures) within spin systems and using the influence parameters to establish conditions that imply mixing. Dyer et al. [26] have also used a block dynamics in the context of analysing the mixing time of a Markov chain for proper colourings of the square lattice. Both of these papers consider a random update Markov chain, however several of the ideas and techniques carry over to the analysis of systematic scan as we shall see. We explore the analysis of systematic scan Markov chains making use of block dynamics in this thesis. In particular we give a new condition based on bounding the influence *on* a site that implies $O(\log n)$ mixing of systematic scan Markov chains using block dynamics on finite spin systems. Applications of this condition give rapid mixing of systematic scan for proper q-colourings of (1) general graphs, (2) trees, and (3) the grid for improved parameters than were previously known. We also apply the condition to H-colourings of the n-vertex path and obtain rapid mixing of systematic scan for a restricted family of graphs. We extend the H-colouring result to general graphs H by placing more restrictions on the scan and using a well-established technique for bounding mixing times of Markov chains called path coupling [5].

While using block dynamics in order to facilitate a better analysis of systematic scan Markov chains is very much a central theme in this thesis we also consider a few single-site dynamics. One of these chains is a chain for sampling proper q-colourings of a tree and another is for sampling proper q-colourings of general bipartite graphs. Both of these results have since been matched or improved by new research in the field, although the single-site systematic scan for sampling proper q-colourings of a bipartite graph that we present remains the only single-site systematic scan Markov chain that mixes in $O(\log n)$ scans when $q = 2\Delta$ in the $\Delta = 3$ and $\Delta = 4$ cases. Note that the grid, which is of significant importance, is included in this result.

1.1 Summary of Results

We now give a brief description of the results to be presented in this thesis.

A Dobrushin Condition for Rapid Mixing of Systematic Scan with Block Dynamics

It is known that, in a single-site setting, the mixing time of systematic scan can be bounded in terms of the influences sites have on each other (see for example Dyer et al. [18]). Some known theorems are of the form: "If the influence on a site is small then a systematic scan Markov chain mixes in $O(\log n)$ scans." This is similar to a condition proved by Dobrushin [12] (although not in the context of studying the mixing time of Markov chains or systematic scan) and we refer to a condition of this form as a *Dobrushin condition*. We generalise this technique for bounding the mixing time of systematic scan to block dynamics, a setting in which a (constant size) set of sites are updated simultaneously. In particular we define an influence parameter α, corresponding to the maximum influence on any site, and show that if $\alpha < 1$ then the corresponding systematic scan Markov chain mixes rapidly. In fact the condition will apply regardless of the specific scan order as we will discuss in more details in due course. As applications of this proof technique we prove $O(\log n)$ mixing of systematic scan (for any scan order) for proper q-colourings of a general graph with maximum vertex-degree Δ when $q \geq 2\Delta$ by considering a chain making heat-bath updates of both endpoints of a single edge at the time. We also apply the method to reduce the number of colours required in order to obtain mixing in $O(H)$ scans for systematic scan on trees, with height H, using some suitable heat-bath block updates.

Sampling H-colourings of the Path

We then considerably widen the setting to general H-colourings but at the expense of restricting the underlying graph of the spin system to the path. We show that systematic scan for sampling from the uniform distribution on H-colourings of the n-vertex path mixes in $O(\log n)$ scans for any fixed H using some suitable block updates. This is a significant improvement over the previous bound on the mixing time which was $O(n^5)$ scans due to Dyer et al. [20]. Note, however, that the Markov chain in Dyer et al. [20] is a single-site chain, whereas our chain uses block dynamics. It is of special interest to observe that we can use block updates to obtain a mixing time that is faster than a known lower bound for

3-colourings of the path that applies to single-site chains. Furthermore we use the influence parameter α to show that for a slightly more restricted family of H (where any two vertices are connected by a 2-edge path) systematic scan also mixes in $O(\log n)$ scans for any scan order. Finally, for completeness, we show that a random update Markov chain mixes in $O(n \log n)$ updates for any fixed H, improving the previous bound on the mixing time which was $O(n^5)$ updates.

Sampling 7-colourings of the Grid

An important problem is to sample from the uniform distribution of proper q-colourings of the grid using as few colours as possible. We consider the $q = 7$ case using systematic scan. The systematic scan Markov chain that we present cycles through subsets consisting of 2×2 sub-grids and updates the colours assigned to the sites using the heat-bath update rule. We give a computer-assisted proof that this systematic scan Markov chain mixes in $O(\log n)$ scans, where n is the size of the rectangular sub-grid. This is the first time that the mixing time of a systematic scan Markov chain for proper colourings of the grid has been shown to mix with less than 8 colours. We also give partial results that underline the challenges of proving rapid mixing of a systematic scan Markov chain for sampling 6-colourings of the grid by considering the possibilities of updating 2×3 and 3×3 sub-grids.

Single-site Systematic Scan for Bipartite Graphs

It remains of natural interest to study Markov chains that make single-site updates. We consider a systematic scan Markov chain that scans each colour class of bipartite graph in turn and show, using path coupling, that it mixes in $O(\log n)$ scans whenever $q \geq 2\Delta$. This result has since been improved by Bordewich et al. [4] for $\Delta \geq 9$ and matched for $5 \leq \Delta < 9$. It remains, however, the only single-site systematic scan that mixes in $O(\log n)$ scans whenever $q = 2\Delta$ and $\Delta \in \{3, 4\}$.

1.2 Plan of Thesis and Biographical Notes

In Chapter 2 we give precise definitions of spin systems and the mixing time of Markov chains. We go on to define the notation required state our conditions for mixing as well as stating our results and placing them in the context of known results in the field. Chapter 3 contains the proof of our condition for rapid mixing

of systematic scan with block dynamics as well as two immediate applications to spin systems corresponding to proper colourings of general graphs and trees. The material from Chapter 3 is published in Pedersen [47]. In Chapter 4 we study the mixing time of systematic scan for sampling from the uniform distribution of H-colourings of the n-vertex path. The material from Chapter 4 has been submitted for publication in Pedersen [48]. Chapter 5 is concerned with sampling from the uniform distribution of 7-colourings of the square grid. The material from Chapter 5 has been submitted for publication in Jalsenius and Pedersen [38] and is joint work with Markus Jalsenius. Both authors made equal contributions to the preparation of that paper. Chapter 6 is concerned with analysing the mixing time of a single-site systematic scan for sampling proper colourings of bipartite graphs. The material from Chapter 6 is unpublished.

Chapter 2

Preliminaries

In this chapter we set the basis for the work presented in this thesis. We give a formal definition of a spin system as well as introducing examples of specific spin systems that we will study in more detail. We go on to introduce important concepts relating to Markov chains and their mixing times, one of the main topics of this thesis. We then formally introduce the concepts of block dynamics and influence parameters. We conclude this chapter by stating the results to be proved in this thesis.

2.1 Spin Systems

Let $C = \{1, \ldots, q\}$ be the set of *spins* and $V = \{1, \ldots, n\}$ be the set of *sites*. The sites are vertices of a connected graph $G = (V, E)$ which is the underlying graph of the spin system. Both of the sets C and V will be finite throughout this thesis. We say that a pair of sites $i, j \in V$ are *adjacent* in the spin system if $(i, j) \in E$. A configuration of the spin system is an assignment of a spin to each site. We let $\Omega^+ = C^V$ be the set of all configurations of a spin system. If $x \in \Omega^+$ is a configuration and $j \in V$ is a site then x_j denotes the spin assigned to site j in configuration x. Adjacent sites interact locally making some sub-configurations more likely than others. In particular, the locality requirement is that the spin assigned to a site j may only depend on the spins assigned at sites adjacent to j. This interaction gives rise to a well-defined probability distribution π on the set of all configurations. Let $\Omega = \{x \in \Omega^+ \mid \pi(x) > 0\} \subseteq \Omega^+$ be the set of configurations with positive measure in π. We refer to Ω as the set of legal configurations.

Example 1. The spin system we will consider in most of our applications is

the q-state anti-ferromagnetic Potts model. This spin system has a set of q distinct spins and interactions between adjacent sites is *antiferromagnetic*, i.e., configurations in which adjacent sites are assigned unequal spins are favored. In particular the probability that the spin system is in a given configuration $x \in \Omega^+$ is given by

$$\pi(x) \propto \exp\left(-\beta \sum_{(i,j)\in E} \mathbf{1}_{x_i=x_j}\right)$$

where $0 \leq \beta \leq \infty$ is the inverse temperature and $\mathbf{1}_{x_i=x_j} = 1$ if and only if $x_i = x_j$. A case of special interest is the zero-temperature case (i.e., $\beta = \infty$) which introduces hard constraints, meaning that no configuration in which *any* pair of adjacent sites are assigned the same spin has positive measure in π. In theoretical computer science this spin system has been well-studied, as a legal configuration corresponds to a *proper q-colouring* of the underlying graph. A proper colouring of a graph is an assignment of a colour (spin) to each vertex (site) such that no to adjacent vertices are assigned the same colour. We also note that in the zero-temperature case π is uniform over the set of proper colourings and zero elsewhere.

Example 2. Another famous example is the hard core model (independent sets) which, in statistical physics, has been used as a model of lattice gasses [30]. This spin system consists of two spins $C = \{0, 1\}$ and we say that a site is "occupied" if it is assigned spin 1 and "unoccupied" if it is assigned spin 0. The specification of the system states that no occupied site may be adjacent to another occupied site. In the computer science literature, a configuration for which this condition holds is called an independent set of the underlying graph. If $\Omega \subseteq \Omega^+$ is the set of independent sets of the underlying graph for the given spin system then the measure of a given independent set $x \in \Omega$ is given by

$$\pi(x) \propto \lambda^{\sum_{i\in V} x_i}$$

where $\lambda > 0$ is the activity parameter (sometimes called the fugacity). For all remaining configurations $x \in \Omega^+ \setminus \Omega$ it holds that $\pi(x) = 0$. Observe that the sum $\sum_{i\in V} x_i$ is the number of sites in the independent set so if λ is big then independent sets with many occupied sits are favoured. Of particular interest to computer scientists is the $\lambda = 1$ case where π is uniform over all independent sets i.e., each independent set is equally probable in π.

Example 3. A natural generalisation of both of the two previous examples is

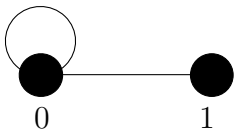

Figure 2.1. The graph describing the *independent sets* model. Sites assigned colour 0 are "unoccupied' and sites assigned 1 are "occupied".

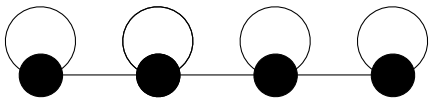

Figure 2.2. The graph describing the *Beach* model.

the H-colouring model. An H-colouring of a graph G is a *homomorphism* from G to some fixed graph H. The vertices of H correspond to spins and the edges of H specify which spins are allowed to be adjacent in an H-colouring of a graph. If H is the q-clique then an H-colouring of a graph is a proper colouring. Similarly H-colourings using the graph H from Figure 2.1 correspond to independent set configurations of a graph. Other well-known examples of H-colouring problems include the *Beach* model introduced by Burton and Steif [7] and the q-particle Widom-Rowlinson due to Widom and Rowlinson [57]. The graph corresponding to the Beach model is shown in Figure 2.2. The Beach model was originally introduced as an example of a physical system, with underlying graph $\mathbb{Z}^d$, which exhibits more than a single measure of maximal entropy when $d > 1$. The q-particle Widom-Rowlinson model is a model of gas consisting of q types of particles that are not allowed to be adjacent to each other. The graph corresponding to the $q = 4$ case is shown in Figure 2.3 where the center vertex represents empty sites and each remaining vertex represents a particle.

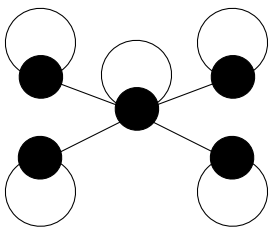

Figure 2.3. The graph describing the 4-particle *Widom-Rowlinson* model.

2.2 Markov Chains and Mixing Time

We are interested in sampling from the probability distribution π, a task that can be carried out by simulating a suitable (finite) Markov chain. A Markov chain $\mathcal{M}$ with state space $\mathcal{S}$ is a sequence of random variables $X_0, X_1, \ldots$ where $X_t \in \mathcal{S}$ for each $t \geq 0$ and which satisfies the following equality

$$\Pr(X_{t+1} = y \mid X_t = x_t, \ldots X_0 = x_0) = \Pr(X_{t+1} = y \mid X_t = x_t)$$

for all $t \geq 0$ and $x_0, x_1, \ldots x_t \in \mathcal{S}$. We consider the case when $\mathcal{S}$ is finite. For the subsequent discussion we do not assume that $\mathcal{S} = \Omega$ although this is our eventual purpose.

The transitions of a Markov chain are defined by a transition matrix P. In particular, P has the property that $P(x, y) = \Pr(X_{t+1} = y \mid X_t = x)$ for all pairs of states $(x, y) \in \mathcal{S} \times \mathcal{S}$. The transition matrix denotes the transition probabilities for a single step of the Markov chain. The t-step transition probabilities P^t of $\mathcal{M}$ are inductively defined by $P^t(x, y) = \sum_{x' \in \mathcal{S}} P^{t-1}(x, x')P(x', y)$ for $t > 0$ where we let $P^0(x, y) = \mathbf{1}_{x=y}$. Hence $P^t(x, y)$ is the probability that the Markov chain moves from state x to state y in exactly t transitions. We let $P^t(x, \cdot)$ be the distribution of the state that the chain is in after making t transitions starting from state $X_0 = x$.

We are interested in the convergence properties of Markov chains. A *stationary distribution* of a Markov chain is a probability distribution μ on $\mathcal{S}$ satisfying

$$\mu(y) = \sum_{x \in \mathcal{S}} \mu(x)P(x, y)$$

for each $y \in \mathcal{S}$. Informally, we can say that once a Markov chain reaches its stationary distribution no transition can change the distribution of the state that the chain is in. A Markov chain that satisfies the following two properties

- *irreducibility*: for all pairs of states $x, y \in \mathcal{S}$ there exists a positive integer t such that $P^t(x, y) > 0$; and
- *aperiodicity*: for all states $x \in \mathcal{S}$ it holds that $\gcd\{t : P^t(x, y) > 0\} = 1$

is said to be ergodic. It is a well-known result from classical Markov chain theory (see for example Aldous [2]) that an ergodic Markov chain has a unique stationary distribution. An ergodic Markov chain hence eventually "forgets" its initial state

and converges to its stationary distribution regardless of which state its starts from.

Given a spin system we can use an ergodic Markov chain to obtain a sample from π as follows. We construct an ergodic Markov chain $\mathcal{M}$ with state space Ω (the set of all legal configurations of the given spin system) such that its (unique) stationary distribution is π. Note that the set of states now corresponds to the set of legal configurations. We simulate $\mathcal{M}$ until the distribution on states is sufficiently close to π in an appropriate sense. Once the distribution on the states of $\mathcal{M}$ is sufficiently close to π we stop the simulation and return the current state of $\mathcal{M}$ as the sample. This type of algorithm is known as a Markov chain Monte Carlo algorithm.

Example 4. Arguably the simplest Markov chain is the *heat-bath Glauber dynamics*. We consider the spin system corresponding to proper q-colourings of a graph $G = (V, E)$ with maximum vertex-degree Δ. Let Ω be the set of all proper q-colourings of G. Recall from Example 1 that π is uniform over Ω in this case. We let Ω be the state space of the heat-bath Glauber dynamics and a transition from a configuration $x \in \Omega$ to $x' \in \Omega$ is made according to the following three step process

1. Select a site $i \in V$ uniformly at random.
2. Select a colour $c \in C_i$ uniformly at random where $C_i = C \setminus \{x_j : (i, j) \in E\}$ is the set of all colours that are not assigned to neighbours of site i.
3. Set $x'_i = c$ and $x'_j = x_j$ for each $j \neq i$.

The heat-bath Glauber dynamics is known to be ergodic provided that $q \geq \Delta + 2$ (Jerrum [39]) and furthermore π is the stationary distribution, which can be verified by observing that π is invariant with respect to the transition matrix P of the heat-bath Glauber dynamics. Since $P(x, y) = P(y, x)$ we have

$$\pi(x)P(x, y) = \pi(y)P(y, x) \tag{2.1}$$

and hence

$$\sum_x \pi(x)P(x, y) = \sum_x \pi(y)P(y, x) = \pi(y)$$

for any configuration $y \in \Omega$. Equation (2.1) is known as detailed balance and holds for so-called *time reversible* Markov chains. Since the heat-bath Glauber dynamics is ergodic it hence eventually converges to π regardless of its initial state.

As illustrated by the above example it is generally straight-forward to ensure, via the construction of the chain, that a Markov chain is ergodic with the desired stationary distribution. An important question that remains is how long we need to simulate a Markov chain for before reaching a distribution that is sufficiently close to stationary. In particular, for the Markov chain Monte Carlo method to be effective we need to ensure that the Markov chain converges in a number of steps that is polynomial in the size of the underlying graph. We call the number of transitions required to become sufficiently close to the stationary distribution of a Markov chain its *mixing time*. Recall that we denote the stationary distribution of $\mathcal{M}$ by μ. Formally the mixing time of $\mathcal{M}$ from an initial state $x \in \mathcal{S}$ is defined, as a function of the deviation ε from stationarity, by

$$\mathrm{Mix}_x(\mathcal{M}, \varepsilon) = \min\{t > 0 : \mathrm{d_{TV}}(P^t(x, \cdot), \mu) \leq \varepsilon\}$$

where

$$\mathrm{d_{TV}}(\theta_1, \theta_2) = \frac{1}{2} \sum_i |\theta_1(i) - \theta_2(i)| = \max_{A \subseteq \mathcal{S}} |\theta_1(A) - \theta_2(A)|$$

is the total variation distance between two distributions θ_1 and θ_2 on $\mathcal{S}$. The mixing time $\mathrm{Mix}(\mathcal{M}, \varepsilon)$ of $\mathcal{M}$ is then obtained my maximising over all possible initial states

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) = \max_{x \in \mathcal{S}} \mathrm{Mix}_x(\mathcal{M}, \varepsilon).$$

We say that $\mathcal{M}$ is *rapidly mixing* if the mixing time of $\mathcal{M}$ is polynomial in n and $\log(\varepsilon^{-1})$ and our goal is to establish rapid mixing of Markov chains for sampling from π. We will mainly be concerned with providing good upper bounds on the mixing time of Markov chains and we now go on to describe a classical method for establishing such bounds.

2.3 Coupling and Path Coupling

A classical method for bounding the mixing time of a Markov chain is the coupling method. A coupling of two distributions θ_1 and θ_2 is a joint distribution whose marginal distributions are θ_1 and θ_2. We will discuss the precise requirements in more detail subsequently. Coupling is a general probabilistic technique and it can be applied to the study of the mixing time of Markov chains by considering two copies of the same Markov chain, $\mathcal{M}$. Let the state space of $\mathcal{M}$ be $\mathcal{S}$ and its transition matrix be P. We denote the two copies of $\mathcal{M}$ by $X = X_0, X_1, \ldots$ and $Y = Y_0, Y_1, \ldots$. Viewed individually X and Y both behave exactly as $\mathcal{M}$,

but when viewed as a coupled process their moves may be correlated. The aim of the coupling is to bring copy X and copy Y together as quickly as possible; note that if $X_t = Y_t$ then it is straightforward to arrange that $X_{t'} = Y_{t'}$ for $t' \geq t$.

In order to construct a coupling for $\mathcal{M}$ we need to define a coupling $\Psi(x, y)$ of the distributions $P(x, \cdot)$ and $P(y, \cdot)$ for each pair $(x, y) \in \mathcal{S} \times \mathcal{S}$. In particular in order for the marginal distributions of $\Psi(x, y)$ to be $P(x, \cdot)$ and $P(y, \cdot)$ we require that

$$P(x, x') = \sum_{y' \in \mathcal{S}} \Pr_{(\sigma,\tau)\in\Psi(x,y)}(\sigma = x', \tau = y') \quad \forall x' \in \mathcal{S}$$

and

$$P(y, y') = \sum_{x' \in \mathcal{S}} \Pr_{(\sigma,\tau)\in\Psi(x,y)}(\sigma = x', \tau = y') \quad \forall y' \in \mathcal{S}$$

where we write $(\sigma, \tau) \in \Psi(x, y)$ when the pair of states (σ, τ) is drawn from $\Psi(x, y)$. Since the coupling $\Psi(x, y)$ is defined for all pairs of states $(x, y) \in \mathcal{S} \times \mathcal{S}$ it is the transition matrix of a Markov chain with state space $\mathcal{S} \times \mathcal{S}$. This type of coupling, which is the transition matrix of a Markov chain, is called *Markovian*. The following lemma, known as the coupling lemma, bounds the mixing time of a Markov chain using coupling (see for example Aldous [2]).

Lemma 5 (Coupling Lemma)**.** *Let (X_t, Y_t) be a coupling for a Markov chain $\mathcal{M}$ on $\mathcal{S}$. Suppose that $t(\varepsilon) : (0, 1) \to \mathbb{N}$ satisfies*

$$\Pr(X_{t(\varepsilon)} \neq Y_{t(\varepsilon)}) \leq \varepsilon$$

for all pairs of initial states $X_0 = x, Y_0 = y \in \mathcal{S}$ and $\varepsilon > 0$. Then the mixing time of $\mathcal{M}$ satisfies

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) \leq t(\varepsilon).$$

Proof. Let P be the transition matrix of $\mathcal{M}$ and $P^t(x, \cdot)$ the t-step distribution of $\mathcal{M}$ starting from state $X_0 = x$. For any $\varepsilon \in (0, 1)$ and some corresponding $t = t(\varepsilon)$ we have

$$\begin{aligned} \mathrm{d_{TV}}(P^t(x, \cdot), P^t(y, \cdot)) &= \max_{A \subseteq \mathcal{S}} |\Pr(X_t \in A) - \Pr(Y_t \in A)| \\ &\leq \max_{A \subseteq \mathcal{S}} |\Pr(X_t \in A, Y_t \notin A)| \\ &\leq \Pr(X_t \neq Y_t) \\ &\leq \varepsilon \end{aligned}$$

for any pair of states $x, y \in \mathcal{S}$. Now suppose that Y_0 has distribution μ, then $\mathrm{d_{TV}}(P^t(x, \cdot), \mu) \leq \varepsilon$ for any initial state $X_0 = x \in \mathcal{S}$. □

The following lemma is useful for establishing the mixing time of a Markov chain (see for example Dyer and Greenhill [22]).

Lemma 6. *Let Φ be an integer valued metric defined on $\mathcal{S} \times \mathcal{S}$ which takes values in $\{0, \ldots, D\}$. Let (X_t, Y_t) be a coupling for a Markov chain $\mathcal{M}$ on $\mathcal{S}$. Suppose that there exists a constant $0 < \beta \leq 1$ such that $\mathbf{E}[\Phi(X_{t+1}, Y_{t+1})] \leq \beta\Phi(X_t, Y_t)$ for all pairs $(X_t, Y_t) \in \mathcal{S} \times \mathcal{S}$. If $\beta < 1$ then the mixing time of $\mathcal{M}$ satisfies*

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{1-\beta}.$$

Furthermore if $\beta = 1$ and there exists a constant $\alpha > 0$ such that

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

for all t then the mixing time of $\mathcal{M}$ satisfies

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Proof. The proof is based on Dyer and Greenhill [22]. Using the fact that Φ is non-negative and only takes integer values we have

$$\Pr(X_t \neq Y_t) \leq \mathbf{E}[\Phi(X_t, Y_t)]$$

by Markov's inequality. Furthermore,

$$\mathbf{E}[\Phi(X_t, Y_t)] \leq \beta^t \Phi(X_0, Y_0) \leq \beta^t D$$

which can be verified by induction on t. Hence if $\beta < 1$ then the coupling lemma (Lemma 5) gives

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{\log(\beta^{-1})} \leq \frac{\log(D\varepsilon^{-1})}{1-\beta}$$

since $1 - \beta \leq |\log(\beta)| = \log(\beta^{-1})$ for $0 < \beta < 1$ which can be verified by considering the series expansion of $\log(1-x)$ where $x = 1 - \beta$.

Dyer and Greenhill also give a proof of the $\beta = 1$ case, however as we will not make use of that case in this thesis we omit the proof. □

A difficulty arising in bounding the mixing time of a Markov chain using coupling is that one needs to specify the coupling for all possible pairs of states. Path coupling, introduced by Bubley and Dyer [5] is a method of reducing the

number of states for which the coupling needs to be specified. The key idea of path coupling is to specify a suitable set of *adjacent* pairs of states that connects the state space and then define a coupling for all pairs of adjacent states. The path coupling machinery then extends the coupling to all pairs of states in the state space. In particular, we need to define a relation $S \subseteq \mathcal{S} \times \mathcal{S}$ which connects the state space and which has the property that for all $(X_t, Y_t) \in \mathcal{S} \times \mathcal{S}$ there exists a path

$$X_t = Z_0, Z_1, \ldots, Z_l = Y_t$$

such that $(Z_i, Z_{i+1}) \in S$ for $0 \leq i < l$. Furthermore, for a metric Φ defined on all pairs in $\mathcal{S} \times \mathcal{S}$ we require that

$$\sum_{i=0}^{l-1} \Phi(Z_i, Z_{i+1}) = \Phi(X_t, Y_t).$$

for the given path between X_t and Y_t. A coupling defined on pairs in S can then be extended to a coupling defined for each pair in $\mathcal{S} \times \mathcal{S}$ by inductively coupling and conditioning on the previous choice along the path of configurations in S.

Theorem 7 (Bubley, Dyer [5])**.** *Let $\mathcal{M}$ be a Markov chain with state space $\mathcal{S}$. Let Φ be an integer valued metric defined on $\mathcal{S} \times \mathcal{S}$ which takes values in $\{0, \ldots, D\}$. Let $S \subseteq \mathcal{S} \times \mathcal{S}$ be a relation with transitive closure $\mathcal{S} \times \mathcal{S}$ such that for all $(X_t, Y_t) \in \mathcal{S} \times \mathcal{S}$ there exists a path*

$$X_t = Z_0, Z_1, \ldots, Z_l = Y_t$$

such that $(Z_i, Z_{i+1}) \in S$ for $0 \leq i < l$ and also

$$\sum_{i=0}^{l-1} \Phi(Z_i, Z_{i+1}) = \Phi(X_t, Y_t)$$

Suppose that $(X, Y) \mapsto (X', Y')$ is a coupling of a Markov chain $\mathcal{M}$ defined for all pairs $(X, Y) \in S$. Then this coupling can be extended to a coupling $(X_t, Y_t) \mapsto (X_{t+1}, Y_{t+1})$ defined for all pairs $(X_t, Y_t) \in \mathcal{S} \times \mathcal{S}$ such that if there exists a constant $0 < \beta \leq 1$ such that $\mathbf{E}[\Phi(X', Y')] \leq \beta\Phi(X, Y)$ for all pairs $(X, Y) \in S$ then

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

Proof. This proof is based on the account of path coupling in Dyer and Greenhill [23]. We extend the existing coupling along the given path to all pairs

$(X_t, Y_t) \in \mathcal{S} \times \mathcal{S}$ as follows. We obtain a new path $Z'_0, Z'_1, \ldots, Z'_l$ by first selecting Z'_0 from the distribution $P(Z_0, \cdot)$ where P is the transition matrix of $\mathcal{M}$. We then select Z'_1 according to the distribution induced by the transition $(Z_0, Z_1) \mapsto (Z'_0, Z'_1)$ in the coupled process conditioned on the choice of Z'_0. Continue to select the states from the distribution induced by the given transition in the coupled process, conditioned on the previous choice. Then let $X_{t+1} = Z'_0$ and $Y_{t+1} = Z'_l$.

Then using the triangle inequality for metrics and linearity of expectation we have

$$\begin{aligned}
\mathbf{E}\left[\Phi(X_{t+1}, Y_{t+1}\right] &\leq \mathbf{E}\left[\sum_{i=0}^{l-1} \Phi(Z'_i, Z'_{i+1})\right] \\
&= \sum_{i=0}^{l-1} \mathbf{E}\left[\Phi(Z'_i, Z'_{i+1})\right] \\
&\leq \beta \sum_{i=0}^{l-1} \Phi(Z_i, Z_{i+1}) \\
&= \beta\Phi(X_t, Y_t)
\end{aligned}$$

which completes the proof. □

In order to take maximum advantage of the path coupling method we need to make the set S as small as possible whilst continuing to satisfy the conditions of Theorem 7. This leads to a trade off between the simplicity of the metric and the relation S. It is often the case that one can define an ergodic Markov chain $\mathcal{M}$ on $\mathcal{S}$ with the desired stationary distribution μ but that it is convenient for technical reasons (such as being able to use a simple metric in a path coupling construction) to extend $\mathcal{M}$ to a Markov chain $\mathcal{M}_{\text{ext}}$ with state space $\mathcal{S}^+ \supseteq \mathcal{S}$ when bounding its mixing time. The state space $\mathcal{S}^+$ of the extended chain is required to be finite which is generally straightforward to ensure. The extended chain $\mathcal{M}_{\text{ext}}$ acts just like the original chain $\mathcal{M}$ when the starting state of both chains is in $\mathcal{S}$ and $\mathcal{M}_{\text{ext}}$ will never make a move from a state in $\mathcal{S}$ to a state in $\mathcal{S}^+ \setminus \mathcal{S}$. Hence all states in $\mathcal{S}^+ \setminus \mathcal{S}$ are transient states with zero measure in the stationary distribution μ_{ext} of $\mathcal{M}_{\text{ext}}$. Intuitively, if $\mathcal{M}_{\text{ext}}$ is rapidly mixing then the original chain $\mathcal{M}$ is also rapidly mixing with at most the same mixing time. Using this kind of extended chain is a standard technique, however for completeness we present a proof that the mixing time of the extended chain is an upper bound on the mixing time of the original chain.

Lemma 8. *Let $\mathcal{M}$ be an ergodic Markov chain on the state space $\mathcal{S}$ and let μ be the unique stationary distribution of $\mathcal{M}$. Let P be the transition matrix of $\mathcal{M}$. Then let the Markov chain $\mathcal{M}_{\text{ext}}$ be an extension of $\mathcal{M}$ to the (finite) state space $\mathcal{S}^+$. In particular, the transition matrix P_{ext} of $\mathcal{M}_{\text{ext}}$ is given by $P_{\text{ext}}(x,y) = P(x,y)$ for all pairs of states $(x,y) \in \mathcal{S} \times \mathcal{S}$. Furthermore let*

$$\lim_{t\to\infty} P_{\text{ext}}^t(x,y) = 0 \tag{2.2}$$

for any states $x \in \mathcal{S}^+$ and $y \in \mathcal{S}^+ \setminus \mathcal{S}$. Let μ_{ext} be the probability distribution on $\mathcal{S}^+$ given by

$$\mu_{\text{ext}}(x) = \begin{cases} \mu(x) & \text{if } x \in \mathcal{S} \\ 0 & \text{if } x \in \mathcal{S}^+ \setminus \mathcal{S}. \end{cases}$$

Then μ_{ext} is the unique stationary distribution of $\mathcal{M}_{\text{ext}}$ and furthermore the mixing time of $\mathcal{M}$ satisfies

$$\text{Mix}(\mathcal{M}, \varepsilon) \leq \text{Mix}(\mathcal{M}_{\text{ext}}, \varepsilon).$$

Proof. We begin by showing that μ_{ext} is a stationary distribution of $\mathcal{M}_{\text{ext}}$. For any state $y \in \mathcal{S}^+$

$$\begin{aligned} \sum_{x\in\mathcal{S}^+} \mu_{\text{ext}}(x) P_{\text{ext}}(x,y) &= \sum_{x\in\mathcal{S}} \mu(x) P_{\text{ext}}(x,y) \\ &= \begin{cases} \mu(y) & \text{if } y \in \mathcal{S} \\ 0 & \text{if } y \in \mathcal{S}^+ \setminus \mathcal{S} \end{cases} \end{aligned}$$

since μ is a stationary distribution of $\mathcal{M}$ and $P_{\text{ext}}(x,y) = 0$ whenever $x \in \mathcal{S}$ and $y \in \mathcal{S}^+ \setminus \mathcal{S}$.

Now suppose that μ' is a stationary distribution of $\mathcal{M}_{\text{ext}}$. First for any $y \in \mathcal{S}^+ \setminus \mathcal{S}$ we have

$$\mu'(y) = \lim_{t\to\infty} \sum_{x\in\mathcal{S}^+} \mu'(x) P_{\text{ext}}^t(x,y) = \sum_{x\in\mathcal{S}^+} \mu'(x) \lim_{t\to\infty} P_{\text{ext}}^t(x,y) = 0 \tag{2.3}$$

since $\mathcal{S}^+$ is finite and using the limit from (2.2). Now suppose that $y \in \mathcal{S}$. Then using (2.3)

$$\mu'(y) = \sum_{x\in\mathcal{S}} \mu'(x) P_{\text{ext}}(x,y) + \sum_{x\in\mathcal{S}^+\setminus\mathcal{S}} \mu'(y) P_{\text{ext}}(x,y) = \sum_{x\in\mathcal{S}} \mu'(x) P(x,y)$$

and hence $\mu'(y) = \mu(y) = \mu_{\text{ext}}(y)$ for each $y \in \mathcal{S}$ since μ is the unique stationary distribution of $\mathcal{M}$. Hence, μ_{ext} is the unique stationary distribution of $\mathcal{M}_{\text{ext}}$.

Thus if the initial state of $\mathcal{M}_{\text{ext}}$ is in $\mathcal{S}$ then the chain behaves exactly as $\mathcal{M}$ and thus converges to μ_{ext}. Otherwise the initial state of the chain is in $\mathcal{S}^+ \setminus \mathcal{S}$ and it eventually makes a transition to a state in $\mathcal{S}$ after which it will converge to μ_{ext} as discussed above.

In order to relate the mixing times of the two chains we need to establish the following fact

$$P^t_{\text{ext}}(x, y) = \begin{cases} P^t(x, y) & \text{if } y \in \mathcal{S} \\ 0 & \text{if } y \in \mathcal{S}^+ \setminus \mathcal{S} \end{cases} \tag{2.4}$$

for every $x \in \mathcal{S}$. We establish (2.4) by strong induction on t. The base case is $t = 1$. When $t = 1$ then the $y \in \mathcal{S}$ case follows directly from the definition of P_{ext} and the case when $y \in \mathcal{S}^+ \setminus \mathcal{S}$ follows since $\sum_{x' \in \mathcal{S}} P_{\text{ext}}(x, x') = \sum_{x' \in \mathcal{S}} P(x, x') = 1$ and thus $P_{\text{ext}}(x, y) = 0$ for any $y \notin \mathcal{S}$.

Now suppose that (2.4) holds for $t - 1$. Then

$$\begin{aligned} P^t_{\text{ext}}(x, y) &= \sum_{x' \in \mathcal{S}^+} P^{t-1}_{\text{ext}}(x, x') P_{\text{ext}}(x', y) \\ &= \sum_{x' \in \mathcal{S}} P^{t-1}_{\text{ext}}(x, x') P_{\text{ext}}(x', y) + \sum_{x' \in \mathcal{S}^+ \setminus \mathcal{S}} P^{t-1}_{\text{ext}}(x, x') P_{\text{ext}}(x', y) \\ &= \begin{cases} \sum_{x' \in \mathcal{S}} P^{t-1}(x, x') P(x', y) & \text{if } y \in \mathcal{S} \\ 0 & \text{if } y \in \mathcal{S}^+ \setminus \mathcal{S} \end{cases} \end{aligned}$$

where the last equality uses the induction hypothesis. Note in particular that $P_{\text{ext}}(x', y) = 0$ when $x' \in \mathcal{S}$ and $y \in \mathcal{S}^+ \setminus \mathcal{S}$ and also that $P^{t-1}_{\text{ext}}(x, x') = 0$ whenever $x' \in \mathcal{S}^+ \setminus \mathcal{S}$.

Hence using (2.4) we have

$$\begin{aligned}
\mathrm{Mix}(\mathcal{M},\varepsilon) &= \max_{x\in\mathcal{S}} \min\{t>0 : \mathrm{d}_{\mathrm{TV}}(P^t(x,\cdot),\mu)\le\varepsilon\} \\
&= \max_{x\in\mathcal{S}} \min\left\{t>0 : \frac{1}{2}\sum_{x'\in\mathcal{S}}\left|P^t(x,x')-\mu(x')\right|\le\varepsilon\right\} \\
&= \max_{x\in\mathcal{S}} \min\left\{t>0 : \frac{1}{2}\sum_{x'\in\mathcal{S}}\left|P^t_{\mathrm{ext}}(x,x')-\mu_{\mathrm{ext}}(x')\right|\le\varepsilon\right\} \\
&\le \max_{x\in\mathcal{S}^+} \min\left\{t>0 : \frac{1}{2}\sum_{x'\in\mathcal{S}^+}\left|P^t_{\mathrm{ext}}(x,x')-\mu_{\mathrm{ext}}(x')\right|\le\varepsilon\right\} \\
&= \mathrm{Mix}(\mathcal{M}_{\mathrm{ext}},\varepsilon)
\end{aligned}$$

where the inequality uses the fact that $\mathcal{S}\subseteq\mathcal{S}^+$. □

Remark. Although the requirements of (2.2) may seem limiting, this condition is generally straightforward to arrange in practice. In particular, (2.2) holds whenever all the states in $\mathcal{S}^+\setminus\mathcal{S}$ are transient (see Corollary 6.2.5 in Grimmett and Stirzaker [35]).

When working with Markov chains whose state space is the set of legal configurations, Ω, of a spin system it is often desirable to use Hamming distance as the metric and let S be the set of configurations that only differ on the spin assigned to a single site. The Hamming distance between two configurations x and y, denoted by $\mathrm{Ham}(x,y)$, is the number of sites that are assigned different spins in x and y. However, for some spin systems it is the case that this choice of metric and definition of S fails to satisfy the conditions of Theorem 7. This is only a minor technical difficulty which is easily solved by extending the state space of the Markov chain in question to Ω^+ as discussed above.

From now on and throughout this thesis we let $S=\bigcup_{j\in V}S_j$ where $S_j\subseteq\Omega^+\times\Omega^+$ is the set of pairs of configurations that differ only on the spin assigned to site j. Hence $S=\{(x,y)\in\Omega^+\times\Omega^+ : \mathrm{Ham}(x,y)=1\}$ is the set of all pairs of configurations that only differ on the spin assigned to a single site. For ease of reference we state the following corollary of Theorem 7 and Lemmas 6 and 8.

Corollary 9. *Let $\mathcal{M}$ be a Markov chain with state space Ω. Suppose that $(x,y)\mapsto(x',y')$ is a coupling of $\mathcal{M}$ defined for all pairs $(x,y)\in S$ and that $\mathbf{E}\left[\mathrm{Ham}(x',y')\right]=(1-\gamma)\mathrm{Ham}(x,y)$ for some $0<\gamma<1$. Then $\mathrm{Mix}(\mathcal{M},\varepsilon)\le\log(n\varepsilon^{-1})/\gamma$.*

2.4 Block Dynamics and Influence Parameters

It is sometimes convenient to consider a Markov chain that updates a set of sites simultaneously during each step rather than just one site. One reason for this is that single-site update Markov chains may in some cases not yield to analysis while a block dynamics may. We will give examples of this phenomena in due course. Furthermore, the analysis of block dynamics is relevant to the study of single-site update Markov chains since it is known that their mixing times are similar, provided that the blocks used are of constant size. In particular, it is possible to obtain a bound on the mixing time of a single-site chain from an existing bound on the mixing time of a block dynamics chain by using some Markov chain comparison techniques, although at the expense of a polynomial factor in the mixing time. For details of the comparison method used to relate the mixing times of these chains consult the survey paper by Dyer, Goldberg, Jerrum and Martin [21]. We now formalise our notion for block dynamics and give some definitions required to specify our conditions for rapid mixing of systematic scan Markov chains that use block dynamics. We will make frequent use of these definitions throughout the thesis. The notation for block dynamics is partly based on notation in Weitz [55] and we also draw from definitions in Dyer et al. [18] in order to define our influence parameters.

We consider a finite collection of m blocks $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ such that each block $\Theta_k \subseteq V$ and Θ covers V. We say that Θ covers V if $\bigcup_{k=1}^{m} \Theta_k = V$. One site may be contained in several blocks and the size of each block is not required to be the same; we do however require that the size of each block is bounded independently of n. This requirement is in order to ensure that a step of the chain can be efficiently implemented. For any block Θ_k and a pair of configurations $x, y \in \Omega^+$ we write "$x = y$ on Θ_k" if $x_i = y_i$ for each $i \in \Theta_k$ and similarly "$x = y$ off Θ_k" if $x_i = y_i$ for each $i \in V \setminus \Theta_k$. We will sometimes saw that x and y "agree" off Θ_k if $x = y$ off Θ_k. We also let $\partial\Theta_k = \{i \in V \setminus \Theta_k \mid \exists j \in \Theta_k : (i, j) \in E\}$ denote the set of sites adjacent to but not included in Θ_k; we will refer to $\partial\Theta_k$ as the *boundary* of Θ_k.

With each block Θ_k, we associate a transition matrix $P^{[k]}$ on state space Ω^+. For ease of reference we say that $P^{[k]}$ is a *valid update rule* if it satisfies the following two properties:

1. If $P^{[k]}(x, y) > 0$ then $x = y$ off Θ_k, and also

2. π is invariant with respect to $P^{[k]}$.

We will always make sure to satisfy these two properties by construction of the update rules. Property 1 ensures that an application of $P^{[k]}$ moves the state of the system from from one configuration to another by only updating the sites contained in the block Θ_k and Property 2 ensures that any dynamics composed solely of transitions defined by $P^{[k]}$ converges to π. While the requirements of Property 1 are clear we take a moment to discuss what we mean in Property 2. Consider the following two step process in which some configuration x is initially drawn from π and then a configuration y is drawn from $P^{[k]}(x, \cdot)$ where $P^{[k]}(x, \cdot)$ is the distribution on configurations resulting from applying $P^{[k]}$ to a configuration x. We say that π is invariant with respect to $P^{[k]}$ (i.e. y has distribution π) if for each configuration $\sigma \in \Omega^+$ we have $\Pr(x = \sigma) = \Pr(y = \sigma)$. That is the distribution on configurations generated by the two-step process is the same as if only the first step was executed. In terms of our dynamics this means that once the distribution of the dynamics reaches π, π will continue be the distribution of the dynamics even after applying $P^{[k]}$ to the state of the dynamics. Our main result (Theorem 14) holds for any choice of update rule $P^{[k]}$ provided that it satisfies these two properties.

The distribution $P^{[k]}(x, \cdot)$, which specifies how the dynamics updates block Θ_k, clearly depends on the specific update rule implemented as $P^{[k]}$. In order to make this idea more clear we give some concrete examples of possible update rules.

Example 10. One of the most natural choices for $P^{[k]}$ is the *heat-bath* update rule. Consider the spin system corresponding to proper q-colourings of a graph, and recall that π is the uniform distribution on the set of all proper colourings. The transition matrix $P^{[k]}$ for a heat-bath move makes the following transition from a given configuration x. Let $\Omega_{\Theta_k}(x) \subseteq \Omega^+$ be the set of configurations that agree with x off Θ_k and where no edge containing a site in Θ_k is monochromatic. An edge is said to be monochromatic if each endpoint is assigned the same colour. If $\Omega_{\Theta_k}(x)$ is not empty then $P^{[k]}$ makes a transition to a uniformly chosen configuration in $\Omega_{\Theta_k}(x)$. Otherwise $P^{[k]}$ leaves the configuration unchanged. The two required properties of $P^{[k]}$ hold for heat-bath updates since (1) only the assignment of the spin to the sites in Θ_k are changed and (2) the new configuration is drawn from an appropriate distribution induced by π. Hence an update rule that performs heat-bath updates is a valid update rule.

Example 11. Another well-known choice for $P^{[k]}$ is the *Metropolis* update rule. Again consider the spin system corresponding to proper q-colourings of a graph.

In this case $P^{[k]}$ makes the following transition from a given configuration x. A configuration x' is chosen uniformly at random from the set of all configurations that agree with x off Θ_k. If no edge containing a site in Θ_k is monochromatic in the configuration x', then the new configuration is x'. Otherwise the new configuration is x.

Recall that throughout this thesis we distinguish between two types of Markov chains namely random update Markov chains and systematic scan Markov chains. We now give a definition for each type of Markov chain in the block setting.

Definition 12. *Given a set of blocks $\Theta = \{\Theta_1, \ldots, \Theta_m)$ with associated valid update rules $P^{[1]}, \ldots, P^{[m]}$, a systematic scan Markov chain is a Markov chain $\mathcal{M}_{\rightarrow}$ with state space Ω^+ and transition matrix $P_{\rightarrow} = \prod_{k=1}^{m} P^{[k]}$.*

Definition 13. *Given a set of blocks $\Theta = \{\Theta_1, \ldots, \Theta_m)$ with associated valid update rules $P^{[1]}, \ldots, P^{[m]}$, a random update Markov chain is a Markov chain $\mathcal{M}_{\mathrm{RU}}$ with state space Ω^+ and transition matrix $P_{\mathrm{RU}} = (1/m) \sum_{k=1}^{m} P^{[k]}$.*

Observe that π is a stationary distribution of both $\mathcal{M}_{\rightarrow}$ and $\mathcal{M}_{\mathrm{RU}}$ as discussed above and if the chains are ergodic then π is unique. It is also worth pointing out that the definition of $\mathcal{M}_{\rightarrow}$ holds for *any* order on the set of blocks. We will refer to one application of $P_{\rightarrow}$ (that is updating each block once) as one *scan* of $\mathcal{M}_{\rightarrow}$. One scan takes $\sum_k |\Theta_k|$ *updates* and it is generally straightforward to ensure, via the construction of the set of blocks, that this sum is of order $O(n)$.

It is well-known that the mixing time of a Markov chain can be bounded by studying the influence that the sites have on each other. This technique arises in both path coupling and *Dobrushin's uniqueness criterion*. Recently Weitz [55] generalised two conditions namely "the influence *on* a site is small" (originally attributed to Dobrushin [12]) and "the influence *of* a site is small" (originally Dobrushin and Shlosman [13]) and showed that both imply mixing of a corresponding random update Markov chain. We call a condition of the form "if the influence *on* a site is small then the corresponding dynamics converges to π quickly" a Dobrushin condition since Dobrushin was originally concerned with establishing conditions that hold when the influence on as site is small. In the context of single-site systematic scan, Dyer et al. [18] have pointed out that the condition "the influence *on* a site is small" implies rapid mixing. Our condition is a generalisation of this condition to block dynamics.

We will now formalise the notion of *influence* sites have on each other. Recall that for each site $j \in V$ we let S_j denote the set of pairs $(x, y) \in \Omega^+ \times \Omega^+$ of

configurations that only differ on the spin assigned to site j, that is $x_i = y_i$ for all $i \neq j$. For any pair of configurations $(x, y) \in \Omega^+ \times \Omega^+$ let $\Psi_k(x, y)$ be a coupling of the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ which we will refer to as "updating block Θ_k". Recall that a coupling $\Psi_k(x, y)$ of $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ is a joint distribution on $\Omega^+ \times \Omega^+$ whose marginal distributions are $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ and that we write $(x', y') \in \Psi_k(x, y)$ when the pair of configurations (x', y') is drawn from $\Psi_k(x, y)$. We define the influence of site i on site j under block Θ_k as

$$\rho^k_{i,j} = \max_{(x,y)\in S_i} \{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j)\}. \tag{2.5}$$

The influence of i on j under Θ_k is hence the maximum probability that two coupled Markov chains differ on the spin assigned to site j following an update of block Θ_k starting from two configurations that only differ on the spin assigned to site i. Using this definition of the influence of i on j it is natural to say that the total influence on site j when updating block Θ_k is $\sum_i \rho^k_{i,j}$. To make the condition more general we assign a positive weight w_i to each site $i \in V$. The maximum (weighted) influence *on* a site, the influence parameter we will denote by α, is then

$$\alpha = \max_k \max_{j\in\Theta_k} \sum_{i\in V} \rho^k_{i,j} \frac{w_i}{w_j}. \tag{2.6}$$

We point out that the weights are purely a proof construct and can be omitted using uniform weights. We also observe at this point that our definition of $\rho^k_{i,j}$ is not the standard definition of ρ used in the literature (see for example Simon [51] or Dyer et al. [18]) since the coupling $\Psi_k(x, y)$ is explicitly included. In the block setting it is, however, necessary to include the coupling directly in the definition of ρ as we will discuss in Chapter 3.

2.5 Statement of Results

We now go on to formally state the results we will present in this thesis as well as to discuss their relation to previous work in the field.

2.5.1 A Dobrushin Condition for Rapid Mixing of Systematic Scan with Block Dynamics

Chapter 3 will be concerned with the development of a new method of proving rapid mixing of systematic scan Markov chains using block dynamics. Our main theorem is concerned with using the influence parameter α (defined in (2.6)) to

bound the mixing time of systematic scan. Informally, we will show that if the weighted influence on any site of the underlying graph is sufficiently small then systematic scan mixes rapidly regardless of the scan order. In particular, the systematic scan Markov chain $\mathcal{M}_{\rightarrow}$ mixes in $O(\log n)$ scans of the graph.

Theorem 14. *Consider any spin system with underlying graph $G = (V, E)$. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be any set of blocks covering V. For each block Θ_k let $P^{[k]}$ be a valid update rule associated with block Θ_k. $\mathcal{M}_{\rightarrow}$ is the systematic scan Markov chain which updates the blocks in the order $\Theta_1, \ldots, \Theta_m$. If $\alpha = \max_k \max_{j \in \Theta_k} \sum_{i \in V} \rho_{i,j}^k w_i / w_j < 1$ then $\mathcal{M}_{\rightarrow}$ is ergodic and its mixing time is at most*

$$\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \leq \frac{\log(n\gamma\varepsilon^{-1})}{1-\alpha}$$

scans of the graph where

$$\gamma = \frac{\max_{i \in V} w_i}{\min_{j \in V} w_j}$$

is the maximum ratio between the weights.

Remark. The fact that Theorem 14 holds regardless of the order of the blocks follows from the observation that the value of the parameter α is a maximum and hence does not depend on the order in which the blocks are updated.

This result is a generalisation of a similar condition for single-site dynamics by Dyer et al. [18] as we will discuss in more detail in Chapter 3. Even though we will mainly be concerned with applying Theorem 14 to spin systems corresponding to proper colourings of graphs we point out that it applies to any spin system.

Chapter 3 also contains two applications of Theorem 14 to spin systems corresponding to proper q-colourings of graphs, both of which improve the parameters for which systematic scan mixes. In these applications we restrict the state space of the systematic scan Markov chains to the set of proper colourings, Ω, of the underlying graph. First we allow the underlying graph to be any finite graph with maximum vertex-degree Δ. Previously, the least number of colours for which systematic scan was known to mix in $O(\log n)$ scans was $q > 2\Delta$ and when $q = 2\Delta$ the best known bound on the mixing time was $O(n^2 \log n)$ scans, both due to Dyer et al. [18]. For completeness we pause to mention that the minimum number of colours required for rapid mixing (in $O(n \log n)$ updates) of a random update Markov chain is $q > (11/6)\Delta$ due to Vigoda [53]. We consider the following Markov chain, *edge scan* denoted $\mathcal{M}_{\mathrm{edge}}$, updating each endpoint of an edge during each update. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be a set of m edges in G such that Θ covers V. In order for the scan to be as efficient as possible it is advantageous

to make m as small as possible and it can always be ensured that $m = O(n)$. Note that it is $P^{[k]}$ is the transition matrix for performing a heat-bath move on the endpoints of the edge Θ_k and it was shown in Example 10 that this choice for $P^{[k]}$ is a valid update rule.

Definition 15. *Let $\mathcal{M}_{\text{edge}}$ be the systematic scan Markov chain with state space Ω and transition matrix $\prod_{k=1}^{m} P^{[k]}$.*

We prove the following theorem, which improves the mixing time of systematic scan by a factor of n^2 for proper colourings of general graphs when $q = 2\Delta$ and matches the existing bound when $q > 2\Delta$.

Theorem 16. *Let G be a graph with maximum vertex-degree Δ. Consider the systematic scan Markov chain $\mathcal{M}_{\text{edge}}$ on Ω. If $q \geq 2\Delta$ then the mixing time of $\mathcal{M}_{\text{edge}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{edge}}, \varepsilon) \leq \Delta^2 \log(n\varepsilon^{-1})$$

scans. If $m = O(n)$ then this corresponds to $O(n \log n)$ block updates.

Next we restrict the class of graphs to trees. It is known that a single-site systematic scan mixes in $O(\log n)$ scans when $q > \Delta + 2\sqrt{\Delta - 1}$ and in $O(n^2 \log n)$ scans when $q = \Delta + 2\sqrt{\Delta - 1}$ is an integer; see e.g. Hayes [36] or Dyer et al. [19]. We present a proof of the first of these claims using our condition, although in our case the mixing time will be $O(H)$ where H is the height of the tree (the maximum number of edges between the root and a leaf). We point out that our proof preceded both of the cited results. We define the systematic scan Markov chain *tree scan*, denoted $\mathcal{M}_{\text{tree}}$, as follows. For each site $k \in V$ we let $\Theta_k = \{k\}$, so this is a single-site Markov chain. $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k so $P^{[k]}$ is a valid update rule.

Definition 17. *Let $\mathcal{M}_{\text{tree}}$ be the systematic scan Markov chain with state space Ω and transition matrix $\prod_{k=1}^{n} P^{[k]}$.*

We prove the following theorem.

Theorem 18. *Let G be a tree with maximum vertex degree $\Delta \geq 3$ and height H. Consider the systematic scan Markov chain $\mathcal{M}_{\text{tree}}$ on Ω. If $q \geq \Delta + 2\sqrt{\Delta - 1 + \delta}$ for $\delta > 0$ then the mixing time of $\mathcal{M}_{\text{tree}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{tree}}, \varepsilon) \leq \max\left(\frac{2(\Delta - 1 + \delta)}{\delta}, 4\right)\left(H \log\left(\frac{q - \Delta}{2(\Delta - 1)}\right) + \log(n\varepsilon^{-1})\right)$$

scans of the tree. Since $\log n \leq H \leq n$, this corresponds to $O(nH)$ updates.

Table 2.1. Optimising the number of colours using blocks

Δ	h	$f(\Delta)$	$\lceil \Delta + 2\sqrt{\Delta - 1} \rceil$
3	15	5	6
4	3	7	8
5	12	8	9
6	3	10	11
7	7	11	12
8	13	12	14
9	85	13	15
10	5	15	16
20	21	27	29
30	117	38	41
40	50	49	53
50	150	60	64
60	51	71	76
100	45	115	120

For completeness we mention that the mixing time of a random update Markov chain for proper colourings on a tree mixes in $O(n \log n)$ updates when $q \geq \Delta+2$, a result due to Martinelli et al. [46], improving a related result by Kenyon et al. [43].

We will use a systematic scan with block updates to reduce the number of colours required for mixing of systematic scan for proper colourings of trees. We construct a set of m blocks, where the height h of each block is defined in Table 2.1. Let a block Θ_k contain a site r along with all sites below r in the tree that are at most $h-1$ edges away from r. The set of blocks Θ must cover the sites of the tree and no block has height less than h. Note that $m = O(n)$. As before $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k which is a valid update rule.

Definition 19. *Let $\mathcal{M}_{\text{BlockTree}}$ be the systematic scan Markov chain with state space Ω and transition matrix $\prod_{k=1}^{m} P^{[k]}$ where m is the number of blocks.*

We prove the following theorem which improves the number of colours required for rapid mixing of systematic scan for the stated values of Δ.

Theorem 20. *Let G be a tree with maximum vertex-degree Δ and height H. Consider the systematic scan Markov chain $\mathcal{M}_{\text{BlockTree}}$ on Ω. If $q \geq f(\Delta)$ where $f(\Delta)$ is specified in Table 2.1 for small Δ then the mixing time of $\mathcal{M}_{\text{BlockTree}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{BlockTree}}, \varepsilon) = O(H + \log(\varepsilon^{-1}))$$

scans of the tree. This corresponds to $O(nH)$ block updates by the construction

of the set of blocks.

2.5.2 Sampling H-colourings of the Path

In Chapter 4 we broaden the type of spin system we consider to general H-colourings, although at the expense of limiting the underlying graph of the spin system to the path. When discussing H-colourings it is again natural to refer to elements of C as colours rather than spins. An H-colouring of a graph G is a *homomorphism* from the graph of interest G to some fixed graph H. The vertices of H correspond to colours and the edges of H specify which colours are allowed to be adjacent in an H-colouring of G. If $H = (C, E_H)$ is any fixed graph then an H-colouring of a graph $G = (V, E)$ is a function $h : V \to C$ such that $(h(v), h(u)) \in E_H$ for all edges $(v, u) \in E$ of G. We will only consider the case when G is the n-vertex path.

We study Markov chains that perform heat-bath moves on a constant number of sites at the time. Like in our other applications we would normally let Ω (the set of all H-colourings of G) be the state space of our Markov chains, however, if H is bipartite then we encounter a minor technical difficulty because the Markov chain may not be ergodic. We overcome this ergodicity issue by partitioning the state space as follows. If C_1 and C_2 are the colour classes of H then $\Omega_1 = \{x \in \Omega : x_1 \in C_1\}$ is the set of H-colourings of the n-vertex path where the first site of the path is assigned a colour from C_1. We let V_1 denote the set of odd-numbered sites of the path and V_2 the set of even-numbered sites. Observe that for each H-colouring in Ω_1 it holds that each site in V_1 is assigned a colour from C_1 and each site in V_2 is assigned a colour from C_2. Similarly $\Omega_2 = \{x \in \Omega : x_1 \in C_2\}$ is the set of H-colourings where the first site is assigned a colour from C_2. Intuitively, Ω_1 and Ω_2 are the two connected components of Ω and we will show (Lemma 63) that the constructed Markov chains are ergodic on both Ω_1 and Ω_2. To see that $\Omega_1 \cup \Omega_2$ contain all H-colourings of the n-vertex path it is enough to observe that if $x \in \Omega$ then any pair of adjacent sites of the n-vertex path must be assigned colours from opposite colour classes of H in x. We let $\Omega_\sim$ be the relevant state space of the Markov chains in order to ensure ergodicity. In particular, if H is non-bipartite then $\Omega_\sim = \Omega$. Otherwise H is bipartite and we let $\Omega_\sim$ be one of Ω_1 and Ω_2. This is the same partition used by Dyer et al. in [20]. See also Cooper et al. [8] for a discussion of a similar issue.

We are now ready to define our systematic scan Markov chains for sampling H-colourings of the n-vertex path and state our results. Let $l_1 = \lceil \Delta_H^2 \log(\Delta_H^2 + 1) \rceil +$

1 where Δ_H is the maximum vertex-degree of H. Then let $\Theta = \{\Theta_1, \ldots, \Theta_{m_1}\}$ be any set of $m_1 = \lceil n/l_1 \rceil$ blocks such that each block consists of exactly l_1 consecutive sites and Θ covers V. For each block Θ_k we define $P^{[k]}$ to be the transition matrix on the state space $\Omega_\sim$ for performing a heat-bath move on Θ_k. As before observe that $P^{[k]}$ is a valid update rule as shown in Example 10.

Definition 21. *Let $\mathcal{M}_{\text{AnyOrder}}$ be the systematic scan Markov chain with state space $\Omega_\sim$ and transition matrix $\prod_{k=1}^{m_1} P^{[k]}$.*

It is worth pointing out that the following result holds for *any order* of the blocks, as is the case for all results obtained by Dobrushin uniqueness.

Theorem 22. *Let H be a fixed connected graph with maximum vertex-degree Δ_H and consider the systematic scan Markov chain $\mathcal{M}_{\text{AnyOrder}}$ on the state space $\Omega_\sim$. Suppose that H is a graph in which every two sites are connected by a 2-edge path. Then the mixing time of $\mathcal{M}_{\text{AnyOrder}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{AnyOrder}}, \varepsilon) \leq \Delta_H^2(\Delta_H^2 + 1)\log(n\varepsilon^{-1})$$

scans of the n-vertex path. This corresponds to $O(n \log n)$ block updates by the construction of the set of blocks.

Remark. Note that each H for which Theorem 22 is valid is non-bipartite so $\Omega_\sim = \Omega$.

Remark. Several well known H-colouring problems satisfy the condition of Theorem 22, for example Widom-Rowlinson configurations, independent set configurations and proper q-colourings for $q \geq 3$. The fact that an H corresponding to 3-colourings satisfies the condition of the theorem is particularly interesting since a lower bound of $\Omega(n^2 \log n)$ scans for single-site systematic scan on the path is proved in Dyer at al. [20]. This means that using a simple single-site coupling cannot be sufficient to establishing Theorem 22 for any family of H including 3-colourings and hence we have to use block updates.

While many natural H-colouring problems belong to the family covered by Theorem 22, others (e.g. Beach configurations) are not included. We go on to show that systematic scan mixes in $O(\log n)$ scans for any fixed graph H by placing more strict restrictions on the construction of the blocks and the order of the scan. Let $s = 4q + 1$, $\beta = \lceil \log(2sq^s + 1) \rceil q^s$ and $l_2 = 2\beta s$. For any integer n consider the following set of $m_2 + 1 = \lfloor 2n/l_2 \rfloor$ blocks $\Theta = \{\Theta_0, \ldots, \Theta_{m_2}\}$ where

$$\Theta_k = \{k\beta s + 1, \ldots, \min((k+2)\beta s, n)\}.$$

We observe that Θ covers V by construction of the set of blocks. Furthermore note that the size of Θ_{m_2} is at least βs and that the size of every other block is exactly l_2.

Definition 23. *Let $\mathcal{M}_{\text{FixedOrder}}$ be the systematic scan Markov chain, with state space $\Omega_\sim$, which performs a heat-bath move on each block in the order $\Theta_0, \ldots, \Theta_{m_2}$.*

We will use path coupling [5] to prove the following theorem, which improves the mixing time from the corresponding result in Dyer et al. [20] from $O(n^5)$ scans to $O(\log n)$ scans.

Theorem 24. *Let H be any fixed connected graph and consider the systematic scan Markov chain $\mathcal{M}_{\text{FixedOrder}}$ on the state space $\Omega_\sim$. The mixing time of $\mathcal{M}_{\text{FixedOrder}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{FixedOrder}}, \varepsilon) \leq (4sq^s + 2)\log(n\varepsilon^{-1})$$

scans of the n-vertex path. This corresponds to $O(n \log n)$ block updates by the construction of the set of blocks.

Remark. It is worth remarking at this point that Theorem 24 eclipses Theorem 22 in the sense that it shows the existence of a systematic scan for a broader family of H than Theorem 22 but with the same (asymptotic) mixing time. The result stated as Theorem 22 however remains interesting in its own right since it applies to any order of the scan. Following the proof of Theorem 22 we will discuss (Observation 60) the obstacles one encounters when attempting to extend Theorem 22 to a larger family of H using the same method of proof.

For completeness we conclude Chapter 4 by considering a random update Markov chain for sampling H-colourings of the n-vertex path. Let $\gamma = 2q^s + 1$, where $s = 4q + 1$ as before, and define the following set of $n + s\gamma - 1$ blocks, which is constructed such that each site is contained in exactly $s\gamma$ blocks

$$\Theta_k = \begin{cases} \{k, \ldots, \min(k + s\gamma - 1, n)\} & \text{when } k \in \{1, \ldots, n\} \\ \{1, \ldots, n + s\gamma - k\} & \text{when } k \in \{n+1, \ldots, n + s\gamma - 1\}. \end{cases}$$

Definition 25. *Let $\mathcal{M}_{\text{RND}}$ be the random update Markov chain, with state space $\Omega_\sim$, which at each step selects a block uniformly at random and performs a heat-bath move on it.*

We will use path coupling [5] to prove the following theorem, which improves the mixing time from the corresponding result in Dyer et al. [20] from $O(n^5)$ updates to $O(n \log n)$ updates.

Theorem 26. *Let H be any fixed connected graph and consider the random update Markov chain $\mathcal{M}_{\text{RND}}$ on the state space $\Omega_\sim$. The mixing time of $\mathcal{M}_{\text{RND}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{RND}}, \varepsilon) \leq \frac{(n + 2sq^s + s - 1)\log(n\varepsilon^{-1})}{s}$$

block updates.

2.5.3 Sampling 7-colourings of the Grid

In Chapter 5 we present a systematic scan Markov chain for sampling from the uniform distribution of proper 7-colourings of the square grid. We let the underlying graph $G = (V, E)$ be be a finite piece of the infinite square grid. In this section Ω is the set of all proper 7-colourings of G. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be a set of m blocks such that each block $\Theta_k \subseteq V$ is a 2×2 sub-grid and Θ covers V. As before it is advantageous to make m as small as possible in order for the scan to be efficient. For each block Θ_k we let $P^{[k]}$ be the transition matrix for performing a heat-bath move on Θ_k. Hence $P^{[k]}$ is a valid update rule.

Definition 27. *Let $\mathcal{M}_{\text{grid}}$ be the systematic scan Markov chain with state space Ω and transition matrix $P_{\text{grid}} = \prod_{k=1}^{m} P^{[k]}$.*

We will prove the following theorem and point out that this is the first proof of rapid mixing of systematic scan for 7-colourings on the grid as it improves the 8-colouring result which is included in Theorem 16. The proof of this theorem is computer-assisted.

Theorem 28. *Let G be a finite and rectangular piece of the infinite square lattice. Consider the systematic scan Markov chain $\mathcal{M}_{\text{grid}}$ on Ω. The mixing time of $\mathcal{M}_{\text{grid}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{grid}}, \varepsilon) \leq 63 \log(n\varepsilon^{-1})$$

scans of the grid. This corresponds to $O(n \log n)$ block updates since each block is of size 4.

As before we wish to compare the systematic scan results to known results for random update Markov chains. In the random update case, Achlioptas et al. [1] gave a computer-assisted proof of mixing in $O(n \log n)$ updates when $q = 6$ by

considering blocks consisting of 2×3 sub-grids. More recently Goldberg et al. [33] gave a hand-proof of mixing in $O(n \log n)$ updates when $q \geq 7$ by establishing strong spatial mixing which in turn implies the stated bound on the mixing time. Previously Salas and Sokal [50] gave a computer-assisted proof of the $q = 7$ case, a result which was also implied by another computer-assisted result due to Bubley, Dyer and Greenhill [6] that applies to 4-regular triangle-free graphs. Finally it is worth pointing out that, in the special case when $q = 3$, two complementary results of Luby, Randall and Sinclair [44] and Goldberg, Martin and Paterson [34] give rapid mixing of a random update chain.

2.5.4 Single-site Systematic Scan for Bipartite Graphs

In Chapter 6 we study a single-site systematic scan Markov chain for sampling from the uniform distribution of proper q-colourings of bipartite graphs. We let $G = (V, E)$ be any bipartite graph with maximum vertex-degree Δ. The colour classes of G are denoted by $L(V)$ and $R(V)$. We let Ω be the set of proper q-colourings of G. We study a Markov chain $\mathcal{M}_{\text{LR}}$, called *left-right scan*, that first updates each site in $L(V)$ using a Metropolis move (see Example 11) and then updates each site in $R(V)$ also using Metropolis.

Definition 29. *Let $\mathcal{M}_{\text{LR}}$ be the systematic scan Markov chain which state space Ω which makes the following transitions:*

1. *for each $i \in L(V)$ make a Metropolis move on site i*
2. *for each $i \in R(V)$ make a Metropolis move on site i.*

We assign weights to each site such that $w_i = \omega_l = q^3 - 4$ for each site $i \in L(V)$ and $w_i = \omega_r = 2\omega_l - 4$ for each site $i \in R(V)$. For technical reasons we only consider the case when $\Delta \geq 3$, but note that tight bounds are given in Dyer et al. [20] for the $\Delta = 2$ case.

Theorem 30. *Let G be any bipartite graph with maximum vertex-degree $\Delta \geq 3$. Consider the systematic scan Markov chain $\mathcal{M}_{\text{LR}}$ on the state space Ω. Let $\gamma = \omega_r \left(1 + \frac{1}{q^3}\right) - \frac{\Delta\omega_l}{q} - \frac{\Delta\omega_r}{q} - \frac{\Delta^2\omega_r}{q^2}$ where $\omega_l = q^3 - 4$ and $\omega_r = 2\omega_l - 4$. If $q \geq 2\Delta$ then $\gamma > 0$ and the mixing time of $\mathcal{M}_{\text{LR}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{LR}}, \varepsilon) \leq \frac{\omega_r \log(n\omega_r\varepsilon^{-1})}{\gamma}$$

scans.

We have previously pointed out that Theorem 30 has since been improved by a result of Bordewich et al. [4] since been improved by a result of Bordewich et al. [4] since been improved by a result of Bordewich et al. [4] when $\Delta \geq 9$ and matched when $5 \leq \Delta < 9$. Theorem 30 remains, however, the only single-site systematic scan that mixes in $O(\log n)$ scans when $q = 2\Delta$ and $\Delta = 3$ or $\Delta = 4$. It is particularly important to note that the $\Delta = 4$ case is included in this result, since this class of graphs contains the grid which is considered an important problem.

Remark. Note that the result from Theorem 16 also matches the result of Theorem 30 as well as holding for general bounded degree graphs. Theorem 30 remains interesting in its own right since it bounds the mixing time of a single-site systematic scan where as Theorem 16 uses a block dynamics. It is possible to obtain rapid mixing of a single-site chain from the result in Theorem 16 by using a comparison technique as previously discussed, however, at the expense of a polynomial factor loss in the mixing time.

Chapter 3

A Dobrushin Condition for Systematic Scan with Block Dynamics

In this chapter we study the mixing time of systematic scan Markov chains on finite spin systems in a general setting. It is known that, for single-site Markov chains, the mixing time of systematic scan can be bounded in terms of the influences sites have on each other. We generalise this technique for bounding the mixing time of systematic scan to block updates, a setting in which a (constant size) set of sites are updated simultaneously. In particular we introduce a parameter α, corresponding to the maximum influence on any site in the system, and show that if $\alpha < 1$ then the corresponding systematic scan Markov chain mixes in $O(\log n)$ scans.

As applications of this method we prove rapid mixing of two systematic scan Markov chains on proper q-colourings of a graph for any scan order. The first systematic scan that we consider performs heat-bath updates on edges of a general graph with maximum vertex-degree Δ and mixes in $O(\log n)$ scans when $q \geq 2\Delta$. The second systematic scan performs heat-bath updates on some suitable block when the graph is a tree with height H. The number of colours required for $O(H)$ mixing of this chain is lower than previous bounds.

We conclude the chapter with a discussion of the influence parameter α and how it relates to the corresponding parameters for the "influence on a site" in Weitz [55] and Dyer et al. [18]. In particular we will show that the condition in Weitz [55], which is for a random update Markov chain, does not imply mixing of systematic scan. We also show that the condition in Dyer et al. [18], for a single-site systematic scan, is a special case of our condition namely $\alpha < 1$.

3.1 Preliminaries

When analysing the mixing time of Markov chains it can be useful, and sometimes necessary, to consider chains that make use of block updates. A block update is a move of the chain that may change the spin assigned to more one site during each step of the process, as long as the number of sites that are being updated is constant. Block updates as a proof technique was used in the mid 1980s by Dobrushin and Shlosman [13] in their study of conditions that imply uniqueness of the Gibbs measure of a spin system, a topic closely related to studying the mixing time of Markov chains. Recently Weitz [55] used block updates in a generalisation of the work of Dobrushin and Shlosman, studying the relationship between two key influence parameters within spin systems and using the influence parameters to establish conditions that imply mixing. We will bound the mixing time of a systematic scan Markov chain by studying one of these influence parameters, although in a slightly different form. We will show that if "the influence *on* a site is small" in an appropriate sense then we can obtain rapid mixing of a systematic scan Markov chain. We call this a Dobrushin condition as it is similar to the types of conditions originally considered by Dobrushin [12].

We begin by reminding the reader of some terms and definitions from Chapter 2. First, recall from Definition 12 that $\mathcal{M}_{\rightarrow}$ is a systematic scan Markov chain with state space Ω^+ and transition matrix $P_{\rightarrow} = \prod_{k=1}^{m} P^{[k]}$ where $P^{[k]}$ is any valid update rule. Also recall from (2.5) that the influence of a site i on a site j under a block Θ_k, denoted by $\rho_{i,j}^k$, is the maximum probability that two coupled Markov chains differ at the spin of site j following an update of Θ_k starting from two configurations that only differ at the spin on site i. That is

$$\rho_{i,j}^k = \max_{(x,y)\in S_i} \{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j)\}.$$

The total (weighted) influence *on* any site in the graph site defined by

$$\alpha = \max_k \max_{j\in\Theta_k} \sum_i \frac{w_i}{w_j} \rho_{i,j}^k$$

where w_i is a positive weight assigned to each site of the spin system. We will use these definitions to prove Theorem 14 namely the following.

Theorem 14. *Consider any spin system with underlying graph $G = (V, E)$. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be any set of blocks covering V. For each block Θ_k let $P^{[k]}$ be a valid update rule associated with block Θ_k. $\mathcal{M}_{\rightarrow}$ is the system-*

atic scan Markov chain which updates the blocks in the order $\Theta_1, \ldots, \Theta_m$. *If* $\alpha = \max_k \max_{j \in \Theta_k} \sum_{i \in V} \rho_{i,j}^k w_i / w_j < 1$ *then* $\mathcal{M}_{\rightarrow}$ *is ergodic and its mixing time is at most*

$$\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \leq \frac{\log(n\gamma\varepsilon^{-1})}{1-\alpha}$$

scans of the graph where

$$\gamma = \frac{\max_{i \in V} w_i}{\min_{j \in V} w_j}$$

is the maximum ratio between the weights.

As previously stated we will apply Theorem 14 to two spin systems corresponding to proper q-colourings of graphs in order to improve the parameters for which systematic scan mixes. In both applications we restrict the state space of the Markov chains to the set of proper colourings, Ω, of the underlying graph. Firstly we allow the underlying graph to be any finite graph with maximum vertex-degree Δ. Recall from Definition 15 that $\mathcal{M}_{\mathrm{edge}}$ is a systematic scan Markov chain that updates each endpoint of an edge during each move. In particular recall that $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ is any set of m edges in G such that Θ covers V and that $P^{[k]}$ is the transition matrix for performing a heat-bath move on the endpoints of the edge Θ_k. The transition matrix of $\mathcal{M}_{\mathrm{edge}}$ is $\prod_{k=1}^m P^{[k]}$. We prove Theorem 16 which, we remind the reader, improves the mixing time of systematic scan by a factor of n^2 for proper colourings of general graphs when $q = 2\Delta$ and matches an existing bound when $q > 2\Delta$.

Theorem 16. *Let* G *be a graph with maximum vertex-degree* Δ. *Consider the systematic scan Markov chain* $\mathcal{M}_{\mathrm{edge}}$ *on* Ω. *If* $q \geq 2\Delta$ *then the mixing time of* $\mathcal{M}_{\mathrm{edge}}$ *is*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{edge}}, \varepsilon) \leq \Delta^2 \log(n\varepsilon^{-1})$$

scans. If $m = O(n)$ *then this corresponds to* $O(n \log n)$ *block updates.*

Next we restrict the class of graphs to trees. Recall from Definition 17 that $\mathcal{M}_{\mathrm{tree}}$ is the (single-site) systematic scan Markov chain with state space Ω and transition matrix $\prod_{k=1}^n P^{[k]}$ where $P^{[k]}$ is the transition matrix for performing a heat-bath move on block $\Theta_k = \{k\}$ for each $k \in V$. We prove Theorem 18 and remind the reader that this theorem matches existing bounds as discussed previously.

Theorem 18. *Let* G *be a tree with maximum vertex degree* $\Delta \geq 3$ *and height* H. *Consider the systematic scan Markov chain* $\mathcal{M}_{\mathrm{tree}}$ *on* Ω. *If* $q \geq \Delta + 2\sqrt{\Delta - 1 + \delta}$

Table 3.1. Optimising the number of colours using blocks

Δ	h	ξ	$f(\Delta)$	$\lceil \Delta + 2\sqrt{\Delta - 1} \rceil$
3	15	$\frac{4}{7}$	5	6
4	3	$\frac{5}{11}$	7	8
5	12	$\frac{5}{11}$	8	9
6	3	$\frac{1}{2}$	10	11
7	7	$\frac{10}{23}$	11	12
8	13	$\frac{1}{3}$	12	14
9	85	$\frac{5}{19}$	13	15
10	5	$\frac{5}{19}$	15	16
20	21	$\frac{3}{20}$	27	29
30	117	$\frac{3}{20}$	38	41
40	50	$\frac{57}{500}$	49	53
50	150	$\frac{101}{1000}$	60	64
60	51	$\frac{19}{200}$	71	76
100	45	$\frac{7}{100}$	115	120

for $\delta > 0$ then the mixing time of $\mathcal{M}_{\text{tree}}$ is

$$\text{Mix}(\mathcal{M}_{\text{tree}}, \varepsilon) \leq \max\left(\frac{2(\Delta - 1 + \delta)}{\delta}, 4\right)\left(H \log\left(\frac{q - \Delta}{2(\Delta - 1)}\right) + \log(n\varepsilon^{-1})\right)$$

scans of the tree. Since $\log n \leq H \leq n$, this corresponds to $O(nH)$ updates.

The number of colours required for rapid mixing of systematic scan for sampling proper colourings of trees can be reduced for individual values of Δ by using some suitable block updates. Recall from Definition 19 that $\mathcal{M}_{\text{BlockTree}}$ is the systematic scan Markov chain with state space Ω and transition matrix $\prod_{k=1}^{m} P^{[k]}$ where $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k. The blocks are constructed as follows. We construct the following set of blocks where the height h of the blocks is defined in Table 2.1 (repeated in Table 3.1). Let a block Θ_k contain a site r along with all sites below r in the tree that are at most $h - 1$ edges away from r. The values for h are given in Table 2.1 (repeated in Table 3.1). The set of blocks Θ is constructed such that it covers the sites of the tree and no block has height less than h. We prove Theorem 20 which improves the number of colours required for rapid mixing of systematic scan for the stated values of Δ.

Theorem 20. *Let G be a tree with maximum vertex-degree Δ and height H. Consider the systematic scan Markov chain $\mathcal{M}_{\text{BlockTree}}$ on Ω. If $q \geq f(\Delta)$ where $f(\Delta)$ is specified in Table* 2.1 *(repeated in Table 3.1) for small Δ then the mixing*

time of $\mathcal{M}_{\text{BlockTree}}$ *is*

$$\text{Mix}(\mathcal{M}_{\text{BlockTree}}, \varepsilon) = O(H + \log(\varepsilon^{-1}))$$

scans of the tree. This corresponds to $O(nH)$ *block updates by the construction of the set of blocks.*

3.2 Bounding the Mixing Time of Systematic Scan

This section contains the proof of Theorem 14. The proof follows the structure of the proof from the single-site setting in Dyer et al. [18], which follows Föllmer's [28] account of Dobrushin's proof presented in Simon's book [51].

We will make use of the following definitions. For any function $f : \Omega^+ \to \mathbb{R}_{\geq 0}$ let $\delta_i(f) = \max_{(x,y)\in S_i} |f(x) - f(y)|$ and $\Delta(f) = \sum_{i\in V} w_i \delta_i(f)$. Also for any transition matrix P define (Pf) as the function from Ω^+ to $\mathbb{R}_{\geq 0}$ given by $(Pf)(x) = \sum_{x'} P(x, x')f(x')$. Finally let $\mathbf{1}_{i\notin\Theta_k}$ be the indicator function given by

$$\mathbf{1}_{i\notin\Theta_k} = \begin{cases} 1 & \text{if } i \notin \Theta_k \\ 0 & \text{otherwise.} \end{cases}$$

We can think of $\delta_i(f)$ as the deviation from constancy of f at site i and $\Delta(f)$ as the aggregated deviation from constancy of f. Now, Pf is a function where $(Pf)(x)$ gives the expected value of f after making a transition starting from x. Intuitively, if t transitions are sufficient for mixing then $P^t f$ is a very smooth function. An application of $P^{[k]}$ fixes the non-constancy of f at the sites within Θ_k although possibly at the cost of increasing the non-constancy at sites on the boundary of Θ_k. Our aim is then to show that one application of $P_{\rightarrow}$ will on aggregate make f smoother i.e., decrease $\Delta(f)$. We will establish the following lemma, which corresponds to Corollary 12 in Dyer et al. [18], from which Section 3.3 of [18] implies Theorem 14.

Lemma 31. *If* $\alpha < 1$ *then*

$$\Delta(P_{\rightarrow} f) \leq \alpha \Delta(f).$$

We begin by bounding the effect on f from one application of $P^{[k]}$. The following lemma is a block-move generalisation of Proposition V.1.7 from Simon [51]

and Lemma 10 from Dyer et al. [18].

Lemma 32. $\delta_i(P^{[k]}f) \leq \mathbf{1}_{i \notin \Theta_k}\delta_i(f) + \sum_{j\in\Theta_k} \rho^k_{i,j}\delta_j(f)$

Proof. Take $\mathbf{E}_{(x',y')\in\Psi_k(x,y))}\left[f(x')\right]$ to be the the expected value of $f(x')$ when a pair of configurations (x', y') are drawn from $\Psi_k(x, y)$. Since $\Psi_k(x, y)$ is a coupling of the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$, the distribution $P^{[k]}(x, \cdot)$ and the first component of $\Psi_k(x, y)$ are the same and hence

$$\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[f(x')\right] = \mathbf{E}_{x'\in P^{[k]}(x,\cdot)}\left[f(x')\right] \tag{3.1}$$

and the same fact holds for the distribution $P^{[k]}(y, \cdot)$ so

$$\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[f(y')\right] = \mathbf{E}_{y'\in P^{[k]}(y,\cdot)}\left[f(y')\right]. \tag{3.2}$$

Using (3.1), (3.2) and linearity of expectation we have

$$\begin{aligned}
\delta_i(P^{[k]}f) &= \max_{(x,y)\in S_i} \left|(P^{[k]}f)(x) - (P^{[k]}f)(y)\right| \\
&= \max_{(x,y)\in S_i} \left|\sum_{x'} P^{[k]}(x,x')f(x') - \sum_{y'} P^{[k]}(y,y')f(y')\right| \\
&= \max_{(x,y)\in S_i} \left|\mathbf{E}_{x'\in P^{[k]}(x,\cdot)}\left[f(x')\right] - \mathbf{E}_{y'\in P^{[k]}(y,\cdot)}\left[f(y')\right]\right| \\
&= \max_{(x,y)\in S_i} \left|\mathbf{E}_{(x',y')\in\Psi_k(x,y))}\left[f(x')\right] - \mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[f(y')\right]\right| \\
&= \max_{(x,y)\in S_i} \left|\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[f(x') - f(y')\right]\right| \\
&\leq \max_{(x,y)\in S_i} \mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[|f(x') - f(y')|\right] \\
&\leq \max_{(x,y)\in S_i} \mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[\sum_{j\in V}\left|f(x'_1\ldots x'_j y'_{j+1}\ldots y'_n) - f(x'_1\ldots x'_{j-1}y'_j\ldots y'_n)\right|\right] \\
&= \max_{(x,y)\in S_i} \sum_{j\in V} \mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[\left|f(x'_1\ldots x'_j y'_{j+1}\ldots y'_n) - f(x'_1\ldots x'_{j-1}y'_j\ldots y'_n)\right|\right].
\end{aligned}$$

Notice that $x = x'$ off Θ_k and $y = y'$ off Θ_k.

We need to bound the expectation

$$\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[\left|f(x'_1\ldots x'_j y'_{j+1}\ldots y'_n) - f(x'_1\ldots x'_{j-1}y'_j\ldots y'_n)\right|\right]$$

for each site $j \in V$. First suppose that $j \in \Theta_k$. By definition of $\rho^k_{i,j}$ the coupling

will yield $x'_j \neq y'_j$ with probability at most $\rho^k_{i,j}$ and so

$$\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[\left|f(x'_1\dots x'_j y'_{j+1}\dots y'_n) - f(x'_1\dots x'_{j-1}y'_j\dots y'_n)\right|\right] \\ \leq \rho^k_{i,j}\max_{(\sigma,\tau)\in S_j}\{|f(\sigma)-f(\tau)|\} = \rho^k_{i,j}\delta_j(f).$$

Otherwise $j \notin \Theta_k$ and we observe that $x_j = x'_j$ since $x = x'$ off Θ_k and similarly $y_j = y'_j$ since $y = y'$ of Θ_k. Since $(x, y) \in S_i$ we can only have $x'_j \neq y'_j$ when $i = j$ and hence

$$\mathbf{E}_{(x',y')\in\Psi_k(x,y)}\left[\left|f(x'_1\dots x'_j y'_{j+1}\dots y'_n) - f(x'_1\dots x'_{j-1}y'_j\dots y'_n)\right|\right] \leq \mathbf{1}_{i=j}\delta_j(f).$$

Adding up the expectations up we get the statement of the lemma. □

We will use Lemma 32 in conjunction with an inductive proof similar to (V.1.16) in Simon [51] in order to establish the following lemma. It is important to note at this point that the result in Simon is presented for single-site heat-bath updates, whereas the following lemma applies to any block dynamics (satisfying the stated assumptions) and weighted sites. This lemma is also a block generalisation of Lemma 11 in Dyer et al. [18].

Lemma 33. *For any $k \in \{1,\dots,m\}$ let $\Gamma(k) = \bigcup_{l=1}^{k}\Theta_l$. If $\alpha < 1$ then*

$$\Delta(P^{[1]}\cdots P^{[k]}f) \leq \alpha\sum_{i\in\Gamma(k)} w_i\delta_i(f) + \sum_{i\in V\setminus\Gamma(k)} w_i\delta_i(f).$$

Proof. Induction on k. Taking $k = 0$ as the base case, we get the definition of Δ.

Assume the statement holds for $k-1$.

$$\begin{aligned}\Delta(P^{[1]}\cdots P^{[k]}f) &\leq \alpha\sum_{i\in\Gamma(k-1)} w_i\delta_i(P^{[k]}f) + \sum_{i\in V\setminus\Gamma(k-1)} w_i\delta_i(P^{[k]}f) \\ &\leq \alpha\sum_{i\in\Gamma(k-1)}\mathbf{1}_{i\notin\Theta_k}w_i\delta_i(f) + \alpha\sum_{i\in\Gamma(k-1)}\sum_{j\in\Theta_k} w_i\rho^k_{i,j}\delta_j(f) \\ &\quad + \sum_{i\in V\setminus\Gamma(k-1)}\mathbf{1}_{i\notin\Theta_k}w_i\delta_i(f) + \sum_{i\in V\setminus\Gamma(k-1)}\sum_{j\in\Theta_k} w_i\rho^k_{i,j}\delta_j(f)\end{aligned}$$

by Lemma 32.

Simplifying and using $\alpha < 1$

$$\begin{aligned}
\Delta(P^{[1]} \cdots P^{[k]} f) &\leq \alpha \sum_{i \in \Gamma(k-1) \setminus \Theta_k} w_i \delta_i(f) + \sum_{i \in \Gamma(k-1)} \sum_{j \in \Theta_k} w_i \rho_{i,j}^k \delta_j(f) \\
&\quad + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k-1)} \sum_{j \in \Theta_k} w_i \rho_{i,j}^k \delta_j(f) \\
&= \alpha \sum_{i \in \Gamma(k-1) \setminus \Theta_k} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f) \\
&\quad + \sum_{j \in \Theta_k} \delta_j(f) \left(\sum_{i \in \Gamma(k-1)} w_i \rho_{i,j}^k + \sum_{i \in V \setminus \Gamma(k-1)} w_i \rho_{i,j}^k \right) \\
&= \alpha \sum_{i \in \Gamma(k-1) \setminus \Theta_k} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f) + \sum_{j \in \Theta_k} \delta_j(f) \sum_{i \in V} w_i \rho_{i,j}^k \\
&\leq \alpha \sum_{i \in \Gamma(k-1) \setminus \Theta_k} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f) + \sum_{j \in \Theta_k} \delta_j(f) \max_l \sum_{i \in V} w_i \rho_{i,j}^l \\
&\leq \alpha \sum_{i \in \Gamma(k-1) \setminus \Theta_k} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f) + \alpha \sum_{j \in \Theta_k} w_j \delta_j(f) \\
&= \alpha \sum_{i \in \Gamma(k)} w_i \delta_i(f) + \sum_{i \in V \setminus \Gamma(k)} w_i \delta_i(f)
\end{aligned}$$

by definition of α. □

Lemma 31 is now a simple consequence of Lemma 33 since

$$\Delta(P_{\rightarrow} f) = \Delta(P^{[1]} \cdots P^{[m]} f) \leq \alpha \sum_{i \in V} w_i \delta_i(f) = \alpha \Delta(f)$$

and Theorem 14 follows as discussed above. For completeness we do however give a proof of Theorem 14. The following lemma is required for technical reasons in that proof.

Lemma 34. $\max_{\omega \in \Omega^+} f(\omega) - \min_{\omega \in \Omega^+} f(\omega) \leq \Delta(f) / \min_{j \in V} w_j$.

Proof. Let $x, y \in \Omega^+$ be such that $\max_{\omega \in \Omega^+} f(\omega) = f(x)$ and $\min_{\omega \in \Omega^+} f(\omega) = f(y)$. For each $i \in \{1, \ldots, n\}$ let $\Theta_i = \{i\}$. Construct a path of colourings

$x = z^0, \ldots, z^n = y$ where $z^i = z^{i-1}$ off Θ_i and $z^i_i = y_i$ for all $i \in \{1, \ldots, n\}$. Then

$$\begin{aligned}\max_{\omega\in\Omega^+} f(\omega) - \min_{\omega\in\Omega^+} f(\omega) &= f(x) - f(y) \\ &= \sum_{i=0}^{n-1} f(z^i) - f(z^{i+1}) \\ &\leq \sum_{i=1}^{n} \delta_i(f) \\ &\leq \frac{1}{\min_{j\in V} w_j} \sum_{i=1}^{n} w_i \delta_i(f) \\ &= \frac{\Delta(f)}{\min_{j\in V} w_j}\end{aligned}$$

by definition of δ and Δ. □

We are now in position to establish a proof of Theorem 14.

Theorem 14. *Consider any spin system with underlying graph $G = (V, E)$. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be any set of blocks covering V. For each block Θ_k let $P^{[k]}$ be a valid update rule associated with block Θ_k. $\mathcal{M}_{\rightarrow}$ is the systematic scan Markov chain which updates the blocks in the order $\Theta_1, \ldots, \Theta_m$. If $\alpha = \max_k \max_{j\in\Theta_k} \sum_{i\in V} \rho^k_{i,j} w_i/w_j < 1$ then $\mathcal{M}_{\rightarrow}$ is ergodic and its mixing time is at most*

$$\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \leq \frac{\log(n\gamma\varepsilon^{-1})}{1-\alpha}$$

scans of the graph where

$$\gamma = \frac{\max_{i\in V} w_i}{\min_{j\in V} w_j}$$

is the maximum ratio between the weights.

Proof. For a test function f, let $f_t(x) = \sum_{\omega\in\Omega^+} P^t_{\rightarrow}(x, \omega) f(\omega)$ with the intention $f_t = P^{[1]} \cdots P^{[m]} f_{t-1}$.

We use a lemma from Aldous and Fill [3] to deduce

$$\begin{aligned}\max_{x\in\Omega^+} \mathrm{d_{TV}}(P^t_{\rightarrow}(x, \cdot), \pi) &\leq \max_{x,y\in\Omega^+} \mathrm{d_{TV}}(P^t_{\rightarrow}(x, \cdot), P^t_{\rightarrow}(y, \cdot)) \\ &= \max_{x,y\in\Omega^+} \max_{A\subseteq\Omega^+} |P^t_{\rightarrow}(x, A) - P^t_{\rightarrow}(y, A)|\end{aligned}$$

using the definition of total variation distance. Letting f be the indicator variable

for being in some subset A of Ω^+ we have

$$\begin{aligned} P^t_{\rightarrow}(x, A) - P^t_{\rightarrow}(y, A) &= \sum_{\omega \in \Omega^+} P^t_{\rightarrow}(x, \omega) f(\omega) - \sum_{\omega \in \Omega^+} P^t_{\rightarrow}(y, \omega) f(\omega) \\ &\leq \max_{\omega \in \Omega^+} f_t(\omega) - \min_{\omega \in \Omega^+} f_t(\omega) \\ &\leq \frac{\Delta(f_t)}{\min_{j \in V} w_j} \end{aligned}$$

by Lemma 34. Applying Lemma 31 t times gives

$$\frac{\Delta(f_t)}{\min_{j \in V} w_j} \leq \frac{\alpha^t \Delta(f_0)}{\min_{j \in V} w_j} \leq \frac{\alpha^t n \max_{i \in V} w_i}{\min_{j \in V} w_j}$$

which is at most ε for $t \geq \frac{\log(n\gamma\varepsilon^{-1})}{1-\alpha}$. □

3.3 Application: Edge Scan on an Arbitrary Graph

In this section we prove Theorem 16. That is, we present a general version of a systematic scan on edges and use Theorem 14 to prove that it mixes in $O(\log n)$ scans whenever $q \geq 2\Delta$. We use uniform weights for the sites and so omit all weights throughout this section. Recall that $\mathcal{M}_{\text{edge}}$ is the systematic scan Markov chain with transition matrix $\prod_{k=1}^{m} P^{[k]}$ where $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ is an ordered set of edges in G that covers V and $P^{[k]}$ is the transition matrix for performing a heat-bath move on the endpoints of the edge Θ_k.

In order to apply Theorem 14 we extend the chain to the state space Ω^+ such that the extended chain is identical to $\mathcal{M}_{\text{edge}}$ on configurations in Ω. Furthermore, the extended chain never makes a transition from a configuration in Ω to a configuration outside Ω. Observe that for any given configuration it is possible to update the endpoints of any edge in G in such a way that both endpoints of that edge are coloured properly. Hence the configurations in $\Omega^+ \setminus \Omega$ are transient states of the extended chain and an upper bound on the mixing time of the extended chain is also an upper bound on the mixing time of $\mathcal{M}_{\text{edge}}$ by Lemma 8. As previously discussed, extending the state space of the chain in this way is a standard technique.

We need to construct a coupling $\Psi_k(x, y)$ of the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ for each pair of configurations $(x, y) \in S_i$ that differ only at the colour assigned to site i. Assume without loss of generality that $x_i = 1$ and $y_i = 2$ and also let j and j' be the endpoints of the edge Θ_k. Recall from Example 10 that,

since the dynamics uses heat-bath updates, $P^{[k]}(x,\cdot)$ is the uniform distribution on configurations that agree with x off Θ_k and where no edge containing j or j' is monochromatic. For ease of notation we let $D_1 = P^{[k]}(x,\cdot)$ and $D_2 = P^{[k]}(y,\cdot)$. We go on to make the following definitions for $l \in \{1,2\}$ and $s \in \Theta_k$. $D_l(s)$ is the distribution of the colour assigned to site s induced by D_l, and $[D_l \mid s = c]$ is the uniform distribution on the set of colourings of the sites in Θ_k where site s is assigned colour c. We also let d_l denote the number of configurations with positive measure in D_l and $d_{l,s=c}$ be the number of configurations that assign colour c to site s and have positive measure in D_l.

Definition 35. *For $c_1, c_2 \in C$ we say that the choice c_1c_2 is "valid" for D_l if there is a configuration with positive measure in D_l in which site j is coloured c_1 and site j' is coloured c_2. Similarly a colour $c \in C$ is "valid" on a site s in D_l if there exists a valid choice for D_l where site s is coloured c.*

3.3.1 Overview of the Coupling

We begin the construction of the coupling $\Psi_k(x,y)$ by giving an overview of the cases we will need to consider and show that they are mutually exclusive and exhaustive of all configurations. It is important to note that, by definition of $\rho^k_{i,j}$, the coupling we define may depend on the initial configurations x and y in the sense that if two pairs of configurations (x_1,y_1) and (x_2,y_2) can be distinguished then the couplings $\Psi_k(x_1,y_1)$ and $\Psi_k(x_2,y_2)$ may be defined differently.

We consider two simple cases in the coupling construction. First, if $i \notin \partial\Theta_k$ then $\Psi_k(x,y)$ is the identity coupling where the same choice is made in both distributions. Hence, for $i \notin \partial\Theta_k$ and $j \in \Theta_k$ we have $\rho^k_{i,j} = 0$. In particular, observe that this case includes the situation when $i \in \Theta_k$.

Now suppose that i is adjacent to at least one site in Θ_k, that is $i \in \partial\Theta_k$. In order to construct a sufficiently good coupling we consider the following five sub-cases, which by construction are exhaustive of all possible configurations and mutually exclusive. In the diagrams that relate to these cases a dotted line between a site $j \in \Theta_k$ and a colour 1, say, denotes that no site adjacent to j on the boundary of Θ_k (other than possibly i) is coloured 1. A full line denotes that some site adjacent to j on the boundary of Θ_k (other than possibly i) is coloured 1. The full details of each case of the coupling will be given in Section 3.3.2 along with bounds on $\rho^k_{i,j}$ and $\rho^k_{i,j'}$ where j and j' are the sites included in Θ_k.

1. Exactly one site in Θ_k is adjacent to i. Let this site be labeled j and let the other site in Θ_k be labeled j'. This is shown in Figure 3.1.

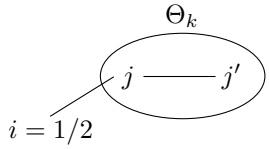

Figure 3.1. Case 1. Exactly one site in Θ_k is adjacent to i. Let this site be labeled j and let the other site in Θ_k be labeled j'.

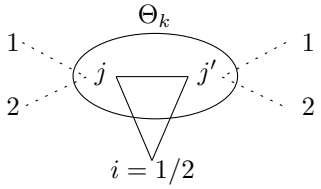

Figure 3.2. Case 2. Both sites in Θ_k are adjacent to i and no other sites in $\partial\Theta_k$ are coloured 1 or 2. The labeling of the sites in Θ_k is arbitrary.

2. Both sites in Θ_k are adjacent to i and no other sites in $\partial\Theta_k$ are coloured 1 or 2. The labeling of the sites in Θ_k is arbitrary. This is shown in Figure 3.2.

3. Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 (or 2). Let this site be labeled j'. The other site in Θ_k is labeled j and it is not adjacent to any site, other than i, coloured 1 or 2. This is shown in Figure 3.3.

4. Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 and no sites that are coloured 2. Let this site be labeled j'. The other site in Θ_k, labeled j, is adjacent to at least one site other than i coloured 2 and no sites coloured 1. This is shown in Figure 3.4.

5. Both sites in Θ_k are adjacent to i and at least one site, other than i coloured 1 (or 2). The labeling of the sites in Θ_k is arbitrary. This is shown in

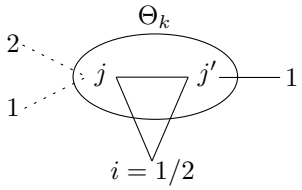

Figure 3.3. Case 3. Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 (or 2). Let this site be labeled j'. The other site in Θ_k is labeled j and it is not adjacent to any site, other than i, coloured 1 or 2.

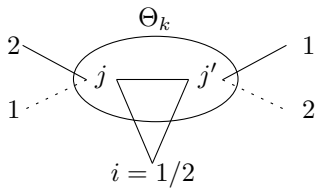

Figure 3.4. Case 4. Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 and no sites that are coloured 2. Let this site be labeled j'. The other site in Θ_k, labeled j, is adjacent to at least one site other than i coloured 2 and no sites coloured 1.

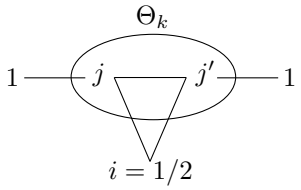

Figure 3.5. Case 5. Both sites in Θ_k are adjacent to i and at least one site, other than i coloured 1 (or 2). The labeling of the sites in Θ_k is arbitrary.

Figure 3.5.

3.3.2 Details of Coupling and Proof of Mixing

We will now give the full details of each case of the coupling and establish the required bounds on the influence of site i on sites j and j'. The following lemma is required to establish the coupling for all the stated cases.

Lemma 36. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i, j) \in E$. Then for each pair of colours $c_1, c_2 \in C \setminus \{1, 2\}$ the choice c_1c_2* is valid for D_1 *if and only if c_1c_2* is valid for D_2.

Proof. We start with the *if* direction. Suppose c_1c_2 is valid in D_2 then no site adjacent to j has colour c_1 in D_2 and since $c_1 \neq 1$ no site adjacent to j has colour c_1 in D_1. Also no site adjacent to j' has colour c_2 in D_2 hence no site adjacent to j' has colour c_2 in D_1 since $c_2 \neq 1$. Since c_1c_2 is valid in D_2 $c_1 \neq c_2$ and so c_1c_2 is valid in D_1.

The *only if* direction is similar. Suppose c_1c_2 is valid in D_1 then no site adjacent to j has colour c_1 in D_1 and since $c_1 \neq 2$ no site adjacent to j has colour c_1 in D_2. Also no site adjacent to j' has colour c_2 in D_1 hence no site adjacent to j' has colour c_2 in D_2 again since $c_2 \neq 2$. Since c_1c_2 is valid in D_1 $c_1 \neq c_2$ and so c_1c_2 is valid in D_2. □

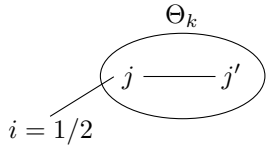

Figure 3.6. Case 1 (repeat of Figure 3.1). Exactly one site in Θ_k is adjacent to i. Let this site be labeled j and let the other site in Θ_k be labeled j'.

Details of case 1. (Repeated in Figure 3.6.) We construct a coupling $\Psi_k(x, y)$ of the distributions D_1 and D_2 using the following two step process. Let ψ_j be a coupling of $D_1(j)$ and $D_2(j)$ which greedily maximises the probability of assigning the same colour to site j in each distribution. Then, for each pair of colours (c, c') drawn from ψ_j, $\Psi_k(x, y)$ is a coupling, minimising Hamming distance, of the conditional distributions $D_1 \mid j = c$ and $D_2 \mid j = c'$.

Remark. The reason for defining the coupling $\Psi_k(x, y)$ recursively is that this particular coupling construction lets us upper bound the probability of a discrepancy at site j in a pair of configurations drawn from the coupling $\Psi_k(x, y)$ by assuming that j' is assigned the worst case colour. This is due to Lemma 13 of Goldberg et al. [33]. For completeness we state a special case of this lemma, which is sufficient for our needs, although we point out that the original lemma is stated for a more general case.

Lemma (Special case of Lemma 13 in Goldberg et al. [33])**.** *Let $\Psi_k(x, y)$ be the above coupling. For any $(\sigma, \tau) \in S_i$, let μ_j be a coupling, minimising Hamming distance at j, of the distributions obtained by performing a heat-bath move on site j starting from configuration σ and τ respectively. Then for any $(x, y) \in S_i$*

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j) \leq \max_{(\sigma,\tau)\in S_i} \Pr_{(\sigma',\tau')\in\mu_j}(\sigma'_j \neq \tau'_j).$$

Lemma 37. *Let j and j' be the endpoints of an edge Θ_k. If $(i, j) \in E$ and $(i, j') \notin E$ then*

$$\rho^k_{i,j} \leq \frac{1}{q - \Delta} \text{ and } \rho^k_{i,j'} \leq \frac{1}{(q - \Delta)^2}.$$

Proof. Assume without loss of generality that $d_1 \geq d_2$, i.e that there are at least as many valid choices for D_1 as for D_2. Since the only site in Θ_k that is adjacent to site i is j, Lemma 13 of Goldberg et al. [33] lets us upper bound the probability of a discrepancy at site j in a pair of configurations drawn from the coupling $\Psi_k(x, y)$ by assuming that j' is assigned the worst case colour. Observe that site

j has at most $\Delta - 1$ neighbours (excluding j') and each of them could invalidate one colour choice for site j in both distributions. If j' is assigned a (worst case) colour not already adjacent to j then site j is adjacent to at most Δ sites each assigned a different colour. This leaves at least $q - \Delta$ valid colours for j in D_1. Since 1 is not valid for j in D_1, Lemma 36 implies that colour 2 is the only valid choice for j in D_1 which would cause a discrepancy at site j since the first step of the coupling is greedy. This establishes the stated bound on $\rho^k_{i,j}$

$$\rho^k_{i,j} = \max_{(x,y)\in S_i} \{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j)\} \leq \frac{1}{q-\Delta}.$$

Now from the definition of the coupling it follows easily that if the same colour, c, is assigned to site j in each distribution during the first step of the coupling then the colour assigned to site j' in the second step will be the same in each distribution since the conditional distributions $D_1 \mid j = c$ and $D_2 \mid j = c$ are the same. If different colours are assigned to j in each distribution then the second step of the coupling is simply the case of colouring one site adjacent to exactly one discrepancy. The argument from above says that at most one colour assigned to j' in D_1 will cause a discrepancy at site j' in the coupling and also that there are at least $q - \Delta$ valid choices for j' in D_1. Hence we have $\max_{(x,y)\in S_i}\{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'} \mid x'_j = c, y'_j = c')\} \leq \frac{1}{q-\Delta}$ and so

$$\begin{aligned}
\rho^k_{i,j'} &= \max_{(x,y)\in S_i} \{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'})\} \\
&= \max_{(x,y)\in S_i} \left\{ \sum_{\substack{c,c' \\ c\neq c'}} \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'} \mid x'_j = c, y'_j = c') \right. \\
&\qquad\qquad \left. \times \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j = c, y'_j = c') \right\} \\
&\leq \frac{1}{q-\Delta} \max_{(x,y)\in S_i} \left\{ \sum_{\substack{c,c' \\ c\neq c'}} \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j = c, y'_j = c') \right\} \\
&\leq \frac{1}{(q-\Delta)^2}
\end{aligned}$$

using the bound from $\rho^k_{i,j}$ which completes the proof. □

The following lemmas are required to define the coupling and bound the influence of a site $i \in \partial\Theta_k$ on sites j and j' when i is adjacent to both sites j and

j'.

Lemma 38. *Let j and j' be the endpoints of an edge and suppose that $(i,j) \in E$ and $(i,j') \in E$. If 1 is valid for j in D_2 and 2 is valid for j in D_1 then the choice $2c_2$ is valid in D_1 if and only if $1c_2$ is valid in D_2.*

Proof. Suppose that $2c_2$ is valid in D_1 then $c_2 \in C \setminus \{1,2\}$ since i is adjacent to j' (and $x_i = 1$). Since 1 is valid for j in D_2 it follows that $1c_2$ is valid in D_2 since the only colour adjacent to j' in D_2 that is (possibly) not adjacent to j' in D_1 is 2, but $c_2 \neq 2$.

For the reverse direction suppose that $1c_2$ is valid in D_2. Then $c_2 \in C \setminus \{1,2\}$ since i is adjacent to j'. Since 2 is valid for j in D_1 it follows that $2c_2$ is valid in D_1 since the only colour adjacent to j' in D_1 that is (possibly) not adjacent to j' in D_2 is 1, but $c_2 \neq 1$. □

Lemma 39. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i,j) \in E$ and $(i,j') \in E$. If 1 is valid for j' in D_2 and 2 is valid for j' in D_1 then the choice $c_1 2$ is valid in D_1 if and only if $c_1 1$ is valid in D_2.*

Proof. Suppose that $c_1 2$ is valid in D_1 then $c_1 \in C \setminus \{1,2\}$ since i is adjacent to j'. Since 1 is valid for j' in D_2 $c_1 1$ is valid in D_2 since the only colour adjacent to j in D_2 that is (possibly) not adjacent to j in D_1 is 2, but $c_1 \neq 2$.

Also, suppose that $c_1 1$ is valid in D_2 then $c_1 \in C \setminus \{1,2\}$ since i is adjacent to j'. Since 2 is valid for j' in D_1 $c_1 2$ is valid in D_1 since the only colour adjacent to j in D_1 that is (possibly) not adjacent to j in D_2 is 1, but $c_1 \neq 1$. □

Lemma 40. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i,j) \in E$ and $(i,j') \in E$.*

(i) Suppose that 1 is valid for j in D_2. For all $c \in C$ where c is valid for j in D_2, if 1 is valid for j' in D_2 then

$$d_{2,j=1} \leq d_{2,j=c} \leq d_{2,j=1} + 1$$

else

$$d_{2,j=1} - 1 \leq d_{2,j=c} \leq d_{2,j=1}.$$

(ii) Suppose that 2 is valid for j in D_1. For all $c \in C$ where c is valid for j in D_1, if 2 is valid for j' in D_1 then

$$d_{1,j=2} \leq d_{1,j=c} \leq d_{1,j=2} + 1$$

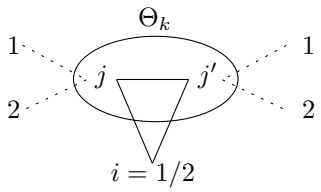

Figure 3.7. Case 2 (repeat of Figure 3.2). Both sites in Θ_k are adjacent to i and no other sites in $\partial\Theta_k$ are coloured 1 or 2. The labeling of the sites in Θ_k is arbitrary.

else

$$d_{1,j=2} - 1 \leq d_{1,j=c} \leq d_{1,j=2}.$$

Proof. Part (i). Consider some valid colour c other than 1 for j in D_2. For each valid choice $1c_2$ for D_2 the choice cc_2 is also valid for D_2 except when $c = c_2$. If 1 is valid for j' in D_2 then the choice $c1$ is also valid for D_2.

Now consider some invalid choice $1c_2$ for D_2 where $c_2 \neq 1$. Since $1c_2$ is not valid for D_2 it follows that c_2 is not valid for j' in D_2 and hence no more choices can be valid for D_2, which guarantees the upper bounds.

Part (ii) is similar. Consider some valid colour c other than 2 for j in D_1. For each valid choice $2c_2$ for D_1 the choice cc_2 is also valid for D_1 except when $c = c_2$. If 2 is valid for j' in D_1 then the choice $c2$ is also valid for D_1.

Finally consider some invalid choice $2c_2$ for D_1 where $c_2 \neq 2$. Since $2c_2$ is not valid for D_1 it follows that c_2 is not valid for j' in D_1 and hence no more choices can be valid for D_1, which guarantees the upper bounds. □

We are now ready to define the coupling for the remaining cases.

Details of case 2. (Repeated in Figure 3.7.) We construct the $\Psi_k(x, y)$ of the distributions D_1 and D_2 as follows. For each valid choice of the form c_1c_2 for D_1 where $c_1 \neq 2$ and $c_2 \neq 2$ Lemma 36 guarantees that c_1c_2 is valid for D_2 so we let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = y' = c_1c_2) = \frac{1}{d_1}.$$

For each valid choice of the form $2c_2$ in D_1 the choice $1c_2$ is valid in D_2 by Lemma 38 so we let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = 2c_2, y' = 1c_2) = \frac{1}{d_1}. \tag{3.3}$$

Lemma 38 also guarantees that there are no remaining valid choices for D_2 of the form $1c_2$. Finally for each valid choice c_12 for D_1 the choice c_11 is valid in D_2 by

Lemma 39 so let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = c_1 2, y' = c_1 1) = \frac{1}{d_1} \tag{3.4}$$

which completes the coupling since $d_1 = d_2$ and all the probability in both D_1 and D_2 has hence been used.

Lemma 41. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i,j) \in E$ and $(i,j') \in E$. If 2 is valid for both j and j' in D_1 and 1 is valid for both j and j' in D_2 then*

$$\rho^k_{i,j} \leq \frac{1}{q-\Delta+1} \text{ and } \rho^k_{i,j'} \leq \frac{1}{q-\Delta}.$$

Proof. This is case 2 of the coupling. Note from Lemma 38 that $d_{1,j=2} = d_{2,j=1}$ so for ease of reference let $d = d_{1,j=2} = d_{2,j=1}$ and let $d' = d_{1,j'=2} = d_{2,j'=1}$ (using Lemma 39). Also let $s = \sum_c d_{2,j=c} - d - d'$ which is the number of valid choices for D_2 other than choices of the form $1c_2$ and $c_1 1$. Note that the number of valid choices for D_1 is $d_1 = s + d + d'$.

As there are no restrictions on colours assigned to the sites in $\partial\Theta_k \setminus \{i\}$ each of the neighbours of j could be assigned a different colour, and the same is true for the neighbours of j'. Hence we get the following lower-bounds on d and d':

$$q - \Delta \leq d \text{ and } q - \Delta \leq d'.$$

To lower bound bound s observe that $s = \sum_c d_{2,j=c} - d - d' = \sum_{c\neq 1} d_{2,j=c} - d'$. Let $J \subseteq C \setminus \{1\}$ be the set of colours, excluding 1, that are valid for j in D_2. By definition of d', at least d' colours other than 1 must be valid for site j in D_2 so the size of J is at least d'. Since 1 is valid for j' in D_2 we use the lower bound on $d_{2,j=c}$ from Lemma 40 (i) and hence

$$\begin{aligned} s &= \sum_{c\in J} d_{2,j=c} - d' \\ &\geq d' \min_{c\in J}\{d_{2,j=c}\} - d' \\ &\geq d'd - d'. \end{aligned}$$

From the coupling, j will be assigned a different colour in each distribution whenever a choice of the form $2c_2$ is made for D_1. From (3.3) this happens with probability $\frac{d}{d_1} = \frac{d}{d+d'+s}$ since d is the number of valid choices for D_1 of the form $2c_2$. Similarly from (3.4), j' will become a discrepancy in the coupling whenever a choice of the form $c_1 2$ is made for D_1, which happens with probability $\frac{d'}{d+d'+s}$.

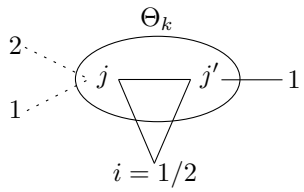

Figure 3.8. Case 3 (repeat of Figure 3.3). Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 (or 2). Let this site be labeled j'. The other site in Θ_k is labeled j and it is not adjacent to any site, other than i, coloured 1 or 2.

Hence

$$\rho^k_{i,j} \leq \frac{d}{d+d'+s} \text{ and } \rho^k_{i,j'} \leq \frac{d'}{d+d'+s}.$$

Starting with $\rho^k_{i,j}$

$$\rho^k_{i,j} \leq \frac{d}{d+d'+s} \leq \frac{d}{d+dd'} \leq \frac{1}{d'+1} \leq \frac{1}{q-\Delta+1}$$

using the lower bounds of s and d'. Similarly using the lower bounds of s and d

$$\rho^k_{i,j'} \leq \frac{d'}{d+d'+s} \leq \frac{d'}{d+dd'} \leq \frac{1}{d} \leq \frac{1}{q-\Delta}$$

which implies the statement of the lemma. □

Details of case 3. (Repeated in Figure 3.8.) We construct the coupling $\Psi_k(x,y)$ of D_1 and D_2 using the following two step process. Let Ψ_j be a coupling of $D_1(j')$ and $D_2(j')$ which greedily maximises the probability of assigning the same colour to site j' in each distribution. Then for each pair of colours (c,c') drawn from Ψ_j we complete $\Psi_k(x,y)$ by letting it be the coupling, greedily minimising Hamming distance, of the conditional distributions $D_1 \mid j' = c$ and $D_2 \mid j' = c'$.

Lemma 42. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i,j) \in E$ and $(i,j') \in E$. Suppose that 2 is valid for j in D_1, 1 is valid for j in D_2 and 1 is not valid for j' in D_2. Then*

$$\rho^k_{i,j'} \leq \frac{1}{q-\Delta+1} \text{ and } \rho^k_{i,j} \leq \frac{1}{q-\Delta}.$$

Proof. This is case 3 of the coupling. Note from Lemma 38 that $d_{1,j=2} = d_{2,j=1}$ and let $s = \sum_c d_{2,j=c} - d_{2,j=1} = \sum_{c\neq 1} d_{2,j=c}$ denote the number of valid choices for D_2 other than choices of the form $1c_2$. The number of valid choices for D_1 is then $d_1 = s + d_{1,j=2} + d_{1,j'=2}$.

Since 1 is not valid for j' in D_2 at least one site other than i on the boundary of Θ_k must be coloured 1 in D_1 and in particular this site is adjacent to j' (we say that some site v on the boundary of Θ_k is coloured c in D_1 if there exists a configuration with positive measure in D_1 in which site v is coloured c). As there are no restrictions on the neighbourhood of j each neighbour of j may be assigned a different colour in D_1. Hence we get the following lower bounds on $d_{1,j=2}$ and $d_{1,j'=2}$

$$q - \Delta + 1 \leq d_{1,j=2} \text{ and } q - \Delta \leq d_{1,j'=2}. \tag{3.5}$$

Next we need to establish a lower bound on s. Let J be the set of colours, excluding 1, that are valid for j in D_2 with the intention that $s = \sum_{c \in J} d_{2,j=c}$. Now observe that there are exactly $d_{1,j'=2}$ colours $c \in J$ for which $d_{2,j=c} > 0$ and hence

$$s = \sum_{c \in J} d_{2,j=c} \geq d_{1,j'=2} \min_{c' \in J} \{d_{2,j=c'}\}.$$

We then use Lemma 40 (i), since 1 is not valid for j' in D_2, to obtain the bound $d_{2,j=1} - 1 \leq d_{2,j=c'}$ for $c' \in J$ which gives the following lower bound on s

$$s \geq d_{1,j'=2} \min_{c' \in J} \{d_{2,j=c'}\} \geq d_{1,j'=2} \left(d_{2,j=1} - 1\right) = d_{1,j'=2} \left(d_{1,j=2} - 1\right) \tag{3.6}$$

since $d_{2,j=1} = d_{1,j=2}$ by Lemma 38 as we have previously noted.

We are now ready to bound the influence of i on j and j'. We consider $\rho^k_{i,j'}$ first. Suppose that a choice of the form $c_1 c_2$ is valid for D_2, in which case $c_1 \neq 2$ and $c_2 \notin \{1, 2\}$ by the conditions of case 3 of the coupling. Firstly if $c_1 \neq 1$ then $c_1 c_2$ is also valid for D_1 by Lemma 36. If $c_1 = 1$ then the choice $2c_2$ is valid for D_1 by Lemma 38 and hence $d_1 \geq d_2$. Note in particular that if a choice $c_1 c_2$ where $c_2 \neq 2$ is valid for D_1 then it is also valid for D_2. Therefore, a different colour will only be assigned to site j' in each distribution if j' is coloured 2 in D_1 during the first step of the coupling since the Hamming distance at site j' is minimised greedily. There are $d_{1,j'=2}$ colourings assigning 2 to j' in D_1 and hence

$$\rho^k_{i,j'} \leq \frac{d_{1,j'=2}}{d_{1,j=2} + d_{1,j'=2} + s} \leq \frac{d_{1,j'=2}}{d_{1,j=2}\left(1 + d_{1,j'=2}\right)} < \frac{1}{d_{1,j=2}} \leq \frac{1}{q - \Delta + 1}$$

where the second inequality uses the lower bound on s from (3.6) and the final inequality uses the lower bound on $d_{1,j=2}$ from (3.5).

Now consider $\rho^k_{i,j}$. Suppose that (c'_1, c'_2) is the pair of colours drawn for site j' in the first step of the coupling. The second step of $\Psi_k(x, y)$ then couples

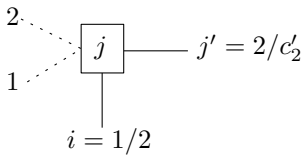

Figure 3.9. The pair of configurations after the colour of site j' has been assigned during the first step of the coupling.

the conditional distributions $D_1 \mid j' = c'_1$ and $D_2 \mid j' = c'_2$ greedily to minimise Hamming distance. First suppose that $c'_1 \neq c'_2$. It was pointed out in the analysis above that if $c'_1 \neq c'_2$ then $c'_1 = 2$ and the resulting configuration is shown in Figure 3.9. We make the following observations about the resulting conditional distributions $D_1 \mid j' = 2$ and $D_2 \mid j' = c'_2$.

- The colour 2 is *not* valid for j in either $D_1 \mid j' = 2$ or $D_2 \mid j' = c'_2$.
- The colour 1 is *not* valid for j in distribution $D_1 \mid j' = 2$ but *could* be valid for j in distribution $D_2 \mid j' = c'_2$.
- The colour c'_2 *could* be valid for j in distribution $D_1 \mid j' = 2$ but is *not* valid for j in distribution $D_2 \mid j' = c'_2$.
- For each $c \in C \backslash \{1, 2, c'_2\}$ the colour c is valid for j in distribution $D_1 \mid j' = 2$ *if and only if* c is valid for j in distribution $D_2 \mid j' = c'_2$.

These observations show that this case is a single-site disagreement sub problem. Furthermore there must be at least $(q-3)-(\Delta-2) = q-\Delta-1$ colours that are valid for j in both conditional distributions since j has at most $\Delta - 2$ neighbours other than i and j'. Finally, there is at most one colour which is valid for j in one distribution but not in the other and since the coupling greedily maximises Hamming distance this implies

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x'_{j'} \neq y'_{j'}) \leq \frac{1}{q-\Delta}.$$

Now suppose that the same colour c, say, is drawn for site j' in both distributions during the first step of the coupling. Then the only site adjacent to i that is coloured differently in the conditional distributions $D_1 \mid j' = c$ and $D_2 \mid j' = c$ is site i, so using a similar reasoning to above we find

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x'_{j'} = y'_{j'}) \leq \frac{1}{q-\Delta}$$

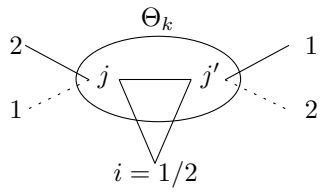

Figure 3.10. Case 4 (repeat of Figure 3.4). Both sites in Θ_k are adjacent to i. One of the sites in Θ_k is adjacent to at least one site, other than i, coloured 1 and no sites that are coloured 2. Let this site be labeled j'. The other site in Θ_k, labeled j, is adjacent to at least one site other than i coloured 2 and no sites coloured 1.

and thus

$$\begin{aligned}
\rho_{i,j}^k &= \max_{(x,y)\in S_i} \left\{ \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j) \right\} \\
&= \max_{(x,y)\in S_i} \{ \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x'_{j'} \neq y'_{j'}) \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'}) \\
&\quad + \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x'_{j'} = y'_{j'}) \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} = y'_{j'}) \} \\
&\leq \max_{(x,y)\in S_i} \left\{ \frac{1}{q-\Delta} \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'}) + \frac{1}{q-\Delta} \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} = y'_{j'}) \right\} \\
&= \frac{1}{q-\Delta} \max_{(x,y)\in S_i} \left\{ \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'}) + \Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} = y'_{j'}) \right\} \\
&= \frac{1}{q-\Delta}
\end{aligned}$$

which completes the proof. □

Details of case 4. (Repeated in Figure 3.10.) We assume without loss of generality that $d_1 \geq d_2$ and construct the coupling $\Psi_k(x,y)$ of D_1 and D_2 as follows. For each valid choice of the form c_1c_2 for D_1 where $c_1 \neq 1$ and $c_2 \neq 2$ Lemma 36 guarantees that c_1c_2 is also valid for D_2 so we construct $\Psi_k(x,y)$ such that

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = y' = c_1c_2) = \frac{1}{d_1}.$$

This leaves the set $Z_1 = \{c_1 2 \mid c_1 2 \text{ valid in } D_1\}$ of valid choices for D_1 and $Z_2 = \{1c_2 \mid 1c_2 \text{ valid in } D_2\} \subseteq D_2$ for D_2. Observe that $z_1 \geq z_2$ where z_1 and z_2 denote the size of Z_1 and Z_2 respectively. Let $Z_1(t)$ denote the t-th element of Z_1 and similarly for Z_2. Then for $1 \leq t \leq z_2$ let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = Z_1(t), y' = Z_2(t)) = \frac{1}{d_1}$$

and for each pair $z_2 + 1 \leq t \leq z_1$ and $h \in D_2$ let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = Z_1(t), y' = h) = \frac{1}{d_1 d_2}.$$

It is straightforward to verify that each valid colouring has the correct weight in $\Psi_k(x, y)$ so this completes the coupling.

Lemma 43. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i, j) \in E$ and $(i, j') \in E$. If 1 is valid for j in D_2, 1 is not valid for j' in D_2, 2 is valid for j' in D_1, and 2 is not valid for j in D_1 then*

$$\rho^k_{i,j} \leq \rho^k_{i,j'} \leq \frac{1}{q - \Delta}.$$

Proof. This is case 4 of the coupling. Let $s = \sum_c d_{2,j=c} - d_{2,j=1}$ be the number of valid choices for D_2 other than choices of the form $1c_2$. Observe that $d_1 = s + d_{1,j'=2}$ and note that $d_{1,j'=2} \geq d_{2,j=1}$ since we have assumed $d_1 \geq d_2$ in the construction of the coupling. At least one neighbour of j', other than i, on the boundary of Θ_k is coloured 1 in D_1 and we get the following lower-bound on $d_{2,j=1}$ since all other neighbours of j' may be assigned a different colour

$$q - \Delta + 1 \leq d_{2,j=1}.$$

We obtain a lower bound on s using an argument similar to the one in the proof of Lemma 42. Let J be the set of colours, excluding 1, that are valid for j in D_2 with the intention that $s = \sum_{c\in J} d_{2,j=c}$. Now observe that there are exactly $d_{1,j'=2}$ colours $c \in J$ for which $d_{2,j=c} > 0$ and hence

$$s = \sum_{c\in J} d_{2,j=c} \geq d_{1,j'=2} \min_{c'\in J}\{d_{2,j=c'}\}.$$

We then use Lemma 40 (i), since 1 is not valid for j' in D_2, to obtain the bound $d_{2,j=1} - 1 \leq d_{2,j=c'}$ for $c' \in J$ which gives the following lower bound on s

$$s \geq d_{1,j'=2} \min_{c'\in J}\{d_{2,j=c'}\} \geq d_{1,j'=2}\,(d_{2,j=1} - 1)\,.$$

We now go on to bound the influence of site i on sites j and j'. Since 2 is not valid for j in D_1 the first $d_{2,j=1}$ choices of the form $c_1 2$ for D_1 are matched with some choice of the form $1c_1$ for D_2 with probability $1/d_1$ resulting in a different colour being assigned to both sites j and j' in each distribution. Each of the $d_{1,j'=2} - d_{2,j=1}$ remaining valid choices for D_1 is matched with each valid choice

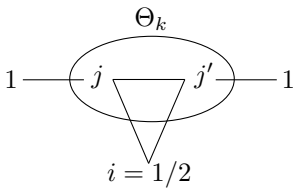

Figure 3.11. Case 5 (repeat of Figure 3.5). Both sites in Θ_k are adjacent to i and at least one site, other than i coloured 1 (or 2). The labeling of the sites in Θ_k is arbitrary.

for D_2 with probability $\frac{1}{d_1 d_2}$ resulting in a disagreement at j' (since 2 is not valid for j' in D_2) and potentially also at j so $\rho^k_{i,j} \leq \rho^k_{i,j'}$. Hence the probability of making a choice of the form $c_1 2$ for D_1

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} = 2) = \frac{d_{1,j'=2}}{d_1}$$

is an upper bound on the disagreement probabilities at both sites j and j'. Using the lower bounds on s and $d_{2,j=1}$ we have

$$\rho^k_{i,j} \leq \rho^k_{i,j'} \leq \frac{d_{1,j'=2}}{d_1} = \frac{d_{1,j'=2}}{d_{1,j'=2} + s} \leq \frac{d_{1,j'=2}}{d_{1,j'=2} + (d_{2,j=1} - 1)d_{1,j'=2}} \leq \frac{1}{q - \Delta}$$

which completes the proof. □

Details of case 5. (Repeated in Figure 3.11.) First observe that 1 is not valid for neither j nor j' so $d_1 = d_2 + d_{1,j=2} + d_{1,j'=2} \geq d_2$ by Lemma 36, since any choice valid for D_2 does not assign colour 2 to any site in Θ_k. Let Z_1 and Z_2 be the sets of colourings valid for D_1 and D_2 respectively. We define the following mutually exclusive subsets of Z_1. $Z_j = \{2c_2 \mid 2c_2 \in Z_1\}$, $Z_{j'} = \{c_1 2 \mid c_1 2 \in Z_1\}$ and $Z = Z_1 \setminus (Z_j \cup Z_{j'}) = Z_2$. By construction, the union of these three subsets is Z_1 and note that the size of Z_j is $d_{1,j=2}$, the size of $Z_{j'}$ is $d_{1,j'=2}$ and the size of Z is d_2.

First we consider choices from Z for D_1. For each choice $h \in Z$ we have $h \in Z_2$ by construction of Z and so we use the identity coupling and let

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x' = y' = h) = \frac{1}{d_1}.$$

We let the remainder of the coupling minimise Hamming distance. First consider the choices for D_1 in Z_j. We construct $\Psi_k(x, y)$ such that it minimises Hamming distance and assigns probability $1/d_1$ to each choice for D_1 in Z_j whilst ensuring

that for each choice $g \in Z_2$ for D_2

$$\sum_{h \in Z_j} \Pr_{(x',y') \in \Psi_k(x,y)}(x' = h, y' = g) = \frac{d_{1,j=2}}{d_1 d_2}.$$

Similarly we assign probability $1/d_1$ to each choice for D_1 in $Z_{j'}$ whilst also requiring that for each choice $g \in Z_2$ for D_2

$$\sum_{h \in Z_{j'}} \Pr_{(x',y') \in \Psi_k(x,y)}(x' = h, y' = g) = \frac{d_{1,j'=2}}{d_1 d_2}.$$

To see that this ensures that the coupling is fair observe that each choice $h \in Z_1$ receives weight $1/d_1$ and each choice $g \in Z_2$ weight

$$\frac{1}{d_1} + \frac{d_{1,j=2}}{d_1 d_2} + \frac{d_{1,j'=2}}{d_1 d_2} = \frac{d_2 + d_{1,j=2} + d_{1,j'=2}}{d_1 d_2} = \frac{1}{d_2}$$

since $d_2 + d_{1,j=2} + d_{1,j'=2} = d_1$.

Remark. Note that a coupling satisfying these requirements always exists. We will not give the detailed construction of $\Psi_k(x, y)$ here, but in the subsequent proof we will consider three cases. In the first two cases *any* coupling minimising Hamming distance will be sufficient to establish the required bounds on the influence of i on j. In the final case we will need a detailed construction of the coupling and so will provide it together with the proof for ease of reference.

Lemma 44. *Let j and j' be the endpoints of an edge Θ_k and suppose that $(i, j) \in E$ and $(i, j') \in E$. If 1 is not valid for j in D_2 and 1 is not valid for j' in D_2 then*

$$\rho^k_{i,j} \leq \frac{1}{q - \Delta + 1} + \frac{1}{(q - \Delta + 1)^2} \text{ and } \rho^k_{i,j'} \leq \frac{1}{q - \Delta + 1} + \frac{1}{(q - \Delta + 1)^2}.$$

Proof. This is case 5 of the coupling. We consider three separate cases. Firstly suppose that 2 is not valid for either j or j' in D_1. Then the only valid choices for D_1 are of the form $c_1 c_2$ where $c_1, c_2 \in C \setminus \{1, 2\}$ and each such choice is also valid in D_2 as observed in the construction of the coupling. The same colouring is selected for each distribution and hence

$$\rho^k_{i,j} = 0 \text{ and } \rho^k_{i,j'} = 0.$$

Next suppose that exactly one site in Θ_k, j' say, is adjacent to some site coloured 2 in D_1. As in the previous case, each choice that is valid in both D_1

and D_2 is matched using the identity matching and does not cause a discrepancy at any site. However if a choice of the form $2c$ is made for D_1 then site j will be coloured differently in each colouring drawn from $\Psi_k(x, y)$ and the colour at site j' may also be different so $\rho^k_{i,j'} \leq \rho^k_{i,j}$. Since all choices of the form $c2$ are not valid for D_1, making a choice of the form $2c$ for D_1 is the only way to create a disagreement at any site in the coupling and so

$$\rho^k_{i,j'} \leq \rho^k_{i,j} \leq \frac{d_{1,j=2}}{d_1}$$

since $d_{1,j=2}$ is the number of valid choices for D_1 of the form $2c$. We need to establish a lower bound of d_1 and observe that, for c valid for j in D_1, $d_{1,j=2} - 1 \leq d_{1,j=c}$ by Lemma 40 (ii) since 2 is not valid for j' in D_1. Let v be the number of colours that are valid for site j in D_1. Then v is lower bounded by $q - \Delta + 2 \leq v$ since at least two of the sites (including i) adjacent to j on the boundary of Θ_k are coloured 1 in D_1. Also, since at least one site (other than j and i) adjacent to j' is coloured 1 and another is coloured 2 in D_1, we have $q - \Delta + 2 \leq d_{1,j=2}$. Using the lower bounds on v and $d_{1,j=c}$ we have, letting J denote the set of colours other than 2 that are valid for j in D_1,

$$\begin{aligned} d_1 = \sum_c d_{1,j=c} &= d_{1,j=2} + \sum_{c \in J} d_{1,j=c} \\ &\geq d_{1,j=2} + \sum_{c \in J} (d_{1,j=2} - 1) \\ &\geq (v-1)(d_{1,j=2} - 1) + d_{1,j=2} \\ &\geq (q - \Delta + 2) d_{1,j=2} - (q - \Delta + 1) \end{aligned}$$

and hence using the lower bound on $d_{1,j=2}$

$$\frac{1}{\rho^k_{i,j}} \geq \frac{(q - \Delta + 2) d_{1,j=2} - (q - \Delta + 1)}{d_{1,j=2}} \geq q - \Delta + 2 - \frac{q - \Delta + 1}{q - \Delta + 2} > q - \Delta + 1$$

which gives the bounds required by the statement of the lemma.

Finally consider the case when the colour 2 is valid for both j and j' in D_1. In this case we will provide details of the construction of $\Psi_k(x, y)$ when required. We begin by establishing some required bounds. Since 1 is not valid for j' in D_2 at least two neighbours of j' (including i) must be coloured 1 in D_1 and the same applies to the neighbourhood of j, so we get the following lower bounds on $d_{1,j=2}$ and $d_{1,j'=2}$

$$q - \Delta + 1 \leq d_{1,j=2} \text{ and } q - \Delta + 1 \leq d_{1,j'=2}. \tag{3.7}$$

We also require bounds on $d_{2,j=c}$ and $d_{2,j'=c}$ for other colours c. Suppose that the choice cc' is valid in D_2 then, since $c, c' \in C \setminus \{1,2\}$, cc' is also valid for D_1 by Lemma 36. Furthermore, the choice $c2$ is valid in D_1 (but not D_2) so $d_{1,j=c} - 1 = d_{2,j=c}$. Lemma 40 (ii) guarantees that $d_{1,j=2} \leq d_{1,j=c} \leq d_{1,j=2} + 1$ so

$$d_{1,j=2} - 1 \leq d_{2,j=c} \leq d_{1,j=2} \tag{3.8}$$

for any c valid for j in D_1. A symmetric argument gives

$$d_{1,j'=2} - 1 \leq d_{2,j'=c} \leq d_{1,j'=c} \tag{3.9}$$

for any colour c valid for j' in D_2. Observe that exactly $d_{1,j'=2}$ colours must be valid for site j in D_2 so using the stated bounds on $d_{2,j=c}$ we have the following bounds on d_2

$$d_{1,j'=2}(d_{1,j=2} - 1) \leq d_2 \leq d_{1,j'=2}d_{1,j=2}. \tag{3.10}$$

We bound the probability of disagreements at sites j and j' from choices made for D_1. From the coupling we again note that if a choice c_1c_2 where $c_1 \neq 2$ and $c_2 \neq 2$ is made for D_1 then there will be no disagreements at any site in Θ_k.

Consider making a valid choice of the form $2c$ for D_1. Firstly, such a choice for D_1 will cause site j to be coloured differently in any pair of colourings drawn from the coupling since 2 is not valid for j in D_2. We construct $\Psi_k(x, y)$ such that the choice $2c$ for D_1 is matched with a choice of the form $c'c$ for D_2 as long as such a choice that has not exceeded it aggregated probability exists. Let J denote the set of choices of the form $c'c$ that are valid for D_2 and note that the size of J is $d_{2,j'=c}$. The total aggregated weight of all choices of the form $c'c$ for D_2 is

$$\sum_{g \in J} \sum_{h \in Z_j} \Pr_{(x',y') \in \Psi_k(x,y)}(x' = h, y' = g) = \sum_{g \in J} \frac{d_{1,j=2}}{d_1 d_2} = \frac{d_{2,j'=c}d_{1,j=2}}{d_1 d_2}$$

so as long as

$$\frac{1}{d_1} \leq \frac{d_{2,j'=c}d_{1,j=2}}{d_1 d_2}$$

there is enough probability available in Z_2 to match all the weight of the choice $2c$ for D_1 with a choice of the form $c'c$ for D_2 and hence assigning the same colour, c, to site j' in any pair of colourings drawn from the coupling. If there is not enough unassigned weight available in Z_2 then the coupling will match as much probability as possible, $\frac{d_{2,j'=c}d_{1,j=2}}{d_1 d_2}$, with choices of the form $c'c$ for Z_2 but the

remaining probability will be matched with choices not assigning colour c to site j' in Z_2. Hence we obtain the following probabilities conditioned on making a choice of the form $2c$ for D_1.

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x' = 2c) = 1$$

and

$$\begin{aligned}\Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'} \mid x' = 2c) &\leq \max\left(0, 1 - \frac{d_{2,j'=c}d_{1,j=2}}{d_2}\right)\\ &\leq \max\left(0, 1 - \frac{(d_{1,j'=2}-1)d_{1,j=2}}{d_{1,j=2}d_{1,j'=2}}\right)\\ &\leq \frac{1}{d_{1,j'=2}}\end{aligned}$$

using the bounds on d_2 and $d_{1,j'=c}$ from (3.10) and (3.9). Lastly observe that there are $d_{1,j=2}$ valid choices for D_1 of the form $2c$ so

$$\sum_c \Pr_{(x',y')\in\Psi_k(x,y)}(x' = 2c) = \frac{d_{1,j=2}}{d_1} = \frac{d_{1,j=2}}{d_{1,j=2} + d_{1,j'=2} + d_2}.$$

The case when making a choice of the form $c2$ for D_1 is symmetric to the case just considered and yields the following conditional probabilities

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x' = c2) \leq \frac{1}{d_{1,j=2}}$$

$$\Pr_{(x',y')\in\Psi_k(x,y)}(x'_{j'} \neq y'_{j'} \mid x' = c2) = 1$$

and

$$\sum_c \Pr_{(x',y')\in\Psi_k(x,y)}(x' = c2) = \frac{d_{1,j'=2}}{d_{1,j=2} + d_{1,j'=2} + d_2}.$$

Using the derived bounds on the conditional probabilities we find

$$\begin{aligned}
\rho_{i,j}^k &= \max_{(x,y)\in S_i} \left\{ \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ \sum_c \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x' = 2c) \Pr_{(x',y')\in\Psi_k(x,y)}(x' = 2c) \right. \\
&\quad \left. + \sum_c \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j \mid x' = c2) \Pr_{(x',y')\in\Psi_k(x,y)}(x' = c2) \right\} \\
&\leq \max_{(x,y)\in S_i} \left\{ \sum_c \left[\Pr_{(x',y')\in\Psi_k(x,y)}(x' = 2c) + \Pr_{(x',y')\in\Psi_k(x,y)}(x' = c2) \frac{1}{d_{1,j=2}} \right] \right\} \\
&\leq \max_{(x,y)\in S_i} \left\{ \frac{d_{1,j=2}}{d_{1,j=2} + d_{1,j'=2} + d_2} + \frac{d_{1,j'=2}}{d_{1,j=2}(d_{1,j=2} + d_{1,j'=2} + d_2)} \right\}.
\end{aligned}$$

Now using the lower bound on d_2 from (3.10) we have

$$\begin{aligned}
\rho_{i,j}^k &\leq \max_{(x,y)\in S_i} \left\{ \frac{d_{1,j=2}}{d_{1,j=2}(1 + d_{1,j'=2})} + \frac{d_{1,j'=2}}{(d_{1,j=2})^2(1 + d_{1,j'=2})} \right\} \\
&< \max_{(x,y)\in S_i} \left\{ \frac{1}{1 + d_{1,j'=2}} + \frac{1}{(d_{1,j=2})^2} \right\} \\
&\leq \frac{1}{q - \Delta + 2} + \frac{1}{(q - \Delta + 1)^2}
\end{aligned}$$

from the lower bounds on $d_{1,j=2}$ and $d_{1,j'=2}$ from (3.7). By symmetry we also have

$$\rho_{i,j'}^k \leq \frac{1}{q - \Delta + 2} + \frac{1}{(q - \Delta + 1)^2}$$

which completes the proof. □

This completes the cases of the coupling and we combine the obtained bounds on $\rho_{i,j}^k$ and $\rho_{i,j'}^k$ in the following corollary of Lemmas 41, 42, 43 and 44 which we use in establishing the mixing time of $\mathcal{M}_{\text{edge}}$.

Corollary 45. *Let j and j' be the endpoints of an edge Θ_k. If $(i,j) \in E$ and $(i,j') \in E$ then*

$$\rho_{i,j}^k \leq \frac{1}{q - \Delta} + \frac{1}{(q - \Delta)^2} \text{ and } \rho_{i,j'}^k \leq \frac{1}{q - \Delta} + \frac{1}{(q - \Delta)^2}.$$

Remark. Note that the bound in Corollary 45 is never tight. This bound could be improved, however this would only allow us to beat the 2Δ bound for special graphs since the bounds in Lemma 37 are tight.

We are now ready to present a proof of Theorem 16.

Theorem 16. *Let G be a graph with maximum vertex-degree Δ. Consider the systematic scan Markov chain $\mathcal{M}_{\text{edge}}$ on Ω. If $q \geq 2\Delta$ then the mixing time of $\mathcal{M}_{\text{edge}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{edge}}, \varepsilon) \leq \Delta^2 \log(n\varepsilon^{-1})$$

scans. If $m = O(n)$ then this corresponds to $O(n \log n)$ block updates.

Proof. Let j and j' be the endpoints of an edge represented by a (worst case) block Θ_k. Let $\alpha_j = \sum_i \rho^k_{i,j}$ be the influence on site j and $\alpha_{j'} = \sum_i \rho^k_{i,j'}$ the influence on j'. Then $\alpha = \max(\alpha_j, \alpha_{j'})$. Suppose that Θ_k is adjacent to t triangles, that is there are t sites $i_1, \ldots, i_t$ such that $(i, j) \in E$ and $(i, j') \in E$ for each $i \in \{i_1, \ldots, i_t\}$. Note that $0 \leq t \leq \Delta - 1$. There are at most $\Delta - 1 - t$ sites adjacent to j that are not adjacent to j' and at most $\Delta - 1 - t$ sites adjacent to j' that are not adjacent to j. From Lemma 37 a site adjacent only to j will emit an influence of at most $\frac{1}{q-\Delta}$ on site j and Lemma 37 also guarantees that a site only adjacent to j' can emit an influence at most $\frac{1}{(q-\Delta)^2}$ on site j. Corollary 45 says that a site adjacent to both j and j' can emit an influence of at most $\frac{1}{q-\Delta} + \frac{1}{(q-\Delta)^2}$ on site j and hence

$$\begin{aligned}\alpha_j &\leq t\left(\frac{1}{q-\Delta} + \frac{1}{(q-\Delta)^2}\right) + (\Delta - 1 - t)\left(\frac{1}{q-\Delta}\right) + (\Delta - 1 - t)\left(\frac{1}{(q-\Delta)^2}\right) \\ &= \frac{\Delta - 1}{q-\Delta} + \frac{\Delta - 1}{(q-\Delta)^2}\end{aligned}$$

and similarly by considering the influence on site j' we find that

$$\alpha_{j'} \leq \frac{\Delta - 1}{q-\Delta} + \frac{\Delta - 1}{(q-\Delta)^2}.$$

Then using our assumption that $q \geq 2\Delta$ we have

$$\alpha = \max(\alpha_j, \alpha_{j'}) \leq \frac{\Delta - 1}{q-\Delta} + \frac{\Delta - 1}{(q-\Delta)^2} \leq \frac{\Delta - 1}{\Delta} + \frac{\Delta - 1}{\Delta^2} = \frac{\Delta^2 - 1}{\Delta^2} = 1 - \frac{1}{\Delta^2}$$

and we obtain the stated bound on the mixing time by applying Theorem 14. □

3.4 Application: Colouring a Tree

In this section we study our two systematic scan Markov chains for sampling from the uniform distribution of proper q-colourings of a tree.

3.4.1 A Single-site Systematic Scan

We begin with the single-site chain. Recall the definition of the systematic scan Markov chain $\mathcal{M}_{\text{tree}}$ where Θ_k is the "block" containing only site k for each $k \in V$. $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k and the transition matrix of $\mathcal{M}_{\text{tree}}$ is $\prod_{k=1}^{n} P^{[k]}$. We will prove Theorem 18, namely that $\mathcal{M}_{\text{tree}}$ mixes in $O(\log n)$ scans whenever $q > \Delta + 2\sqrt{\Delta - 1}$. We will use Theorem 14 to bound the mixing time and assign a weight $w_i = \left(\frac{q-\Delta}{2(\Delta-1)}\right)^{d_i} = \omega^{d_i}$ to each site $i \in V$ where d_i is the distance (number of edges) from i to the root. As usual we extend the state space of the chains to Ω^+ in order to use Theorem 14 in the analysis and remind the reader that an upper bound on the mixing time of the extended chain is also an upper bound on the mixing time of the original chain by Lemma 8.

We define the coupling $\Psi_j(x, y)$ on pairs of colourings $(x, y) \in S_i$ by updating block Θ_j (i.e. site j) using a heat-bath move. Assume without loss of generality that $x_i = 1$ and $y_i = 2$ and let Z_1 be the set of colours that are valid for j when site i is coloured 1 and similarly Z_2 the set of colours valid for site j when i is coloured 2. We denote by z_1 and z_2 the sizes of Z_1 and Z_2 respectively. Firstly if $(i, j) \notin E$ then $Z_1 = Z_2$ and we use the identity coupling where the same colour is assigned to j in each copy.

Now suppose that i and j are adjacent in G. Without loss of generality we can assume that $z_1 \geq z_2$. Every colour $c \in Z_1 \cap Z_2$ is valid for j in both distributions so for each $c \in Z_1 \cap Z_2$ we let

$$\Pr_{(x',y') \in \Psi_j(x,y)}(x'_j = y'_j = c) = \frac{1}{z_1}.$$

If $Z_1 \neq Z_2$ then $Z_1 \setminus Z_2 = \{2\}$ since every other colour is either valid in both distributions or in none and since $z_1 \geq z_2$ there is at most one colour in the set $Z_2 \setminus Z_1$. Firstly if $Z_2 \setminus Z_1 = \{1\}$ then we let

$$\Pr_{(x',y') \in \Psi_j(x,y)}(x'_j = 2, y'_j = 1) = \frac{1}{z_1}$$

which completes the coupling since $z_1 = z_2$. Otherwise $Z_2 \setminus Z_1 = \emptyset$ and for each $c \in Z_2$ we let

$$\Pr_{(x',y') \in \Psi_j(x,y)}(x'_j = 2, y'_j = c) = \frac{1}{z_1 z_2}$$

which completes the coupling.

The following lemma upper bounds the probability of disagreement at site j

in the coupling.

Lemma 46. *Suppose* $(x, y) \in S_i$. *Then*

$$\rho_{i,j}^{j} \leq \begin{cases} \frac{1}{q-\Delta} & \text{if } (i,j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Proof. It is trivial to see that if i and j are not adjacent then j will not become a disagreement since the same colour is used in both copies. Now consider the coupling when i and j are adjacent. From the definition of the coupling the probability of assigning a different colour to site j in each copy is at most

$$\Pr_{(x',y') \in \Psi_j(x,y)}(x'_j \neq y'_j) \leq \frac{1}{z_1}.$$

This bound is only tight when $2 \in Z_1$ which means that no neighbours of j (other than i) can be assigned colour 2. Site j has at most $\Delta - 1$ neighbours other than i, each of which potentially being assigned a different colour so there are at least $z_1 \geq q - (\Delta - 1) - 1 = q - \Delta$ colours in Z_1 and the statement of the lemma follows. □

We now use Lemma 46 to prove Theorem 18.

Theorem 18. *Let* G *be a tree with maximum vertex degree* $\Delta \geq 3$ *and height* H. *Consider the systematic scan Markov chain* $\mathcal{M}_{\text{tree}}$ *on* Ω. *If* $q \geq \Delta + 2\sqrt{\Delta - 1 + \delta}$ *for* $\delta > 0$ *then the mixing time of* $\mathcal{M}_{\text{tree}}$ *is*

$$\text{Mix}(\mathcal{M}_{\text{tree}}, \varepsilon) \leq \max\left(\frac{2(\Delta - 1 + \delta)}{\delta}, 4\right)\left(H \log\left(\frac{q - \Delta}{2(\Delta - 1)}\right) + \log(n\varepsilon^{-1})\right)$$

scans of the tree. Since $\log n \leq H \leq n$, *this corresponds to* $O(nH)$ *updates.*

Proof. We consider the influence on every site in the tree. First consider the root of the tree r. The root (which has weight 1) has at most Δ neighbours each of which has weight $\frac{q-\Delta}{2(\Delta-1)}$. Thus, using Lemma 46, the influence on the root α_{root} is at most

$$\alpha_{\text{root}} = \sum_{i \in adj(r)} \rho_{i,r}^{r} \frac{w_i}{w_r} \leq \frac{\Delta}{q - \Delta} \frac{q - \Delta}{2(\Delta - 1)} = \frac{\Delta}{2(\Delta - 1)} \leq \frac{3}{4}$$

since $\Delta \geq 3$.

Then consider a leaf l which has distance d to the root. A leaf has exactly one neighbour, which has distance $d - 1$ to the root. Thus, using Lemma 46, the

influence on a leaf α_{leaf} is at most

$$\alpha_{\text{leaf}} = \sum_{i \in adj(l)} \rho_{i,l}^{l} \frac{w_i}{w_l} \leq \frac{1}{q-\Delta} \frac{\omega^{d-1}}{\omega^d} = \frac{1}{q-\Delta} \frac{2(\Delta-1)}{q-\Delta} < \frac{2(\Delta-1)}{4(\Delta-1)} = \frac{1}{2}$$

since $q > \Delta + 2\sqrt{\Delta - 1}$.

Finally consider the influence on a general site j in the tree with distance d to the root. Site j has one parent and at most $\Delta - 1$ downward neighbours. Thus, using the bounds from Lemma 46, the influence α_j on a general site is at most

$$\begin{aligned} \alpha_j &\leq \frac{1}{q-\Delta} \frac{\omega^{d-1}}{\omega^d} + \frac{\Delta-1}{q-\Delta} \frac{\omega^{d+1}}{\omega^d} \\ &= \frac{1}{q-\Delta} \frac{2(\Delta-1)}{q-\Delta} + \frac{\Delta-1}{q-\Delta} \frac{q-\Delta}{2(\Delta-1)} \leq \frac{1}{2} \left(\frac{\Delta-1}{\Delta-1+\delta} + 1 \right) \end{aligned}$$

since $q \geq \Delta + 2\sqrt{\Delta - 1 + \delta}$. Rewriting the fraction we find

$$\alpha_j \leq \frac{1}{2} \left(1 - \frac{\delta}{\Delta - 1 + \delta} + 1 \right) = 1 - \frac{\delta}{2(\Delta - 1 + \delta)}$$

and so

$$\alpha = \max(\alpha_{\text{root}}, \alpha_{\text{leaf}}, \alpha_j) \leq \max \left(1 - \frac{\delta}{2(\Delta - 1 + \delta)}, \frac{3}{4} \right).$$

Finally observe that $0 \leq d_i \leq H$ and so

$$\frac{\max_i w_i}{\min_i w_i} \leq \left(\frac{q-\Delta}{2(\Delta-1)} \right)^H$$

which, using Theorem 14, completes the proof. □

Remark. Note that when $\Delta > \frac{4}{\varepsilon^2}$ then $\Delta + 2\sqrt{\Delta - 1} < (1+\varepsilon)\Delta$ for $\varepsilon > 0$.

3.4.2 A Systematic Scan with Block Dynamics

We now go on to consider a systematic scan using block updates, in particular we will will present a proof of Theorem 20 which improves the least number of colours required for mixing of systematic scan on a tree for individual values of Δ. Recall the definition of the systematic scan $\mathcal{M}_{\text{BlockTree}}$ where the set of blocks Θ is defined as follows. Let the block Θ_k contain a site r along with all sites below r in the tree that are at most $h-1$ edges away from r. We call h the *height* of the blocks and h is defined for each Δ in Table 2.1 (repeated in Table 3.1). The set of blocks Θ covers the sites of the tree and we construct Θ such that no block

has height less than h. $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k and hence $P^{[k]}(x,\cdot)$ is the uniform distribution on the set of configurations that agree with x off Θ_k and where no edge incident to a site in Θ_k is monochromatic (see Example 10). The transition matrix of the Markov chain $\mathcal{M}_{\text{BlockTree}}$ is $\prod_{k=1}^{m} P^{[k]}$ where m is the number of blocks.

We will use standard terminology when discussing the structure of the tree. In particular will say that a site i is a *descendant* of a site j (or j is a *predecessor* of i) if j is on the simple path from the root of the tree to i. We will call a site j a *child* of a site i (or i is the *parent* of j) if i and j are adjacent and j is a descendant of i. Finally $N_k(j) = \{i \in \partial\Theta_k \mid i \text{ is a descendant of } j\}$ is the set of descendants of j on the boundary of Θ_k.

The following lemma will provide upper bounds on the probability of disagreement at any site in the block.

Lemma 47. *Let $(x,y) \in S_i$ and suppose that i is adjacent to exactly one site in a block Θ_k. Then there exists a coupling ψ of $D_1 = P^{[k]}(x,\cdot)$ and $D_2 = P^{[k]}(y,\cdot)$ in which*

$$\Pr_{(x',y')\in\psi}(x'_j \neq y'_j) \leq \frac{1}{(q-\Delta)^{d(i,j)}}$$

for all $j \in \Theta_k$ where $d(i,j)$ is the edge distance from i to j.

Proof. We construct a coupling ψ of D_1 and D_2 based on the recursive coupling defined in Goldberg et al. [33]. The following definitions are based on Figure 3.12. Let $R \subseteq V$ be a set of sites. Also let (X, X') be a pair of colourings of the sites on the boundary of R (recall that the boundary of R is the set of sites that are not included in R but are adjacent to some site in R) which use the same colour for every site, *except* for one site u which is coloured l in X and l' in X'. We then say that $A(R,(X,X'),u,(l,l'))$ is a boundary pair. For a boundary pair $A(R,(X,X'),u,(l,l'))$ we let $v \in R$ be the site in R that is adjacent to u. We think of v as the root of R and note that we may need to turn the original tree "upside down" in order to achieve this, however the meaning should be clear. We then label the children (in R) of v as $v_1,\ldots,v_d$ and let $T = \{R_1,\ldots,R_d\}$ be the set of d subtrees of R that do not contain site v, that is for $R_k \in T, 1 \leq k \leq d$ we define $R_k = \{j \in R \mid j = v_k \text{ or } j \text{ is a descendant of } v_k\}$. Finally let D and D' be the uniform distributions on colourings of R consistent with the boundary colourings X and X' respectively and let $D(v)$ (respectively $D'(v)$) be the distribution on the color at site v induced by D (respectively D'). Then Ψ_R is the recursive coupling of D and D' summarised as follows.

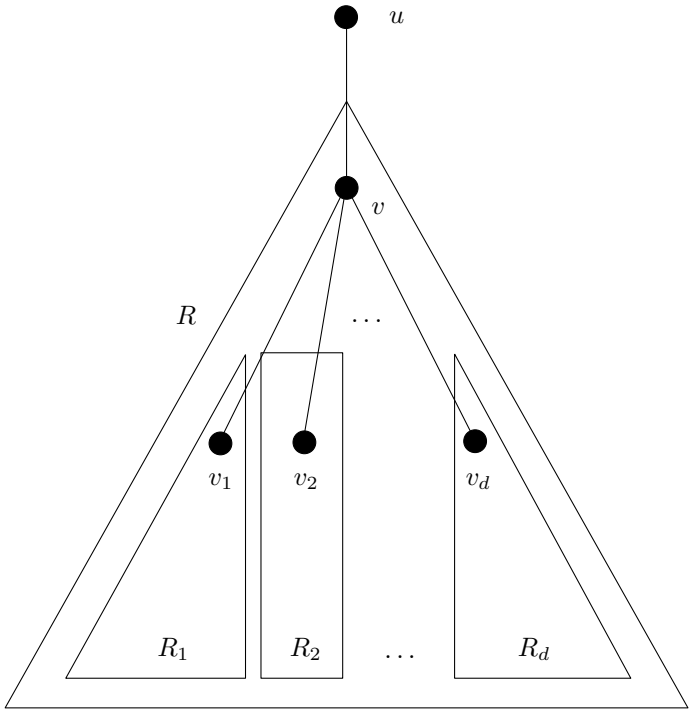

Figure 3.12. The region defined in a boundary pair and the construction of the subtrees.

1. If $l = l'$ then the distributions D and D' are the same and we use the identity coupling, in which the same colouring is used in both copies. Otherwise we couple $D(v)$ and $D'(v)$ greedily to maximise the probability of assigning the same colour to site v in both distributions. If R consists of just one site then this completes the coupling.

2. Suppose that the pair of colours (c, c') were drawn for v in the coupling from step 1. For each subtree $R' \in \{R_1, \dots R_d\}$ we have a well defined boundary pair $A(R', (X_{R'}, X'_{R'}), v, (c, c'))$ where $X_{R'}$ is the boundary colouring X restricted to the sites on the boundary of R'. For each pair of colours (c, c') and $R' \in T$ we recursively construct a coupling $\Psi_{R'}(c, c')$ of the distributions induced by the boundary pair $A(R', (X_{R'}, X'_{R'}), v, (c, c'))$.

Initially we let the boundary pair be $A(R = \Theta_k, (X = x, Y = y), u = i, (l = x_i, l' = y_i))$ and our coupling ψ of D_1 and D_2 is thus the recursive coupling Ψ_{Θ_k} constructed above.

We prove the statement of the lemma by induction on $d(i, j)$. The base case is $d(i, j) = 1$. Applying Lemma 13 from Goldberg et al. [33] we can upper bound the probability of $x'_j \neq y'_j$ where (x', y') is drawn from ψ by assigning the worst possible colouring to neighbours of j in Θ_k. Site j has at most $\Delta - 1$

neighbours (other than i) so there are at least $q - \Delta$ colours available for j in both distributions. There is also at most one colour which is valid for j in x but not in y (and vice versa) so

$$\Pr_{(x',y')\in\psi}(x'_j \neq y'_j) \leq \frac{1}{q-\Delta}.$$

Now let R' be the subtree of Θ_k containing site j and let v be the site in Θ_k adjacent to i. Assume that for $d(v,j) = d(i,j) - 1$

$$\Pr_{(x',y')\in\Psi_{R'}(c,c')}(x'_j \neq y'_j) \leq \frac{1}{(q-\Delta)^{d(v,j)}}.$$

Now for $(x,y) \in S_i$

$$\begin{aligned}
\Pr_{(x',y')\in\psi}(x'_j \neq y'_j) &= \Pr_{(x',y')\in\Psi_{\Theta_k}}(x'_j \neq y'_j) \\
&= \sum_{\substack{c,c' \\ c\neq c'}} \Pr_{(x',y')\in\Psi_R}(x'_v = c, y'_v = c')\Pr_{(x',y')\in\Psi_{R'}(c,c')}(x'_j \neq y'_j) \\
&\leq \frac{1}{(q-\Delta)^{d(i,j)-1}} \sum_{\substack{c,c' \\ c\neq c'}} \Pr_{(x',y')\in\Psi_R}(x'_v = c, y'_v = c') \\
&\leq \frac{1}{(q-\Delta)^{d(i,j)}}
\end{aligned}$$

where the first inequality is the inductive hypothesis and the last is a consequence of the base case. □

We will now use the coupling from Lemma 47 to define the coupling $\Psi_k(x,y)$ of the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ for $(x,y) \in S_i$. If $i \in \partial\Theta_k$ then it is adjacent to exactly one site in Θ_k and we use the coupling from Lemma 47. If $i \notin \partial\Theta_k$ then the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ are the same since we are using heat-bath updates and so we can use the identity coupling. We summarise the bounds on $\rho^k_{i,j}$ in the following corollary of Lemma 47.

Corollary 48. *Let $d(i,j)$ denote the number of edges between i and j. Then for $j \in \Theta_k$*

$$\rho^k_{i,j} \leq \begin{cases} \frac{1}{(q-\Delta)^{d(i,j)}} & \text{if } i \in \partial\Theta_k \\ 0 & \text{otherwise.} \end{cases}$$

We are now ready to present a proof of Theorem 20.

Theorem 20. *Let G be a tree with maximum vertex-degree Δ and height H. Consider the systematic scan Markov chain $\mathcal{M}_{\text{BlockTree}}$ on Ω. If $q \geq f(\Delta)$ where*

$f(\Delta)$ is specified in Table 2.1 *(repeated in Table 3.1 on page 42) for small Δ then the mixing time of $\mathcal{M}_{\text{BlockTree}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{BlockTree}}, \varepsilon) = O(H + \log(\varepsilon^{-1}))$$

scans of the tree. This corresponds to $O(nH)$ block updates by the construction of the set of blocks.

Proof. We will use Theorem 14 and assign a weight to each site i such that $w_i = \xi^{d_i}$ where d_i is the edge distance from i to the root and ξ is defined in Table 3.1 for each Δ. For a block Θ_k and $j \in \Theta_k$ we let

$$\alpha_{k,j} = \frac{\sum_i w_i \rho^k_{i,j}}{w_j}$$

denote the total weighted influence on site j when updating block Θ_k. For each block Θ_k and each site $j \in \Theta_k$ we will upper bound $\alpha_{k,j}$ and hence obtain an upper bound on $\alpha = \max_k \max_{j \in \Theta_k} \alpha_{k,j}$. Note from Corollary 48 that $\rho^k_{i,j} = 0$ when $i \in \Theta_k$ so we only need to bound $\rho^k_{i,j}$ for $i \in \partial\Theta_k$.

We first consider a block Θ_k that does not contain the root. The following labels refer to Figure 3.13 in which a solid line is an edge and a dotted line denotes the existence of a simple path between two sites. Let $p \in \partial\Theta_k$ be the predecessor of all sites in Θ_k and $d_r - 1$ be the distance from p to the root of the tree i.e., $w_p = \xi^{d_r - 1}$. The site $r \in \Theta_k$ is a child of p. Now consider a site $j \in \Theta_k$ which has distance d to r, hence $w_j = \xi^{d+d_r}$ and $d(j,p) = d+1$. From Corollary 48 it then follows that the weighted influence of p on j when updating Θ_k is at most

$$\rho^k_{p,j} \frac{w_p}{w_j} \leq \frac{1}{(q-\Delta)^{d(j,p)}} \frac{\xi^{d_r-1}}{\xi^{d_r+d}} = \frac{1}{(q-\Delta)^{d+1}} \frac{1}{\xi^{d+1}}.$$

Now consider some site $u \in N_k(j)$ which is on the boundary of Θ_k. Since $u \in N_k(j)$ it has weight $w_u = \xi^{d_r+h}$ and so $d(j,u) = h - d$. Hence Corollary 48 says that the weighted influence of u on j is at most

$$\rho^k_{u,j} \frac{w_u}{w_j} \leq \frac{1}{(q-\Delta)^{d(j,u)}} \frac{\xi^{d_r+h}}{\xi^{d_r+d}} = \frac{1}{(q-\Delta)^{h-d}} \xi^{h-d}.$$

Every site in Θ_k has at most $\Delta - 1$ children so the number of sites in $N_k(j)$ is at most $|N_k(j)| \leq (\Delta-1)^{h-d}$ and so, summing over all sites $u \in N_k(j)$, the total

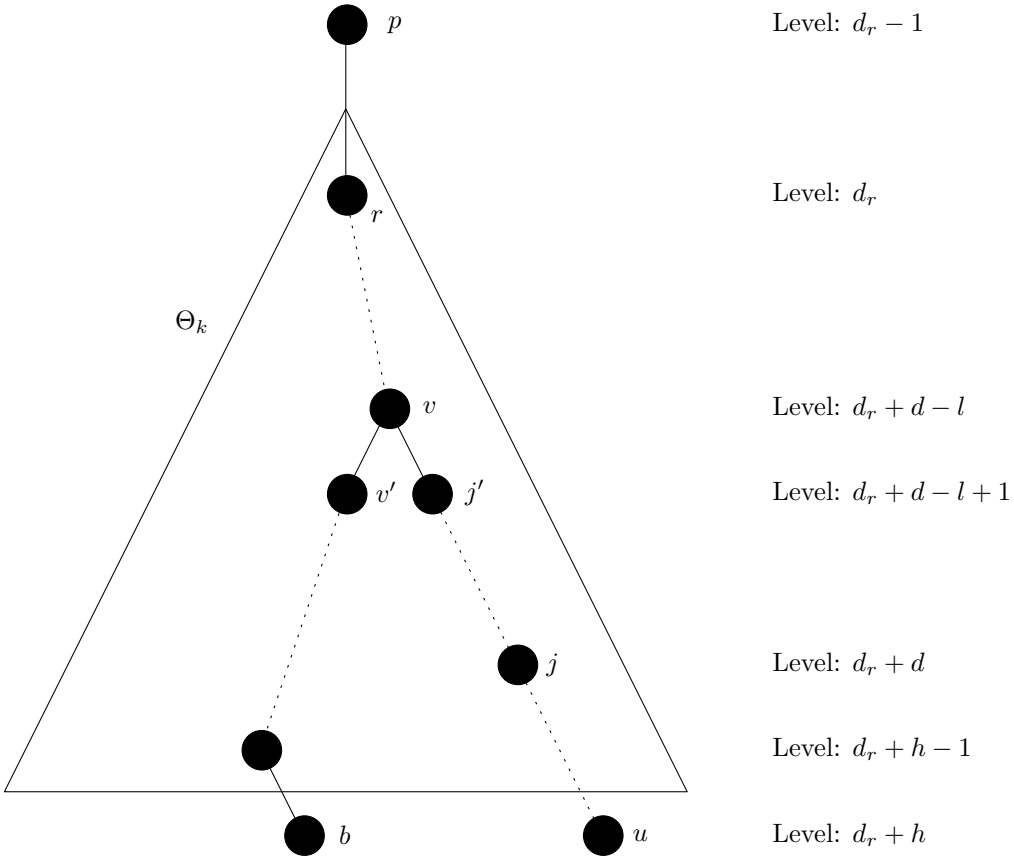

Figure 3.13. A block in the tree. A solid line indicates an edge and a dotted line the existence of a path.

weighted influence on j from sites in $N_k(j)$ when updating Θ_k is at most

$$\sum_{u\in N_k(j)} \rho^k_{u,j}\frac{w_u}{w_j} \leq \sum_{u\in N_k(j)} \frac{1}{(q-\Delta)^{h-d}}\xi^{h-d} \leq \frac{(\Delta-1)^{h-d}}{(q-\Delta)^{h-d}}\xi^{h-d}.$$

The influence on j from sites in $\partial\Theta_k \setminus (N_k(j)\cup\{p\})$ will now be considered. These are the sites on the boundary of Θ_k that are neither descendants or predecessors of j. For each site v between j and p, we will bound the influence on site j from sites $b \in N_k(v)$ that contain v on the simple path between b and j. We call this the influence on j via v. Referring to Figure 3.13 let $v \in \Theta_k$ be a predecessor of j such that $d(j,v) = l$ and observe that v is on level $d_r + d - l$ in the tree and also that $1 \leq l \leq d$ since v is between p and j in the tree. If v is not the parent of j (that is $l \neq 1$) then let j' be the child of v which is also a predecessor of j, that is j' is on the simple path from v to j. If $l = 1$ we let $j' = j$. Also let v' be any child of v other than j' and observe that v' and j' are both on level $d_r + d - l + 1$. Now let $b \in N_k(v')$ be a descendant of v' and note as before that $w_b = \xi^{d_r+h}$. The distance between b and v' is

$$d(v',b) = d_r + h - (d_r + d - l + 1) = h - d + l - 1$$

and so the number of descendants of v' is at most $|N_k(v')| \leq (\Delta-1)^{h-d+l-1}$ since each site has at most $\Delta - 1$ children. Site v has at most $\Delta - 2$ children other than j' so the number of sites on the boundary of Θ_k that are descendants of v but not j' is at most

$$|N_k(v) \setminus N_k(j')| \leq (\Delta-2)|N_k(v')| \leq (\Delta-2)(\Delta-1)^{h-d+l-1}.$$

Finally the only simple path from b to j goes via v and the number of edges on this path is

$$d(j,b) = d(j,v) + d(v,v') + d(v',b) = l + 1 + (h - d + l - 1) = h - d + 2l$$

so, using Corollary 48, the weighted influence of b on site j when updating block Θ_k is at most

$$\rho^k_{b,j}\frac{w_b}{w_j} \leq \frac{\xi^{d_r+h}}{\xi^{d_r+d}}\frac{1}{(q-\Delta)^{d(j,b)}} \leq \frac{\xi^{h-d}}{(q-\Delta)^{h-d+2l}}$$

and summing over all descendants of v (other than descendants of j') on the

boundary of Θ_k we find that the influence on j via site v is at most

$$\sum_{b\in N_k(v)\setminus N_k(j')} \rho^k_{b,j}\frac{w_b}{w_j} \le \sum_{b\in N_k(v)\setminus N_k(j')} \frac{\xi^{h-d}}{(q-\Delta)^{h-d+2l}} \le \xi^{h-d}\frac{(\Delta-2)(\Delta-1)^{h-d+l-1}}{(q-\Delta)^{h-d+2l}}. \tag{3.11}$$

Summing (3.11) over $1 \le l \le d$ gives an upper bound on the the total weighted influence of sites in $\partial\Theta_k \setminus (N_k(j)\cup\{p\})$ on site j when updating Θ_k

$$\sum_{b\in\partial\Theta_k\setminus(N_k(j)\cup\{p\})} \rho^k_{b,j}\frac{w_b}{w_j} \le \xi^{h-d}\sum_{l=1}^{d}\frac{(\Delta-2)(\Delta-1)^{h-d+l-1}}{(q-\Delta)^{h-d+2l}}$$

and adding the derived influences we find that the influence on site j (on level d_r+d) when updating Θ_k is at most

$$\begin{aligned}\alpha_{k,j} &= \frac{\rho^k_{p,j}w_p}{w_j} + \sum_{u\in N_k(j)}\frac{\rho^k_{u,j}w_u}{w_j} + \sum_{b\in\partial\Theta_k\setminus(N_k(j)\cup\{p\})}\frac{\rho^k_{b,j}w_b}{w_j}\\ &\le \frac{1}{(q-\Delta)^{d+1}}\frac{1}{\xi^{d+1}} + \frac{(\Delta-1)^{h-d}}{(q-\Delta)^{h-d}}\xi^{h-d} + \xi^{h-d}\sum_{l=1}^{d}\frac{(\Delta-2)(\Delta-1)^{h-d+l-1}}{(q-\Delta)^{h-d+2l}}.\end{aligned}$$

Now consider the block containing the root of the tree, r. Let this be block Θ_0 and note that $w_r = 1$. The only difference between Θ_0 and any other block is that r may have Δ children. There are at most $\Delta(\Delta-1)^{h-1}$ descendants of r in $\partial\Theta_0$, each of which has weight ξ^h so, using Corollary 48, the weighted influence on the root is at most

$$\alpha_{0,r} = \sum_{b\in N_0(r)}\rho^0_{b,r}\frac{w_b}{w_r} \le \frac{\Delta(\Delta-1)^{h-1}}{(q-\Delta)^h}\xi^h.$$

Now consider a site j on level $d \ne 0$ in block Θ_0. As in the general case considered above there is an influence of at most

$$\sum_{b\in N_0(j)}\frac{\rho^0_{b,j}w_b}{w_j} \le \frac{(\Delta-1)^{h-d}}{(q-\Delta)^{h-d}}\xi^{h-d}$$

on j from the sites in $N_0(j)$. Now consider the influence on site j from $\partial\Theta_0\setminus N_0(j)$. We first consider the influence on j via r, which is shown in Figure 3.14. Site r has at most $\Delta-1$ children other than the site j' which is the child of r that is on the path from r to j. Each child of r has at most $(\Delta-1)^{h-1}$ descendants in $\partial\Theta_0$ and each such descendant has distance $h+d$ to j. Hence, from Corollary 48, the

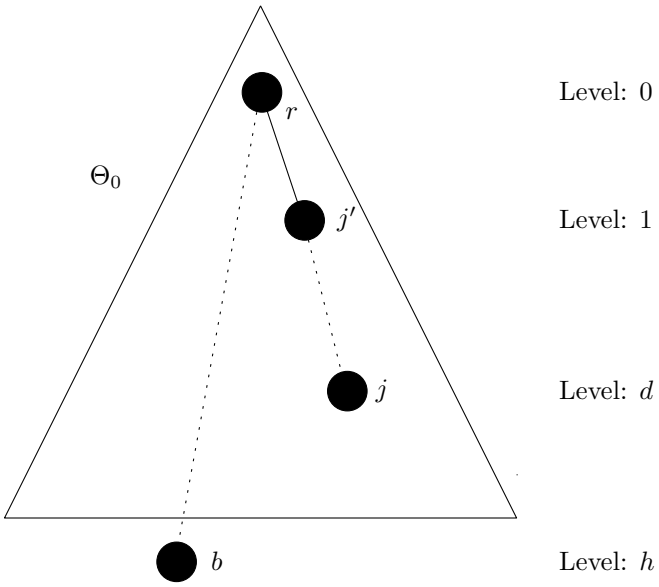

Figure 3.14. The influence on site j via the root. A line denotes an edge and a dotted line the existence of a simple path.

influence on j via the root is at most

$$\sum_{b\in N_0(r)\setminus N_0(j')} \frac{\rho^0_{b,j} w_b}{w_j} \leq \sum_{b\in N_0(r)\setminus N_0(j')} \frac{\xi^h}{\xi^d} \frac{1}{(q-\Delta)^{d(b,j)}} \leq \frac{(\Delta-1)^h}{(q-\Delta)^{h+d}} \xi^{h-d}.$$

Finally consider then influence on j from the remaining sites, which are in the set $R = \partial\Theta_0 \setminus (N_0(j) \cup (N_0(r) \setminus N_0(j')))$. Again consider a site $v \neq r \in \Theta_0$ where v is a predecessor of j and $d(j,v) = l$. In this case we have $1 \leq l \leq d-1$ since $l = d$ is the root which has already been considered. This is the same situation as arose in the general case considered above (see Figure 3.13) so (3.11) is an upper bound on the influence on j via v and so summing (3.11) over $1 \leq l \leq d-1$ and adding the other influences on j we obtain an upper bound on the total weighted influence on site j when updating block Θ_0

$$\begin{aligned}
\alpha_{0,j} &= \sum_{b\in N_0(j)} \frac{\rho^0_{b,j} w_b}{w_j} + \sum_{b\in N_0(r)\setminus N_0(j')} \frac{\rho^0_{b,j} w_b}{w_j} + \sum_{b\in R} \frac{\rho^0_{b,j} w_b}{w_j} \\
&\leq \frac{(\Delta-1)^{h-d}}{(q-\Delta)^{h-d}} \xi^{h-d} + \frac{(\Delta-1)^h}{(q-\Delta)^{h+d}} \xi^{h-d} + \xi^{h-d} \sum_{l=1}^{d-1} \frac{(\Delta-2)(\Delta-1)^{h-d+l-1}}{(q-\Delta)^{h-d+2l}}.
\end{aligned}$$

We require $\alpha < 1$ which we obtain by satisfying the system of inequalities

given by setting

$$\alpha_{k,j} < 1 \tag{3.12}$$

for all blocks Θ_k and sites $j \in \Theta_k$. In particular we need to find an assignment to ξ and h that satisfies (3.12) given Δ and q. Table 3.1 shows the least number of colours $f(\Delta)$ required for mixing for small Δ along with a weight, ξ, that satisfies the system of equations and the required height of the blocks, h. These values were verified by checking the resulting $2h$ inequalities for each Δ using Mathematica; the source of the program is available at `http://www.csc.liv.ac.uk/~kasper/tree_scan/`. The least number of colours required for mixing in the single-site setting is also included in the table for comparison.

Finally observe that $0 \leq d_i \leq H$ and so

$$\frac{\max_i w_i}{\min_i w_i} \leq \left(\frac{1}{\xi}\right)^H$$

which, by Theorem 14, yields a mixing time of

$$\begin{aligned} O(\log(n\xi^{-H}\varepsilon^{-1})) &= O(H\log\xi^{-1} + \log n + \log\varepsilon^{-1}) \\ &= O(H + \log\varepsilon^{-1}) \end{aligned}$$

since $\log n \leq H \leq n$. This completes the proof. □

3.5 A Comparison of Influence Parameters

We conclude this chapter with a discussion of our choice of influence parameter α denoting the maximum influence on any site in the graph. As we will be comparing the condition $\alpha < 1$ to the corresponding, but unweighted, conditions in Dyer et al. [18] and Weitz [55] we will let $w_i = 1$ for each site and omit the weights from now on. Recall our definitions of $\rho^k_{i,j}$ and α

$$\rho^k_{i,j} = \max_{(x,y)\in S_i} \{\Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j)\} \text{ and } \alpha = \max_k \max_{j\in\Theta_k} \sum_{i\in V} \rho^k_{i,j}$$

where $\Psi_k(x,y)$ is a coupling of the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$. We have previously stated that this is not the standard way to define the influence of i on j since the coupling is directly included in the definition of $\rho^k_{i,j}$. It is worth pointing out, however, that the corresponding definition in Weitz [55], which is also for block dynamics, also makes explicit use of the coupling. In the single-site

setting (Dyer et al. [18]) the influence of i on j, which we will denote $\hat{\rho}_{i,j}$, is defined by

$$\hat{\rho}_{i,j} = \max_{(x,y)\in S_i} \mathrm{d_{TV}}(\mu_j(x), \mu_j(y))$$

where $\mu_j(x)$ is the distribution on spins at site j induced by $P^{[j]}(x,\cdot)$. The corresponding condition is $\hat{\alpha} = \max_j \sum_{i\in V} \hat{\rho}_{i,j} < 1$. We will show (Lemma 49) that $\hat{\rho}_{i,j}$ is a special case of $\rho^j_{i,j}$ when $\Theta_j = \{j\}$ and $\Psi_j(x,y)$ is a coupling minimising the Hamming distance at site j. This will prove our claim that our condition $\alpha < 1$ is a generalisation of the single-site condition $\hat{\alpha} < 1$.

Before demonstrating this fact we will discuss the need to include the coupling in the definition of ρ in the block setting. Consider a pair of distinct sites $j \in \Theta_k$ and $j' \in \Theta_k$ and a pair of configurations $(x,y) \in S_i$. When updating block Θ_k the dynamics needs to draw a pair of new configurations (x',y') from the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ as previously specified. Hence the interaction between j and j' has to be according to these distributions and so it is not possible to consider the influence of i on j and the influence of i on j' separately. In the context of our definition of $\rho^k_{i,j}$ this means that the influence of i on j and the influence of i on j' have to be defined using the *same coupling*. This is to say that the coupling $\Psi_k(x,y)$ can only depend on the block Θ_k and the initial pair of configurations x and y, which in turn specify which site is labeled i. It is important to note that the coupling can not depend on j, since otherwise having a small influence on a site would not imply rapid mixing of systematic scan (or indeed random update). The reason why we need to make this distinction when working with block dynamics but not the single-site dynamics is that in the single-site setting $\hat{\rho}_{i,j}$ is the influence of site i on j when updating site j and hence whichever coupling is used must implicitly depend on j. Since the coupling can depend on j in the single-site case it is natural to use the "optimal" coupling, which minimises the probability of having a discrepancy at site j. By definition of total variation distance, the probability of having a discrepancy at site j under the optimal coupling is $\mathrm{d_{TV}}(\mu_j(x), \mu_j(y)) = \hat{\rho}_{i,j}$ (see e.g. Aldous [2]). We will now show that $\hat{\rho}_{i,j}$ is a special case of $\rho^j_{i,j}$ in the way described above.

Lemma 49. *Suppose that for each site $j \in V$ we have a block $\Theta_j = \{j\}$ and that $\Theta = \{\Theta_1, \ldots, \Theta_n\}$. Also suppose that for each pair $(x,y) \in S_i$ of configurations $\Psi_j(x,y)$ is a coupling of $P^{[j]}(x,\cdot)$ and $P^{[j]}(y,\cdot)$ in which, for each $c \in C$,*

$$\Pr_{(x',y')\in\Psi_j(x,y)}(x'_j = y'_j = c) = \min(\Pr_{\mu_j(x)}(c), \Pr_{\mu_j(y)}(c))$$

where $\Pr_{\mu_j(x)}(c)$ *is the probability of drawing colour* c *from distribution* $\mu_j(x)$. *Then* $\rho^j_{i,j} = \hat{\rho}_{i,j}$.

Proof. To see that the coupling $\Psi_j(x, y)$ always exists it is sufficient to observe that $\Pr_{x' \in P^{[j]}(x,\cdot)}(x'_j = c) = \Pr_{\mu_j(x)}(c)$ and similarly $\Pr_{y' \in P^{[j]}(y,\cdot)}(y'_j = c) = \Pr_{\mu_j(y)}(c)$ since j is the only site in Θ_j. The following sequence of equalities establish the proof.

$$\begin{aligned}
\rho^j_{i,j} &= \max_{(x,y)\in S_i} \left\{ \Pr_{(x',y')\in\Psi_j(x,y)}(x'_j \neq y'_j) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ 1 - \sum_{c\in C} (\Pr_{(x',y')\in\Psi_j(x,y)}(x'_j = y'_j = c)) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ 1 - \sum_{c\in C} \min(\Pr_{\mu_j(x)}(c), \Pr_{\mu_j(y)}(c)) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ \sum_{c\in C} \Pr_{\mu_j(x)}(c) - \min(\Pr_{\mu_j(x)}(c), \Pr_{\mu_j(y)}(c)) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ \sum_{c\in C^+} \Pr_{\mu_j(x)}(c) - \Pr_{\mu_j(y)}(c) \right\} \\
&= \max_{(x,y)\in S_i} \left\{ \frac{1}{2} \sum_{c\in C} |\Pr_{\mu_j(x)}(c) - \Pr_{\mu_j(y)}(c)| \right\} \\
&= \max_{(x,y)\in S_i} \mathrm{d_{TV}}(\mu_j(x), \mu_j(y)) \\
&= \hat{\rho}_{i,j}
\end{aligned}$$

where $C^+ = \{c \mid \Pr_{\mu_j(x)}(c) \geq \Pr_{\mu_j(y)}(c)\}$. □

We have previously pointed out that using influence parameters to bound the mixing time of Markov chains is a technique that has been used in recent times. In the context of systematic scan, Dyer et al. [18] have pointed out that the condition "the influence *on* a site is small" implies rapid mixing of systematic scan in the single-site setting. In particular they use the parameter $\alpha_{\mathrm{DGJ}} = \max_{j\in V} \sum_{i\in V} \hat{\rho}_{i,j}$ which denotes the influence on a site, and observe that if the condition $\alpha_{\mathrm{DGJ}} < 1$ is satisfied then any systematic scan Markov chain mixes in $O(\log n)$ scans for the given spin system. Our condition, namely $\alpha < 1$, is then a generalisation of the condition $\alpha_{\mathrm{DGJ}} < 1$ to block dynamics. It is straightforward to verify that if each block contains exactly one site and the coupling minimises the Hamming distance then $\alpha = \alpha_{\mathrm{DGJ}}$ by Lemma 49. Hence the single-site case is a special case of our condition.

Dyer et al. [18] also considered the parameter $\alpha'_{\mathrm{DGJ}} = \max_{i \in V} \sum_{j \in V} \hat{\rho}_{i,j}$ denoting the influence *of* a site. This parameter comes from Föllmer's [28] account of Dobrushin's proof presented by Simon [51]. The condition $\alpha'_{\mathrm{DGJ}} < 1$ is similar in nature to the condition used in path coupling and implies rapid mixing of a random update Markov chain. They go on to show that if $\alpha'_{\mathrm{DGJ}} < 1$ then it is possible to find a set of weights assigned to each site that ensures that $\alpha_{\mathrm{DGJ}} < 1$ (in a weighted setting similar to ours) and hence that systematic scan mixes rapidly. They call their approach *matrix balancing* since in the single-site case it is convenient to represent the influences that sites have on each other by an $n \times n$ matrix, which we call R, in which $R_{i,j} = \hat{\rho}_{i,j}$. The parameter α_{DGJ} then corresponds to the largest column sum of R and α'_{DGJ} is the largest row sum of R. This result has since been improved by Hayes [36] who showed that it is sufficient to bound the second largest eigenvalue (known as the operator norm) of R below one for the same conclusions to hold. This result has in turn been further generalised by Dyer et al. [19] who show that if one can bound any matrix norm below one then both the random update and systematic scan Markov chains are rapidly mixing.

We now return to our discussion of block dynamics and Weitz's conditions for rapid mixing. We can use the definition of $\rho^k_{i,j}$ to translate Weitz's conditions into notation that is easily comparable with our influence parameter α. Weitz's parameter α'_{W}, which represents the influence of a site, is defined as

$$\alpha'_{\mathrm{W}} = \max_{i \in V} \sum_{k=1}^{m} \sum_{j \in \Theta_k} \frac{\rho^k_{i,j}}{b(i)}$$

where $B(j)$ is the set of block indices that contain site j and $b(j)$ the size of this set. Weitz's parameter representing the influence on a site, which we denote by α_{W}, is defined as

$$\alpha_{\mathrm{W}} = \max_{j \in V} \sum_{k \in B(j)} \sum_{i \in V} \frac{\rho^k_{i,j}}{b(j)}.$$

Remark. Weitz's parameters are actually slightly more general than we have presented them here. In particular Weitz [55] states his conditions for general metrics whereas we have implicitly used Hamming distance. Using Hamming distance is also how the corresponding condition is defined in Dyer et al. [18] and Simon [51] for the single-site case.

Weitz [55] proves that each of the conditions $\alpha'_{\mathrm{W}} < 1$ and $\alpha_{\mathrm{W}} < 1$ imply spatial mixing (and hence that the Gibbs measure is unique which is what he is concerned with). For completeness we present proofs that these conditions

also imply rapid mixing of a random update Markov chain; these proofs of rapid mixing are based on a proof outline in Weitz [55]. Recall that, for any set of m blocks Θ, $\mathcal{M}_{\mathrm{RU}}$ is the random update Markov chain with transition matrix $(1/m)\sum_{k=1}^{m} P^{[k]}$.

Theorem 50 (Weitz [55])**.** *Suppose $\alpha'_{\mathrm{W}} = 1 - \gamma$ for some $0 < \gamma < 1$. Then the mixing time of $\mathcal{M}_{\mathrm{RU}}$ is*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{RU}}, \varepsilon) \leq \frac{m \log(n\varepsilon^{-1})}{\min_i b(i)\gamma}.$$

Proof. We prove the claim using path coupling. Consider a pair of configurations $(x, y) \in S_i$ that differ only on the colour at site i. Let (x', y') be the pair of configurations obtained from one step of the coupling starting at (x, y). We will prove that if $\alpha'_{\mathrm{W}} = 1 - \gamma$ then

$$\mathbf{E}\left[\mathrm{Ham}(x', y')\right] \leq 1 - \frac{\min_i b(i)\gamma}{m}$$

which implies the statement of the theorem by Corollary 9.

Denote by $A(i)$ the set of blocks indices that are adjacent to (but do not include) site i and a_i the size of this set, note that $A(i) \cap B(i) = \emptyset$. Suppose that a block Θ_k has been selected for update. There are three cases:

- $k \in A(i)$. In this case site i is unchanged and each site $j \in \Theta_k$ becomes a disagreement with probability at most $\rho^k_{i,j}$. This gives an expected Hamming distance (conditioned on selecting block Θ_k) of $1 + \sum_{j \in \Theta_k} \rho^k_{i,j}$ using linearity of expectation.

- $k \in B(i)$. In this case i is updated and remains a disagreement with probability at most $\rho^k_{i,i}$. Again each site $j \neq i \in \Theta_k$ becomes a disagreement with probability at most $\rho^k_{i,j}$. Using linearity of expectation this gives an expected Hamming distance of $\sum_{j \in \Theta_k} \rho^k_{i,j}$ after updating Θ_k.

- $k \notin A(i) \cup B(i)$. In this case i is unchanged so the (expected) Hamming distance after updating Θ_k is 1.

Each block is updated with probability $1/m$ so using the expectations from the

three cases we have

$$\begin{aligned}\mathbf{E}\left[\mathrm{Ham}(x',y')\right] &\leq \frac{1}{m}\sum_{k\in A(i)}\left(1+\sum_{j\in\Theta_k}\rho_{i,j}^k\right)+\frac{1}{m}\sum_{k\in B(i)}\left(\sum_{j\in\Theta_k}\rho_{i,j}^k\right)+\frac{1}{m}\sum_{k\notin A(i)\cup B(i)}1\\ &= \frac{1}{m}\left(a(i)+\sum_{k\in A(i)}\sum_{j\in\Theta_k}\rho_{i,j}^k+\sum_{k\in B(i)}\sum_{j\in\Theta_k}\rho_{i,j}^k+m-a(i)-b(i)\right)\\ &\leq \frac{1}{m}\left(m-b(i)+\sum_{k}\sum_{j\in\Theta_k}\rho_{i,j}^k\right).\end{aligned}$$

Now for all i note that $\sum_k\sum_{j\in\Theta_k}\rho_{i,j}^k \leq b(i)-b(i)\gamma$ (since $\alpha'_{\mathrm{W}}=1-\gamma$) so

$$\max_i \mathbf{E}\left[\mathrm{Ham}(x',y')\right] \leq \max_i \frac{1}{m}\left(m-b(i)\gamma\right) \leq 1-\frac{\min_i b(i)\gamma}{m}$$

which completes the proof. □

It is straightforward to obtain mixing of a random update Markov chain using path coupling and the condition $\alpha_{\mathrm{W}}<1$.

Theorem 51 (Weitz [55])**.** *Suppose that $\alpha_{\mathrm{W}}=1-\gamma$ for some $0<\gamma<1$. Then the mixing time of $\mathcal{M}_{\mathrm{RU}}$ is*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{RU}},\varepsilon) \leq \frac{m\log(n\varepsilon^{-1})}{\min_j b(j)\gamma}.$$

We prove Theorem 51 using (a block generalisation of) the method and notation from Section 7 of Dyer et al. [18]. First we use path coupling to specify a coupling $\psi_k(x,y)$ on block Θ_k of two configurations differing at arbitrarily many sites. Consider pairs of configurations (x,y) that agree on $\Theta_k\cup\partial\Theta_k$, that is $x=y$ on $\Theta_k\cup\partial\Theta_k$. In this case $\psi_k(x,y)$ is obtained by choosing the same configuration for Θ_k in both copies.

Now consider coupled chains X_t, Y_t and let the path coupling be given by choosing the same block Θ_k in both chains and coupling the choice of spins maximally as follows. Let $P_t=(X_t=Z_0,\ldots,Z_\ell=Y_t \text{ on } \Theta_k\cup\partial\Theta_k)$ be a sequence of configurations such that $\mathrm{Ham}(Z_{r-1},Z_r)=1$ for $1\leq r\leq\ell$. (To ease the notation we do not include as notation that both the states of the path as well as the path length ℓ depend on t.) Now observe that the couplings $\psi_k(Z_{r-1},Z_r)$ for $1\leq r\leq\ell$ are well defined in the sense that we have bounds on the resulting variation distance in the form of our definition of ρ. The coupling $\psi_k(Z_\ell,Y_t)$ is defined above.

We then construct the couping $\psi_k(X_t, Y_t)$ as follows. Initially choose a configuration W_0 from $P^{[k]}(X_t, \cdot)$ which is the equivalent of taking one step of the Markov chain starting at state X_t. Then inductively (for a step r) choose a configuration W_r from the coupling $\psi_k(Z_{r-1}, Z_r)$ conditioned on configuration W_{r-1}. The final step is choosing a configuration $W_{\ell+1}$ from the coupling $\psi_k(Z_\ell, Y_t)$ conditioned on W_ℓ. This is a standard path coupling construction.

Now, the initial states X_0, Y_0 have shortest path P_0 and the length of P_0 is $\mathrm{Ham}(X_0, Y_0)$. Consider the evolution of this path at time t to P_t with length $\ell \geq \mathrm{Ham}(X_t, Y_t)$. We do not optimise the path length at each time step, rather just allow the path to evolve. For any edge (Z_{r-1}, Z_r) in P_t say that it is in S_i if $(Z_{r-1}, Z_r) \in S_i$ and let ν_i^t be the number of edges of P_t in S_i. We prove the following lemma which is an analogue of Lemma 3.3 of Weitz [55].

Lemma 52.

$$\mathbf{E}\left[\nu_j^{t+1}\right] \leq \left(1 - \frac{b(j)}{m}\right) \mathbf{E}\left[\nu_j^t\right] + \sum_{k \in B(j)} \sum_i \frac{\rho_{i,j}^k}{m} \max_i \mathbf{E}\left[\nu_i^t\right]$$

Proof. Suppose Θ_k be the block selected for update. There are two cases. First suppose that $j \notin \Theta_k$. In this case site j does not get updated in either copy of the chain and so for every existing edge in S_j an edge in S_j persists and no new edges in S_j appear. There are $m - b(j)$ such blocks. Second suppose that $j \in \Theta_k$. In this case each edge in S_j persists with at most probability $\rho_{j,j}^k$ and for each edge in S_i (for $i \neq j$) a new edge in S_j is appears with probability at most $\rho_{i,j}^k$. Hence adding up the edges in S_j we have

$$\begin{aligned}
\mathbf{E}\left[\nu_j^{t+1}\right] &\leq \left(1 - \frac{b(j)}{m}\right) \mathbf{E}\left[\nu_j^t\right] + \sum_{k \in B(j)} \sum_i \frac{\rho_{i,j}^k}{m} \mathbf{E}\left[\nu_i^t\right] \\
&\leq \left(1 - \frac{b(j)}{m}\right) \mathbf{E}\left[\nu_j^t\right] + \sum_{k \in B(j)} \sum_i \frac{\rho_{i,j}^k}{m} \max_i \mathbf{E}\left[\nu_i^t\right].
\end{aligned}$$

□

We can now use Lemma 52 to prove Theorem 51.

Proof of Theorem 51. We need to bound $\max_j \mathbf{E}\left[\nu_j^{t+1}\right]$. Using $\sum_{k \in B(j)} \sum_i \rho_{i,j}^k \leq$

$b(j) - b(j)\gamma$ for all j and Lemma 52 we have

$$\begin{aligned}
\max_j \mathbf{E}\left[\nu_j^{t+1}\right] &\leq \max_j \left[\left(1 - \frac{b(j)}{m}\right)\mathbf{E}\left[\nu_j^t\right] + \sum_{k\in B(j)}\sum_i \frac{\rho_{i,j}^k}{m}\max_i \mathbf{E}\left[\nu_i^t\right]\right] \\
&\leq \max_j \left[\left(1 - \frac{b(j)}{m}\right)\max_i \mathbf{E}\left[\nu_i^t\right] + \sum_{k\in B(j)}\sum_i \frac{\rho_{i,j}^k}{m}\max_i \mathbf{E}\left[\nu_i^t\right]\right] \\
&\leq \max_i \mathbf{E}\left[\nu_i^t\right]\max_j\left(1 + \frac{\sum_{k\in B(j)}\sum_i \rho_{i,j}^k - b(j)}{m}\right) \\
&\leq \max_i \mathbf{E}\left[\nu_i^t\right]\max_j\left(1 - \frac{\gamma b(j)}{m}\right) \\
&= \max_i \mathbf{E}\left[\nu_i^t\right]\left(1 - \frac{\min_j b(j)\gamma}{m}\right).
\end{aligned}$$

Initially $\max_i \mathbf{E}\left[\nu_i^0\right] \leq 1$ for all i since the we use the shortest path between states X_0 and Y_0, so after t updates $\max_i \mathbf{E}\left[\nu_i^t\right] \leq \left(1 - \frac{\min_j b(j)\gamma}{m}\right)^t$ which can be verified by induction on t. Finally $\ell = \sum_{i=1}^n \nu_i^t$ and so $\mathbf{E}\left[\ell\right] \leq n \max_i \mathbf{E}\left[\nu_i^t\right]$. Using this bound

$$\begin{aligned}
\mathrm{d_{TV}}(X_t, Y_t) &\leq \Pr(X_t \neq Y_t) \leq \mathbf{E}\left[\mathrm{Ham}(X_t, Y_t)\right] \leq \mathbf{E}\left[\ell\right] \\
&\leq n\max_i \mathbf{E}\left[\nu_i^t\right] \leq n\left(1 - \frac{\min_j b(j)\gamma}{m}\right)^t
\end{aligned}$$

and the statement of the theorem follows. □

Whilst Weitz's results are not concerned with systematic scan they remain of interest to us since they make use of block dynamics. It is not, however, possible (at least in a general setting) to use Weitz's condition in order to obtain rapid mixing of systematic scan with block dynamics. An inspection of the definitions of α and α_{W} reveals that $\alpha_{\mathrm{W}} \leq \alpha$ and we now exhibit a spin system for which $\alpha_{\mathrm{W}} < 1$ and $\alpha = 1$ but systematic scan does not mix rapidly. It is sufficient to show that a specific systematic scan Markov chain does not mix for the given spin system since it is in the nature of the Dobrushin condition that any mixing result holds for any scan order.

Observation 53. *There exists a spin system for which $\alpha_{\mathrm{W}} < 1$ and $\alpha = 1$ but systematic scan does not mix.*

Consider the following spin system. Let G be the n-vertex cycle and label the sites $0, \ldots, n-1$ and C be the set of q spins. Then Θ_i (which has an associated

transition matrix $P^{[i]}$) is the block containing site i and $i+1 \mod n$ and it is updated as follows:

1. The spin at site i is copied to site $i+1$;
2. a spin is assigned to site i uniformly at random from the set of all spins.

The stationary distribution, π, of the spin system is the uniform distribution on all configurations of G. Clearly $P^{[i]}$ satisfies property (1) of the update rule, namely that only sites within the block may change during the update. To see that π is invariant under each $P^{[i]}$ observe that site $i+1$ takes the spin of site i in the original configuration and site i receives a spin drawn uniformly at random. This ensures that each site has probability $1/q$ of having each spin and that they are independent.

We define the ρ values for this spin system by using the following coupling. Consider a block Θ_j for update. The spin at site $j+1$ is deterministic in both copies, and each copy selects the same colour for site j when drawing uniformly at random from C. First suppose that site j is the discrepancy between two configurations. Then, since the spin at j is copied to site $j+1$, the spin of site $j+1$ becomes a disagreement in the coupling and hence $\rho^j_{j,j+1} = 1$. The spin at j is drawn uniformly at random from C in both copies and coupled perfectly so $\rho^j_{j,j} = 0$. Now suppose that the two configurations differ at a site $i \neq j$. Then $\rho^j_{i,j+1} = 0$ since both configurations have the same colour for site j, and $\rho^j_{i,j} = 0$ since the spins at site j are coupled perfectly. Using the values of ρ we deduce that

$$\alpha_{\mathrm{W}} = \max_j \sum_{k \in B(j)} \sum_i \frac{\rho^k_{i,j}}{b(j)} = \frac{1}{2}\left(\rho^{j-1}_{j-1,j} + \sum_{i \neq j-1} \rho^{j-1}_{i,j} + \sum_i \rho^j_{i,j}\right) = \frac{1}{2}$$

and $\alpha = \max_k \max_{j \in \Theta_k} \sum_i \rho^k_{i,j} = 1$.

Let $\mathcal{M}_{\rightarrow}$ be the systematic scan Markov chain that updates the blocks in the order $\Theta_0, \Theta_1, \ldots, \Theta_{n-1}$. For each block Θ_i note that if a configuration y is obtained from updating block Θ_i starting from x then $y_{i+1} = x_i$. Hence when performing the systematic scan, the spin of site 0 in the original configuration moves around the ring ending at site $n-1$ before the update of block Θ_{n-1} moves it on to site 0. Hence if configuration x' is obtained from one complete scan starting from a configuration x we have $x'_0 = x_0$ and the systematic scan Markov chain does not mix since site 0 will always be assigned the same spin after each complete scan.

Observation 54. *The spin system from Observation 53 also gives*

$$\alpha'_{\mathrm{W}} = \max_i \sum_k \sum_{j \in \Theta_k} \frac{\rho^k_{i,j}}{b(i)} = 1/2.$$

Hence, since the given systematic scan does not mix, it is not possible to find any set of weights that gives $\alpha < 1$.

Remark. It is worth remarking that our observations above do not rule out the possibility of a condition of the form $\max_k \max_i \sum_{j \in \Theta_k} \rho^k_{i,j} < 1$ implies that $\alpha < 1$ and hence that systematic scan mixes. It would however require finding a general method for simultaneously balancing all k influence matrices which seems a difficult task. Furthermore, in the single-site case much of the reason for the interest in matrix balancing is the similarity between the condition $\alpha'_{\mathrm{DGJ}} < 1$ and the path coupling condition, where as in the block case we have shown that the condition $\alpha'_{\mathrm{W}} < 1$ (which is similar to path coupling) does not in general imply rapid mixing of systematic scan.

Chapter 4

Sampling H-colourings of the n-vertex Path

In this chapter we bound the mixing times of systematic scan Markov chains for general H-colourings although at the expense of restricting the class of graphs to paths. We will show that a systematic scan for sampling H-colourings of the n-vertex path mixes in $O(\log n)$ scans for any fixed H which is a significant improvement over the previous bound on the mixing time which was $O(n^5)$ scans. Furthermore we show that for a slightly more restricted family of H (where any two vertices are connected by a 2-edge path) systematic scan also mixes in $O(\log n)$ scans for any scan order using a Dobrushin condition. For completeness we make a small digression to show that a random update Markov chain mixes in $O(n \log n)$ updates for any fixed H, improving the previous bound on the mixing time from $O(n^5)$ updates.

4.1 Preliminaries

Many combinatorial problems are of interest to computer scientists both in their own right and due to their natural applications to statistical physics. Such problems can often be studied by considering *homomorphisms* from the graph of interest G to some fixed graph H. This is known as an H-colouring of G. The vertices of H correspond to colours and the edges of H specify which colours are allowed to be adjacent in an H-colouring of a graph. Let $H = (C, E_H)$ by any fixed graph. Formally an H-colouring of a graph $G = (V, E)$ is a function $h : V \to C$ such that $(h(v), h(u)) \in E_H$ for all edges $(v, u) \in E$ of G. Examples of H-colouring problems are proper q-colourings, independent set configurations, Widom-Rowlinson configurations and the Beach model (see Chapter 2 for details).

Consider a fixed (and connected) graph $H = (C, E_H)$ with maximum vertex-degree Δ_H. Let $C = \{1, \ldots, q\}$ be referred to as the set of colours. Also let $V = \{1, \ldots, n\}$ be the set of sites of the n-vertex path and in particular let V_1 be the set of sites with odd indices and V_2 the set of sites with even indices. We formally say that an H-colouring of the n-vertex path is a function h from V to C such that $(h(i), h(i+1)) \in E_H$ for all $i \in V \setminus \{n\}$. Let Ω^+ be the set of all configurations (all possible assignments of colours to the sites) of the n-vertex path and Ω be the set of all H-colourings of the n-vertex path for the given H. Recall that π is the uniform distribution on Ω. Also recall from previous notation that if $x \in \Omega^+$ is a configuration and $j \in V$ is a site then x_j denotes the colour assigned to j in configuration x. Furthermore, for any set $\Lambda \subseteq V$ let $x_\Lambda = \bigcup_{v \in \Lambda} \{x_v\}$ be the set of colours assigned to sites in Λ. For colours $c, d \in C$ and an integer l let $D^{(l)}_{c,d}$ be the uniform distribution on H-colourings of the region of consecutive sites $L = \{v_1, \ldots, v_l\} \subset V$ consistent with site v_1 being adjacent to a site $i \in V \setminus L$ assigned colour c and site v_l being adjacent to a site in $V \setminus L$ assigned colour d. Also let $D^{(l)}_{c,d}(v_j)$ be the distribution on the colour assigned to site v_j induced by $D^{(l)}_{c,d}$. Observe that for $s < l$

$$\left[D^{(l)}_{c,d} \mid v_1 = c_1, \ldots, v_s = c_s \right] = D^{(l-s)}_{c_s,d}$$

where $D^{(l)}_{c,d} \mid v_1 = c_1, \ldots, v_s = c_s$ is the uniform distribution on H-colourings of L conditioned on site v_1 being assigned colour c_1, v_2 colour c_2 and so on until v_s being assigned colour c_s.

We remind the reader that due to a potential technical difficulty with ensuring the ergodicity of the defined Markov chains we let $\Omega_\sim$ be the state space of the Markov chains in this chapter. Recall that if H is non-bipartite then $\Omega_\sim = \Omega$. Otherwise H is bipartite and we let $\Omega_\sim$ be one of Ω_1 and Ω_2 where $\Omega_1 = \{x \in \Omega : x_1 \in C_1\}$ is the set of H-colourings of the n-vertex path where the first site of the path is assigned a colour from C_1 and similarly $\Omega_2 = \{x \in \Omega : x_1 \in C_2\}$. The sets C_1 and C_2 are the colour classes of H. We will show (Lemma 63) that the constructed Markov chains are ergodic on either Ω_1 or Ω_2 in the bipartite case.

Now recall the definitions of the Markov chains we will study in this chapter. Let $l_1 = \lceil \Delta_H^2 \log(\Delta_H^2 + 1) \rceil + 1$ and let $\Theta = \{\Theta_1, \ldots, \Theta_{m_1}\}$ be any set of $m_1 = \lceil n/l_1 \rceil$ blocks such that each block consists of exactly l_1 consecutive sites and Θ covers V. If $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k then $\mathcal{M}_{\text{AnyOrder}}$ is the systematic scan Markov chain with state space $\Omega_\sim$ and transition matrix $\prod_{k=1}^{m_1} P^{[k]}$. The following bound on the mixing time of

$\mathcal{M}_{\text{AnyOrder}}$ holds for *any order* of the blocks, as is the case for all results obtained by Dobrushin uniqueness.

Theorem 22. *Let H be a fixed connected graph with maximum vertex-degree Δ_H and consider the systematic scan Markov chain $\mathcal{M}_{\text{AnyOrder}}$ on the state space $\Omega_\sim$. Suppose that H is a graph in which every two sites are connected by a 2-edge path. Then the mixing time of $\mathcal{M}_{\text{AnyOrder}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{AnyOrder}}, \varepsilon) \leq \Delta_H^2(\Delta_H^2 + 1)\log(n\varepsilon^{-1})$$

scans of the n-vertex path. This corresponds to $O(n \log n)$ block updates by the construction of the set of blocks.

Remark. We again point out that several well known H-colouring problems satisfy the condition of Theorem 22, for example Widom-Rowlinson configurations, independent set configurations and proper q-colourings for $q \geq 3$. The fact that an H corresponding to 3-colourings satisfies the condition of the theorem is particularly interesting since a lower bound of $\Omega(n^2 \log n)$ scans for single site systematic scan on the path is proved in Dyer at al. [20]. This means that using a simple single site coupling cannot be sufficient to establishing Theorem 22 for any family of H including 3-colourings and hence we have to use block updates.

We go on to show that systematic scan mixes in $O(\log n)$ scans for any fixed graph H by placing more strict restrictions on the construction of the blocks and the order of the scan. Let $s = 4q + 1$, $\beta = \lceil \log(2sq^s + 1) \rceil q^s$ and $l_2 = 2\beta s$. For any integer n consider the following set of $m_2 + 1 = \lfloor 2n/l_2 \rfloor$ blocks $\{\Theta_0, \ldots, \Theta_{m_2}\}$ where

$$\Theta_k = \{k\beta s + 1, \ldots, \min((k+2)\beta s, n)\}.$$

We observe that the set of blocks covers V by construction. Furthermore note that the size of Θ_{m_2} is at least βs and that the size of every other block is exactly l_2. Recall that $\mathcal{M}_{\text{FixedOrder}}$ is the systematic scan Markov chain, with state space $\Omega_\sim$, which performs a heat-bath move on each block in the order $\Theta_0, \ldots, \Theta_{m_2}$. The following theorem improves the mixing time from the corresponding result in Dyer et al. [20] from $O(n^5)$ scans to $O(\log n)$ scans.

Theorem 24. *Let H be any fixed connected graph and consider the systematic scan Markov chain $\mathcal{M}_{\text{FixedOrder}}$ on the state space $\Omega_\sim$. The mixing time of $\mathcal{M}_{\text{FixedOrder}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{FixedOrder}}, \varepsilon) \leq (4sq^s + 2)\log(n\varepsilon^{-1})$$

scans of the n-vertex path. This corresponds to $O(n \log n)$ block updates by the construction of the set of blocks.

Remark. We repeat our earlier remark that although Theorem 24 eclipses Theorem 22 in the sense that it shows the existence of a systematic scan for a broader family of H than Theorem 22 but with the same (asymptotic) mixing time, Theorem 22 remains interesting in its own right since it applies to any order of the scan. Following the proof of Theorem 22 we will discuss (Observation 60) the obstacles one encounters when attempting to extend Theorem 22 to a larger family of H using the same method of proof.

We conclude this chapter by bounding the mixing time of a random update Markov chain for sampling H-colourings of the n-vertex path. Let $\gamma = 2q^s + 1$ and define the following set of $n + s\gamma - 1$ blocks, which is constructed such that each site is contained in exactly $s\gamma$ blocks

$$\Theta_k = \begin{cases} \{k, \ldots, \min(k + s\gamma - 1, n)\} & \text{when } k \in \{1, \ldots, n\} \\ \{1, \ldots, n + s\gamma - k\} & \text{when } k \in \{n+1, \ldots, n + s\gamma - 1\}. \end{cases}$$

Recall that $\mathcal{M}_{\mathrm{RND}}$ is the random update Markov chain, with state space $\Omega_\sim$, which at each step selects a block uniformly at random and performs a heat-bath move on it. The following theorem improves the mixing time from the corresponding result in Dyer et al. [20] from $O(n^5)$ updates to $O(n \log n)$ updates (although as previously remarked the Markov chain presented by Dyer et al. is a single-site chain).

Theorem 26. *Let H be any fixed connected graph and consider the random update Markov chain $\mathcal{M}_{\mathrm{RND}}$ on the state space $\Omega_\sim$. The mixing time of $\mathcal{M}_{\mathrm{RND}}$ is*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{RND}}, \varepsilon) \leq \frac{(n + 2sq^s + s - 1)\log(n\varepsilon^{-1})}{s}$$

block updates.

For technical reasons we extend the state space of the Markov chains as follows. Let Ω_1^+ be the set of configurations where each site in V_1 is assigned a colour from C_1 and each site in V_2 is assigned a colour from C_2 (recall that C_1 and C_2 are the colour classes of H). Similarly, Ω_2^+ is the set of configurations where each site in V_1 is assigned a colour from C_2 and each site in V_2 is assigned a colour from C_1. Formally

$$\Omega_1^+ = \{x \in \Omega^+ : x_{V_1} \subseteq C_1, x_{V_2} \subseteq C_2\}$$

and

$$\Omega_2^+ = \{x \in \Omega^+ : x_{V_1} \subseteq C_2, x_{V_2} \subseteq C_1\}.$$

We then extend the state space of the Markov chains to $\Omega_\sim^+$ where $\Omega_\sim^+ = \Omega^+$ if H is not bipartite and $\Omega_\sim^+$ is one of Ω_1^+ or Ω_2^+ when H is bipartite. The extended Markov chains make the same transitions as the original Markov chains on configurations in $\Omega_\sim$ and hence the extended chains do not make transitions from configurations in $\Omega_\sim$ to configurations outside $\Omega_\sim$. The stationary distributions of the extended chains are uniform over the configurations in $\Omega_\sim$ and zero elsewhere. This approach is standard and the mixing times of the original chains are bounded above by the mixing time of corresponding chain on the extended state space as shown in Lemma 8. For each site $j \in V$, let $S_j^\sim$ denote the set of pairs $(x, y) \in \Omega_\sim^+ \times \Omega_\sim^+$ of configurations that only differ on the colour assigned to site j, that is $x_i = y_i$ for all $i \neq j$. Also let $S_\sim = \bigcup_{j \in V} S_j^\sim$ be the set of all such pairs of configurations. For completeness we show that $S_\sim$ connects the state space $\Omega_\sim^+$ which is required in path coupling applications.

Lemma 55. *The transitive closure of $S_\sim$ is the whole of $\Omega_\sim^+ \times \Omega_\sim^+$.*

Proof. Recall that $S_\sim = \bigcup_{i \in V} S_j^\sim$ where $S_j^\sim \subseteq \Omega_\sim^+ \times \Omega_\sim^+$ is the set of pairs $(x, y) \in \Omega_\sim^+ \times \Omega_\sim^+$ of configurations that differ only on the colour assigned to site j. To establish the lemma it is sufficient, for any pair of configurations $(x, y) \in \Omega_\sim^+ \times \Omega_\sim^+$, to construct a path $x = z^0, z^1, \ldots, z^n = y$ such that $(z^{j-1}, z^j) \in S_j^\sim$ for each $j \in \{1, \ldots, n\}$. We define z^j for $j \in \{1, \ldots, n\}$ as follows

$$z_i^j = \begin{cases} y_i & \text{for } 1 \leq i \leq j \\ x_i & \text{for } j < i \leq n. \end{cases}$$

Informally, configuration z^j agrees with configuration y from site 1 to j and with configuration x from site $j + 1$ to n.

By definition of the configurations $z^0, \ldots, z^n$ it follows that z^{j-1} and z^j only differ on the colour assigned to site j for each $j \in \{1, \ldots, n\}$. Hence we only need to check that $z^j \in \Omega_\sim^+$ for each j. If H is non-bipartite then $\Omega_\sim^+ = \Omega^+$ so $z^j \in \Omega_\sim^+$ for each $j \in \{1, \ldots, n\}$. If H is bipartite then $\Omega_\sim^+$ is one of Ω_1^+ or Ω_2^+. Suppose without loss of generality that $\Omega_\sim^+ = \Omega_1^+$. Then for each $j \in \{1, \ldots n\}$ it holds by definition of Ω_1^+ that the colours x_j and y_j must be from the same colour class of H and hence have $z^j \in \Omega_1^+$. □

4.2 H-colourings of the Path for a Restricted Family of H

This section contains the proof of Theorem 22, namely that $\mathcal{M}_{\text{AnyOrder}}$ mixes in $O(\log n)$ scans when H is a graph in which any two colours are connected via a 2-edge path. Observe that each H for which Theorem 22 is valid is non-bipartite so we let $\Omega_\sim = \Omega$ and as a result $S_j^\sim = S_j$ throughout this section. Recall that Δ_H denotes the maximum vertex-degree of some fixed graph H and that $l_1 = \lceil \Delta_H^2 \log(\Delta_H^2 + 1) \rceil + 1$. The systematic scan Markov chain $\mathcal{M}_{\text{AnyOrder}}$ on $\Omega_\sim$ has transition matrix $\prod_{k=1}^{m_1} P^{[k]}$ where $P^{[k]}$ is the transition matrix for performing a heat-bath move on block Θ_k from a set of $m_1 = \lceil n/l_1 \rceil$ size l_1 blocks covering the n-vertex path. We will bound the mixing time of $\mathcal{M}_{\text{AnyOrder}}$ by bounding the influence on a site and begin by establishing some lemmas required to construct the coupling needed in the proof of Theorem 22.

Lemma 56. *Suppose that for any $c_1, c_2 \in C$ there is a 2-edge path in H from c_1 to c_2. Then for any $c_1, c_2, d \in C$ and integer $s' \geq 2$ there exists a coupling $\psi(D^{(s')}_{c_1,d}, D^{(s')}_{c_2,d})$ of $D^{(s')}_{c_1,d}$ and $D^{(s')}_{c_2,d}$ such that*

(i) $\Pr_{(x',y') \in \psi(D^{(s')}_{c_1,d}, D^{(s')}_{c_2,d})}(x'_{v_1} \neq y'_{v_1}) \leq 1 - \frac{1}{\Delta_H^2}$ *and*

(ii) $\Pr_{(x',y') \in \psi(D^{(2)}_{c_1,d}, D^{(2)}_{c_2,d})}(x'_{v_2} \neq y'_{v_2}) \leq 1 - \frac{1}{\Delta_H^2}$.

Proof. By the condition of the lemma there exists some $c' \in C$ adjacent to both c_1 and c_2 in H. We prove the statement by considering two cases on s'.

First suppose that $s' = 2$. By the condition of the lemma there is some colour d' adjacent to both c' and d in H. There are at most Δ_H^2 valid H-colourings of the sites v_1, v_2 in either of the distributions $D^{(2)}_{c_1,d}$ and $D^{(2)}_{c_2,d}$, and hence the colouring h, which assigns c' to v_1 and d' to v_2, has weight at least $1/\Delta_H^2$ in both. We construct a coupling $\psi(D^{(2)}_{c_1,d}, D^{(2)}_{c_2,d})$ such that

$$\Pr_{(x',y') \in \psi(D^{(2)}_{c_1,d}, D^{(2)}_{c_2,d})}(x' = y' = h) \geq \frac{1}{\Delta_H^2}.$$

The rest of the coupling is arbitrary. This gives the following bounds on the disagreement probabilities at v_1 and v_2

$$\Pr_{(x',y') \in \psi(D^{(2)}_{c_1,d}, D^{(2)}_{c_2,d})}(x'_{v_1} = y'_{v_1}) \geq \Pr_{(x',y') \in \psi(D^{(2)}_{c_1,d}, D^{(2)}_{c_2,d})}(x'_{v_1} = y'_{v_1} = c') \geq \frac{1}{\Delta_H^2}$$

which establishes (i) for $s' = 2$ and

$$\Pr_{(x',y')\in\psi(D^{(2)}_{c_1,d},D^{(2)}_{c_2,d})}(x'_{v_2} = y'_{v_2}) \geq \Pr_{(x',y')\in\psi(D^{(2)}_{c_1,d},D^{(2)}_{c_2,d})}(x'_{v_2} = y'_{v_2} = d') \geq \frac{1}{\Delta_H^2}$$

which establishes (ii).

Now suppose $s' > 2$. Let $\mathrm{adj}(c)$ denote the set of colours adjacent to c in H and n_k the number of H-colourings on the sites $v_4, \ldots, v_{s'}$ consistent with v_3 being assigned colour $k \in C$ and $v_{s'}$ being adjacent to a site (outside the block) coloured d. Also let $p_{c,k}$ be the number of H-colourings of v_1, v_2, v_3 assigning colour c to v_1 and k to v_3 without regard to other sites. Finally let z_i be the number of H-colourings with positive measure in $D^{(s')}_{c_i,d}$ and assume without loss of generality that $z_1 \geq z_2$.

There are at most Δ_H colours available for each site in the block which gives $p_{c,k} \leq \Delta_H$ for any $c, k \in C$ and hence

$$z_1 = \sum_{c\in\mathrm{adj}(c_1)} \sum_{k\in C} p_{c,k} n_k \leq \Delta_H \sum_{c\in\mathrm{adj}(c_1)} \sum_{k\in C} n_k \leq \Delta_H^2 \sum_{k\in C} n_k.$$

Now let $H(c')$ be the set of all H-colourings with positive measure in $D^{(s')}_{c_1,d}$ that assign colour c' to site v_1. Let $h(c')$ denote the size of this set. Now $p_{c,k} \geq 1$ for any $c, k \in C$ since there is a 2-edge path in H between any two colours and hence

$$h(c') = \sum_{k\in C} p_{c',k} n_k \geq \sum_{k\in C} n_k.$$

Observe that, for any $h \in H(c')$, h is at least as likely in $D^{(s')}_{c_2,d}$ as in $D^{(s')}_{c_1,d}$ since we have assumed $z_1 \geq z_2$ without loss of generality. We construct a coupling $\psi(D^{(s')}_{c_1,d}, D^{(s')}_{c_2,d})$ of $D^{(s')}_{c_1,d}$ and $D^{(s')}_{c_2,d}$ in which for each $h \in H(c')$

$$\Pr_{(x',y')\in\psi(D^{(s')}_{c_1,d},D^{(s')}_{c_2,d})}(x' = y' = h) \geq \frac{1}{z_1}.$$

The rest of the coupling is arbitrary. Hence

$$\begin{aligned}\Pr_{(x',y')\in\psi(D^{(s')}_{c_1,d},D^{(s')}_{c_2,d})}(x'_{v_1} = y'_{v_1}) &\geq \sum_{h\in H(c')} \Pr_{(x',y')\in\psi(D^{(s')}_{c_1,d},D^{(s')}_{c_2,d})}(x' = y' = h) \\ &\geq \frac{h(c')}{z_1} \\ &\geq \frac{1}{\Delta_H^2}\end{aligned}$$

using the bounds on z_1 and $h(c')$. This completes the proof. □

We then use Lemma 56 to bound the disagreement probabilities at each site of of the block when a pair of configurations are drawn from a recursively constructed coupling.

Lemma 57. *Suppose that for any $c_1, c_2 \in C$ there is a 2-edge path in H from c_1 to c_2. Then for all $c_1, c_2, d \in C$ and integers $l' \geq 2$ there exists a coupling $\Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})$ of $D^{(l')}_{c_1,d}$ and $D^{(l')}_{c_2,d}$ in which for $j \in \{1, \ldots, l'-1\}$*

$$\Pr_{(x',y') \in \Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \leq \left(1 - \frac{1}{\Delta_H^2}\right)^j$$

and

$$\Pr_{(x',y') \in \Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})}(x'_{v_l} \neq y'_{v_l}) \leq \left(1 - \frac{1}{\Delta_H^2}\right)^{l'-1}.$$

Proof. We recursively construct a coupling $\Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})$ of $D^{(l')}_{c_1,d}$ and $D^{(l')}_{c_2,d}$ using the method set out in Goldberg et al. [33] as follows. Firstly $l' = 2$ is the base case and we use the coupling from Lemma 56. For $l' \geq 3$ we construct a coupling using the following two step process.

1. Couple $D^{(l')}_{c_1,d}(v_1)$ and $D^{(l')}_{c_2,d}(v_1)$ greedily to maximise the probability of assigning the same colour to site v_1 in both distributions.

2. If the same colour c was chosen for v_1 in both distributions in step 1 then the set of valid H-colourings of the remaining sites are the same in both distributions. Hence the conditional distributions $D^{(l')}_{c_1,d} \mid v_1 = c$ and $D^{(l')}_{c_2,d} \mid v_1 = c$ are the same and the rest of the coupling is trivial. Otherwise, for all pairs (c'_1, c'_2) of distinct colours recursively couple $\left[D^{(l')}_{c_1,d} \mid v_1 = c'_1\right] = D^{(l'-1)}_{c'_1,d}$ and $\left[D^{(l')}_{c_2,d} \mid v_1 = c'_2\right] = D^{(l'-1)}_{c'_2,d}$ which is a sub problem of size $l' - 1$.

This completes the coupling construction.

Now for $j \in \{1, \ldots, l'-1\}$ we prove by induction that

$$\Pr_{(x',y') \in \Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \leq \left(1 - \frac{1}{\Delta_H^2}\right)^j. \tag{4.1}$$

The base case, $j = 1$, follows from Lemma 56 since we couple the colour at site v_1 greedily to maximise the probability of agreement at v_1 in the first step of the

recursive coupling. Now suppose that (4.1) is true for $j-1$ then

$$\begin{aligned}
&\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_j}\neq y'_{v_j})\\
&\quad=\sum_{c'_1,c'_2}\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j-1}}=c'_1,y'_{v_{j-1}}=c'_2)\\
&\qquad\times\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d}|v_{j-1}=c'_1,D^{(l')}_{c_2,d}|v_{j-1}=c'_2)}(x'_{v_j}\neq y'_{v_j})\\
&\quad=\sum_{c'_1,c'_2}\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j-1}}=c'_1,y'_{v_{j-1}}=c'_2)\\
&\qquad\times\Pr_{(x',y')\in\Psi(D^{(l'-j+1)}_{c'_1,d},D^{(l'-j+1)}_{c'_2,d})}(x'_{v_1}\neq y'_{v_1})\\
&\quad\leq\sum_{c'_1,c'_2}\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j-1}}=c'_1,y'_{v_{j-1}}=c'_2)\left(1-\frac{1}{\Delta_H^2}\right)\\
&\quad\leq\left(1-\frac{1}{\Delta_H^2}\right)^j
\end{aligned}$$

where the first inequality uses Lemma 56 and the second is the inductive hypothesis.

The $j=l'$ case is similar.

$$\begin{aligned}
&\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_l}\neq y'_{v_l})\\
&\quad=\sum_{c'_1,c'_2}\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{l'-2}}=c'_1,y'_{v_{l'-2}}=c'_2)\\
&\qquad\times\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d}|v_{l'-2}=c'_1,D^{(l')}_{c_2,d}|v_{l'-2}=c'_2)}(x'_{v_{l'}}\neq y'_{v_{l'}})\\
&\quad=\sum_{c'_1,c'_2}\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{l'-2}}=c'_1\wedge y'_{v_{l'-2}}=c'_2)\\
&\qquad\times\Pr_{(x',y')\in\Psi(D^{(2)}_{c'_1,d},D^{(2)}_{c'_2,d})}(x'_{v_2}\neq y'_{v_2})\\
&\quad\leq\left(1-\frac{1}{\Delta_H^2}\right)^{l'-2}\left(1-\frac{1}{\Delta_H^2}\right)=\left(1-\frac{1}{\Delta_H^2}\right)^{l'-1}
\end{aligned}$$

where the inequality uses Lemma 56 and (4.1). □

We can then use the coupling constructed in Lemma 57 to construct a coupling $\Psi_k(x,y)$ of the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ for each pair of configurations $(x,y)\in S_i$. We summarise the disagreement probabilities in this coupling in the following corollary (of Lemma 57).

Corollary 58. *For any sites $i,j\in V$ let $d(i,j)$ denote the edge distance between them and suppose that for any $c,d\in C$ there exists a 2-edge path in H from c to*

d. Then

$$\rho_{i,j}^k \leq \begin{cases} \left(1-\frac{1}{\Delta_H^2}\right)^{d(i,j)} & \text{if } i \text{ is on the boundary of } \Theta_k \text{ and } d(i,j) < l_1 \\ \left(1-\frac{1}{\Delta_H^2}\right)^{l_1-1} & \text{if } i \text{ is on the boundary of } \Theta_k \text{ and } d(i,j) = l_1 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. For each block Θ_k we need to specify a coupling $\Psi_k(x,y)$ of the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ for each pair of configurations $(x,y) \in S_i$ and each $i \in V$. Trivially if $i \in \Theta_k$ then the set of H-colourings with positive measure in each distribution is the same and the same H-colouring can be chosen for each distribution. The same holds when i is not on the boundary of Θ_k.

Suppose that i is on the boundary of Θ_k. Let the other site on the boundary of Θ_k be coloured d in both x and y and hence $P^{[k]}(x,\cdot) = D_{x_i,d}^{(l_1)}$ and $P^{[k]}(y,\cdot) = D_{y_i,d}^{(l_1)}$. We then let $\Psi_k(x,y) = \Psi(D_{x_i,d}^{(l_1)}, D_{y_i,d}^{(l_1)})$ which is the coupling constructed in Lemma 57 and gives the stated bounds on the disagreement probabilities. □

Remark. It is important to note that, given distinct sites i and i' both on the boundary of Θ_k, we may use a different coupling for $\rho_{i,j}^k$ and $\rho_{i',j}^k$. This is the case since, by definition of ρ, the coupling may depend on both the block and the two initial configurations x and y (which in turn determine i). Since x and y only differ on the colour assigned to site i, the coupling is defined to start from the site in Θ_k immediately adjacent to i, and thus we can use a different coupling for $\rho_{i,j}^k$ and $\rho_{i',j}^k$.

The following technical lemma is required in the proof of Theorem 22.

Lemma 59. *For any $0 \leq p \leq 1$ and $j, l \in \mathbb{N}$ where $l \geq 2j$*

$$p^j + p^{l-j+1} \geq p^{j+1} + p^{l-(j+1)+1}.$$

Proof.

$$\begin{aligned} p^j + p^{l-j+1} - p^{j+1} - p^{l-j} &= p^j(1-p) - p^{l-j}(1-p) \\ &= (p^j - p^{l-j})(1-p) \\ &= p^j(1-p^{l-2j})(1-p) \geq 0 \end{aligned}$$

since $0 \leq p \leq 1$ where the last equality uses the fact $l \geq 2j$. □

We are now ready to prove Theorem 22.

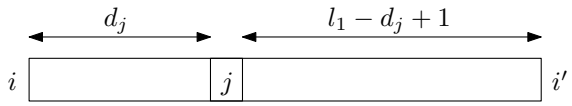

Figure 4.1. A block Θ_k of length l_1.

Theorem 22. *Let H be a fixed connected graph with maximum vertex-degree Δ_H and consider the systematic scan Markov chain $\mathcal{M}_{\text{AnyOrder}}$ on the state space $\Omega_\sim$. Suppose that H is a graph in which every two sites are connected by a 2-edge path. Then the mixing time of $\mathcal{M}_{\text{AnyOrder}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{AnyOrder}}, \varepsilon) \leq \Delta_H^2(\Delta_H^2 + 1)\log(n\varepsilon^{-1})$$

scans of the n-vertex path. This corresponds to $O(n \log n)$ block updates by the construction of the set of blocks.

Proof. We will show that $\alpha < 1$ and then use Theorem 14 to obtain the stated bound on the mixing time. Consider some site $j \in \Theta_k$ and let d_j denote the number of edges between j and the nearest site $i \notin \Theta_k$ on the boundary of Θ_k. Then the distance to the other site, i', on the boundary of Θ_k is $l_1 - d_j + 1$ as shown in Figure 4.1. Notice that $d_j \leq \lceil l_1/2 \rceil$. By Corollary 58 we have

$$\rho_{i,j}^k \leq \left(1 - \frac{1}{\Delta_H^2}\right)^{d_j} \text{ and } \rho_{i',j}^k \leq \mathbf{1}_{d_j \geq 2}\left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - d_j + 1} + \mathbf{1}_{d_j = 1}\left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1}.$$

Now let

$$\alpha_{j,k} = \rho_{i,j}^k + \rho_{i',j}^k \leq \left(1 - \frac{1}{\Delta_H^2}\right)^{d_j} + \mathbf{1}_{d_j \geq 2}\left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - d_j + 1} + \mathbf{1}_{d_j = 1}\left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1}$$

be the influence on site j. Then

$$\alpha = \max_k \max_{j \in \Theta_k} \alpha_{j,k} \leq \max\Bigg\{ \max_{\lceil \frac{l_1}{2} \rceil \geq d_j \geq 2} \left\{ \left(1 - \frac{1}{\Delta_H^2}\right)^{d_j} + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - d_j + 1} \right\}, \left(1 - \frac{1}{\Delta_H^2}\right) + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1} \Bigg\}.$$

Since $d_j \leq \lceil l_1/2 \rceil$ the conditions of Lemma 59 are satisfied for $2 \leq d_j \leq \lceil l_1/2 \rceil - 1$. In particular taking $d_j = \lceil l_1/2 \rceil - 1$, which satisfies the requirements, gives

$$\left(1 - \frac{1}{\Delta_H^2}\right)^{\lceil l_1/2 \rceil - 1} + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - \lceil l_1/2 \rceil + 2} \geq \left(1 - \frac{1}{\Delta_H^2}\right)^{\lceil l_1/2 \rceil} + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - \lceil l_1/2 \rceil + 1}$$

and hence

$$\max_{\lceil \frac{l_1}{2} \rceil \geq d_j \geq 2} \left\{ \left(1 - \frac{1}{\Delta_H^2}\right)^{d_j} + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - d_j + 1} \right\} \leq \left(1 - \frac{1}{\Delta_H^2}\right)^{2} + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1}$$
$$\leq \left(1 - \frac{1}{\Delta_H^2}\right) + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1}$$

which gives

$$\begin{aligned}
\alpha &\leq \left(1 - \frac{1}{\Delta_H^2}\right) + \left(1 - \frac{1}{\Delta_H^2}\right)^{l_1 - 1} \\
&= 1 - \frac{1}{\Delta_H^2} + \left(1 - \frac{1}{\Delta_H^2}\right)^{\lceil \Delta_H^2 \log(\Delta_H^2 + 1) \rceil} \\
&< 1 - \frac{1}{\Delta_H^2} + \frac{1}{\Delta_H^2 + 1} \\
&= 1 - \frac{1}{\Delta_H^2(\Delta_H^2 + 1)}
\end{aligned}$$

by substituting the definition of l_1 and using the fact $(1 - 1/x)^x < e^{-1}$ for $x > 0$. The statement of the theorem now follows by Theorem 14. □

We now take a moment to show that we are unable to use Theorem 14 to prove rapid mixing for systematic scan on H-colourings of the n-vertex path for any H that does not have a 2-edge path between all pairs of colours. This motivates the use of path coupling (at the expense of enforcing a specific scan order) in the subsequent section.

Observation 60. *Let $H = (C, E_H)$ be some fixed and connected graph in which there is no 2-edge path from c_1 to c_2 for some distinct $c_1, c_2 \in C$. Then for any set of m blocks with associated transition matrices $P^{[1]} \dots P^{[m]}$ and any coupling $\Psi_k(x, y)$ for $1 \leq k \leq m$ and $(x, y) \in S_i^{\sim}$ we have $\alpha \geq 1$ in the unweighted setting.*

Proof. Recall that $S_i^{\sim} \subseteq \Omega_{\sim}^+ \times \Omega_{\sim}^+$ where $\Omega_{\sim}^+$ is the set of all configurations (except when H is bipartite in which case $\Omega_{\sim}^+$ is one of Ω_1^+ and Ω_2^+ as described earlier). Note in particular that any given configuration in $\Omega_{\sim}^+$ need not be an H-colouring of the n-vertex path. Also recall that $\rho_{i,j}^k$ is the maximum probability of disagreement at j when drawing from a coupling starting from two configurations $(x, y) \in S_i^{\sim}$. Let x be any proper H-colouring with $x_i = c_1$ and y be the configuration with $y_j = x_j$ for $j \neq i$ and $y_i = c_2$ (If H is bipartite then c_2 is from the same colour class of H as c_1). Note that y is not a proper H-colouring as both edges $(y_{i-1}, y_i) \notin E_H$ and $(y_i, y_{i+1}) \notin E_H$, otherwise the 2-edge paths

$(x_i, x_{i+1} = y_{i+1}, y_i)$ and $(x_i, x_{i-1} = y_{i-1}, y_i)$ would exist in H. However, x and y are both configurations in $\Omega_\sim^+$ and they only differ at the colour of site i so (x, y) is a valid pair in $S_i^\sim$.

Now assume that $\alpha < 1$. Fix some block $\Theta_k = \{i+1, \ldots, i+l\}$ of length l and let $P^{[k]}$ be the transition matrix associated with Θ_k. Also let $\Psi_k(x, y)$ be any coupling of $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$. Since $\alpha < 1$ it must hold that $\rho_{i,j}^k < 1$ for each $j \in \Theta_k$. In particular $\rho_{i,i+1}^k = \Pr_{(x',y') \in \Psi_k(x,y)}(x'_{i+1} \neq y'_{i+1}) < 1$ and so (letting $\mathrm{adj}(c)$ denote the set of colours adjacent to c in H) the set $\mathrm{adj}(c_1) \cap \mathrm{adj}(c_2)$ must be non-empty since there is a positive probability of assigning the same colour to site $i+1$ in both distributions. However take any $d \in \mathrm{adj}(c_1) \cap \mathrm{adj}(c_2)$, then (c_1, d, c_2) is a 2-edge path from c_1 to c_2 in H contradicting the restriction imposed on H and hence $\alpha \geq 1$. □

Remark. It remains to be seen if adding weights will allow a proof in the Dobrushin setting for classes of H not containing 2-edge paths between all colours. However, this can be done using path coupling as we will show in Section 4.3.

4.3 H-colourings of the Path for any H

Recall that $\mathcal{M}_{\mathrm{FixedOrder}}$ is the systematic scan on $\Omega_\sim$ defined as follows. Let $s = 4q+1$, $\beta = \lceil \log(2sq^s+1) \rceil q^s$ and $l_2 = 2\beta s$. Then $\mathcal{M}_{\mathrm{FixedOrder}}$ is the systematic scan which performs a heat-bath move on each of the $m_2 + 1 = \lfloor 2n/l_2 \rfloor$ blocks in the order $\Theta_0, \ldots, \Theta_{m_2}$ where

$$\Theta_k = \{k\beta s + 1, \ldots, \min((k+2)\beta s, n)\}.$$

Note that the size of Θ_{m_2} is at least βs and that every other block is of size l_2. We will prove Theorem 24 which bounds the mixing time of $\mathcal{M}_{\mathrm{FixedOrder}}$. Our method of proof will be path coupling [5] and we begin by establishing some lemmas required to define the coupling we will use in the proof of Theorem 24. The constructions used in the following two lemmas are similar to the ones from Lemma 27 in Dyer et al. [20].

Lemma 61. *If H is not bipartite then for all $c_1, c_2 \in C$ there is an s-edge path in H from c_1 to c_2.*

Proof. Let $c \in C$ be some site on an odd-length cycle in H and let d_1 be the shortest edge-distance from c_1 to c and d_2 the shortest edge-distance from c to c_2. We construct the path as follows. Go from c_1 to c using d_1 edges. If $d_1 + d_2$

is even then go around the cycle using an odd number $q' \leq q$ of edges. Go from c to c_2 in d_2 edges and observe that the constructed path is of odd length. Also the length of the path is at most

$$d_1 + d_2 + q' < 3q.$$

Finally go back and forth on the last edge on the path to make the total length s. □

Lemma 62. *If H is bipartite with colour classes C_1 and C_2 then for all $c_1 \in C_1$ and $c_2 \in C_2$ there is an s-edge path in H from c_1 to c_2.*

Proof. Go from c_1 to c_2 in at most $q-1$ edges and note that the number of edges is odd. Then go back and forth on the last edge to make the total path length equal to s. □

For completeness we present a proof that $\mathcal{M}_{\text{FixedOrder}}$ is ergodic on $\Omega_\sim$.

Lemma 63. *The Markov chain $\mathcal{M}_{\text{FixedOrder}}$ is ergodic on $\Omega_\sim$.*

Proof. Let $P_{\text{FixedOrder}}$ be the transition matrix of $\mathcal{M}_{\text{FixedOrder}}$. We need to show that $\mathcal{M}_{\text{FixedOrder}}$ satisfies the following properties

- irreducible: $P^t_{\text{FixedOrder}}(x, y) > 0$ for each pair $(x, y) \in \Omega_\sim \times \Omega_\sim$ and some integer $t > 0$
- aperiodic: $\gcd\{t : P^t_{\text{FixedOrder}}(x, x) > 0\} = 1$ for each $x \in \Omega_\sim$.

In an application of $P_{\text{FixedOrder}}$ a heat-bath move is made on each block in the order $\Theta_0, \ldots, \Theta_m$. A heat-bath move on any block starting from an H-colouring has a positive probability of self-loop which ensures aperiodicity of the chain. To see that $\mathcal{M}_{\text{FixedOrder}}$ is irreducible consider any pair of H-colourings $(x, y) \in \Omega_\sim \times \Omega_\sim$. We exhibit a sequence of H-colourings $x = \sigma^0, \ldots, \sigma^{m_2+1} = y$ such that $\sigma^k_j = \sigma^{k+1}_j$ for each $0 \leq k \leq m_2$ and $j \in V \setminus \Theta_k$. Using this sequence we observe that $P_{\text{FixedOrder}}(x, y) > 0$ since, for each $0 \leq k \leq m_2$, performing a heat-bath move on block Θ_k to $\sigma^k \in \Omega_\sim$ results in the H-colouring $\sigma^{k+1} \in \Omega_\sim$ with positive probability. Recall that $\Theta_k = \{k\beta s + 1, \ldots, \min((k+2)\beta s, n)\}$. Then let σ^k be given by

$$\sigma^k_i = \begin{cases} y_i & \text{if } 1 \leq i \leq \min((k+2)\beta s - s + 1, n) \\ x_i & \text{if } (k+2)\beta s + 1 \leq i \leq n \\ p(i - (k+2)\beta s + s - 1) & \text{if } (k+2)\beta s - s + 1 < i \leq \min((k+2)\beta s, n) \end{cases}$$

where $p(j)$ is the j-th in the sequence of colours on the s-edge path in H between $p(0) = y_{(k+2)\beta s-s+1}$ and $p(s) = x_{(k+2)\beta s+1}$ given by Lemmas 61 and 62 (since $p(0)$ and $p(s)$ are in opposite colour classes of H in the bipartite case) respectively. □

The following lemma is an analogue of Lemma 13 in Goldberg et al. [33].

Lemma 64. *For any $c_1, c_2, d \in C$ and positive integer $s' \geq s$ such that both $D^{(s')}_{c_1,d}$ and $D^{(s')}_{c_2,d}$ are non-empty there exists a coupling $\psi(D^{(s')}_{c_1,d}, D^{(s')}_{c_2,d})$ of $D^{(s')}_{c_1,d}$ and $D^{(s')}_{c_2,d}$ such that*

$$\Pr_{(x',y')\in\psi(D^{(s')}_{c_1,d},D^{(s')}_{c_2,d})}(x'_{v_s} \neq y'_{v_s}) \leq 1 - \frac{1}{q^s}.$$

Proof. For ease of notation let D_1 denote $D^{(s')}_{c_1,d}$ and D_2 denote $D^{(s')}_{c_2,d}$. For $s' > s$, let n_k be the number of H-colourings on $v_{s+1}, \ldots, v_{s'}$ consistent with v_s being assigned colour $k \in C$ and $v_{s'}$ adjacent to a site (not in L) coloured d. If both $s' = s$ and k is adjacent to d in H then $n_k = 1$. If $s' = s$ but k is not adjacent to d in H then $n_k = 0$. The following definitions are for $i \in \{1, 2\}$. Let $l_i(k)$ be the number of H-colourings on $v_1, \ldots, v_s$ assigning colour k to site v_s and consistent with v_1 being adjacent to a site (not in L) coloured c_i. We also let Z_i be the set of H-colourings on L with positive measure in D_i and z_i be the size of this set. Note that D_i is the uniform distribution on Z_i so for each $x \in Z_i$ $\Pr_{D_i}(x) = 1/z_i$. For each $k \in C$ let $Z_i(k) \subseteq Z_i$ be the set of H-colourings with positive measure in D_i that assign colour k to site v_s and let $z_i(k)$ be the size of this set. Note that $l_i(k)n_k = z_i(k)$ and $\sum_k z_i(k) = z_i$. Let $C^*_i = \{k \in C \mid z_i(k) > 0\}$ be the set of valid colours for v_s in D_i and let $C^* = C^*_1 \cup C^*_2$.

We define a coupling ψ of D_1 and D_2 as follows. Assume without loss of generality that $z_1 \geq z_2$. We create the following mutually exclusive subsets of Z_i. For each $k \in C^*$ let $f(k) = \min(z_1(k), z_2(k))$ and let $F_1(k) = \{\sigma^{(k)}(1), \ldots, \sigma^{(k)}(f(k))\} \subseteq Z_1(k)$ be any subset of H-colourings in Z_1 assigning the colour k to site v_s. Also let $F_2(k) = \{\tau^{(k)}(1), \ldots, \tau^{(k)}(f(k))\} \subseteq Z_2(k)$ and observe that $F_1(k)$ and $F_2(k)$ are of the same size. We then construct ψ such that for each $k \in C^*$ and $j \in \{1, \ldots, f(k)\}$

$$\Pr_{(x',y')\in\psi}(x' = \sigma^{(k)}(j), y' = \tau^{(k)}(j)) = \frac{1}{z_1}.$$

The rest of the coupling is arbitrary. For example let $R_i = Z_i \setminus \left(\bigcup_{k\in C^*} F_i(k)\right)$ be the set of (valid) H-colourings not selected in any of the above subsets of Z_i and the size of R_i be r_i, observing that $r_1 \geq r_2$. Let $R'_1 = \{\sigma(1), \ldots, \sigma(r_2)\} \subseteq R_1$ and

enumerate R_2 such that $R_2 = \{\tau(1), \ldots, \tau(r_2)\}$. Then for $1 \leq j \leq r_2$ let

$$\Pr_{(x',y')\in\psi}(x' = \sigma(j), y' = \tau(j)) = \frac{1}{z_1}.$$

Finish off the coupling by, for each pair $(\sigma \in R_1 \setminus R_1', \tau \in Z_2)$ of H-colourings, letting

$$\Pr_{(x',y')\in\psi}(x' = \sigma, y' = \tau) = \frac{1}{z_1 z_2}.$$

From the construction we can verify that the weight of each colouring $x \in Z_1$ in the coupling is $1/z_1$ and the weight of each colouring $y \in Z_2$ is

$$\frac{1}{z_1} + \frac{z_1 - z_2}{z_1 z_2} = \frac{1}{z_2}$$

since the size of $R_1 \setminus R_1'$ is $z_1 - z_2$. This completes the construction of the coupling.

We will require the following bounds on $l_i(k)$ for each $k \in C^*$

$$1 \leq l_i(k) \leq q^s. \tag{4.2}$$

There are at most q colours available for each site in the block and hence at most q^s valid H-colourings of $v_1, \ldots, v_s$ which gives the upper bound. We establish the lower bound by showing the existence of an s-edge path in H from both c_1 and c_2 to any $k \in C^*$. Suppose that H is non-bipartite, then Lemma 61 guarantees the existence of an s-edge path in H between any two colours in H, satisfying our requirement.

Now suppose that H is bipartite with colour classes C_1 and C_2. Without loss of generality suppose that $c_1 \in C_1$. Since both D_1 and D_2 are non-empty there exists a $(2s'+2)$-edge path in H from c_1 to c_2 (via d) so $c_2 \in C_1$. Let $k \in C^*$ then $k \in C_2$ since there is an s-edge path in H from c_1 to k and s is odd. Lemma 62 implies the existence of an s-edge path between each $c \in C_1$ and each $k \in C_2$ which establishes (4.2).

Using (4.2) to see that $n_k \leq f(k) \leq q^s n_k$ for each $k \in C^*$ we have

$$\begin{aligned}
\Pr_{(x',y')\in\psi}(x'_{v_s} = y'_{v_s}) &= \sum_{k\in C^*} \Pr_{(x',y')\in\psi}(x'_{v_s} = y'_{v_s} = k) \\
&\geq \sum_{k\in C^*} \frac{f(k)}{z_1} \\
&\geq \sum_{k\in C^*} \frac{n_k}{\sum_{k'\in C^*} l_1(k')n_{k'}} \\
&\geq \sum_{k\in C^*} \frac{n_k}{q^s \sum_{k'\in C^*} n_{k'}} \\
&= \frac{1}{q^s}
\end{aligned}$$

which completes the proof. □

Lemma 65. *For any $c_1, c_2, d \in C$ and any positive integer $l' \leq l_2$ such that both $D^{(l')}_{c_1,d}$ and $D^{(l')}_{c_2,d}$ are non-empty there exists a coupling Ψ of $D^{(l')}_{c_1,d}$ and $D^{(l')}_{c_2,d}$ in which for $1 \leq j \leq l'$*

$$\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \leq \left(1 - \frac{1}{q^s}\right)^{\lfloor \frac{j}{s} \rfloor}.$$

Proof. We construct a coupling $\Psi(D^{(l')}_{c_1,d}, D^{(l')}_{c_2,d})$ of $D^{(l')}_{c_1,d}$ and $D^{(l')}_{c_2,d}$ using the following two step process, based on the recursive coupling in Goldberg et al. [33].

1. If $l' < s$ then couple the distributions any valid way which completes the coupling. Otherwise, couple $D^{(l')}_{c_1,d}(v_s)$ and $D^{(l')}_{c_2,d}(v_s)$ greedily to maximise the probability of assigning the same colour to site v_s in both distributions. Then, independently in each distribution, colour the sites $v_1, \ldots, v_{s-1}$ consistent with the uniform distribution on H-colourings. Note that it is possible to do this since we obtained the colour for site v_s in each distribution from the induced distribution on that site. If $l' = s$ this completes the coupling.

2. If the same colour is assigned to v_s then the remaining sites can be coloured the same way in both distributions since the conditional distributions are the same. Otherwise, for all pairs (c'_1, c'_2) of distinct colours the coupling is completed by recursively constructing a coupling of $\left[D^{(l')}_{c_1,d} \mid v_s = c'_1\right] = D^{(l'-s)}_{c'_1,d}$ and $\left[D^{(l')}_{c_2,d} \mid v_s = c'_2\right] = D^{(l'-s)}_{c'_2,d}$.

This completes the coupling construction and we will prove by strong induction

that for $j \in \{1, \ldots, l'\}$

$$\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \leq \left(1 - \frac{1}{q^s}\right)^{\lfloor \frac{j}{s} \rfloor}. \tag{4.3}$$

Firstly the cases $1 \leq j \leq s-1$ are established by observing that $\lfloor j/s \rfloor = 0$ and the probability of disagreement at any site is at most 1. The case $j = s$ is established in Lemma 64. Now for $s < j \leq l'$, suppose that (4.3) holds for all positive integers less than j. Let $S_- = \{s, 2s, \ldots\}$ and define the quantities j_- and a_j by $j_- = \max\{x \in S_- \mid x < j\} = a_j s$ observing that $1 \leq j - j_- \leq s$. Now

$$\begin{aligned}
&\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \\
&\quad= \sum_{c'_1,c'_2} \Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j_-}} = c'_1, y'_{v_{j_-}} = c'_2) \\
&\qquad\times \Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d}|v_{j_-}=c'_1,D^{(l')}_{c_2,d}|v_{j_-}=c'_2)}(x'_{v_j} \neq y'_{v_j}) \\
&\quad= \sum_{c'_1,c'_2} \Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j_-}} = c'_1, y'_{v_{j_-}} = c'_2) \\
&\qquad\times \Pr_{(x',y')\in\Psi(D^{(l'-j_-)}_{c'_1,d},D^{(l'-j_-)}_{c'_2,d})}(x'_{v_{j-j_-}} \neq y'_{v_{j-j_-}}).
\end{aligned}$$

Observe that for any pair (c'_1, c'_2) of colours, if the probabilities of assigning c'_1 to v_{j_-} in $D^{(l')}_{c_1,d}$ and c'_2 to v_{j_-} in $D^{(l')}_{c_2,d}$ are both non-zero then the distributions $D^{(l'-j_-)}_{c'_1,d}$ and $D^{(l'-j_-)}_{c'_2,d}$ are both non-empty and hence, using Lemma 64 for $l' - j_- \geq s$ and upper-bounding probability of disagreement by one otherwise, we get

$$\begin{aligned}
&\Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_j} \neq y'_{v_j}) \\
&\quad\leq \sum_{c'_1,c'_2} \Pr_{(x',y')\in\Psi(D^{(l')}_{c_1,d},D^{(l')}_{c_2,d})}(x'_{v_{j_-}} = c'_1, y'_{v_{j_-}} = c'_2) \left(\mathbf{1}_{j-j_-=s}(1 - 1/q^s) + \mathbf{1}_{j-j_-\neq s}\right) \\
&\quad\leq \begin{cases} \left(1 - \frac{1}{q^s}\right)^{\lfloor \frac{j_-}{s} \rfloor + 1} & \text{if } j - j_- = s \\ \left(1 - \frac{1}{q^s}\right)^{\lfloor \frac{j_-}{s} \rfloor} & \text{if } j - j_- \neq s \end{cases}
\end{aligned} \tag{4.4}$$

where last inequality is the inductive hypothesis since $j_- < j$.

First consider the case $j - j_- \neq s$ in which we have $j_- + b = j$ for some $1 \leq b \leq s-1$. Then

$$\left\lfloor \frac{j_- - 1}{s} \right\rfloor = \left\lfloor \frac{a_j s - 1}{s} \right\rfloor = a_j - 1 < a_j = \left\lfloor \frac{a_j s}{s} \right\rfloor = \left\lfloor \frac{j_-}{s} \right\rfloor$$

and so for $1 \leq b \leq s-1$

$$\left\lfloor \frac{j_- + b}{s} \right\rfloor = \left\lfloor \frac{j_-}{s} \right\rfloor$$

which implies that

$$\left\lfloor \frac{j_-}{s} \right\rfloor = \left\lfloor \frac{j}{s} \right\rfloor. \tag{4.5}$$

Now suppose $j - j_- = s$ which substituting for j_- gives

$$\left\lfloor \frac{j_-}{s} \right\rfloor = \left\lfloor \frac{j-s}{s} \right\rfloor = \left\lfloor \frac{j}{s} \right\rfloor - 1. \tag{4.6}$$

Substituting (4.5) and (4.6) in (4.4) completes the proof. □

We are now ready to define the coupling of the distributions of configurations obtained from one complete scan of the Markov chain $\mathcal{M}_{\text{FixedOrder}}$. The coupling is defined for pairs $(x, y) \in S_i^{\sim}$. We will let (x', y') denote the pair of configurations after one complete scan of $\mathcal{M}_{\text{FixedOrder}}$ starting from (x, y) and let (x^k, y^k) be the pair of configurations obtained by updating blocks $\Theta_0, \ldots, \Theta_{k-1}$ starting from $(x, y) = (x^0, y^0)$. Observe that (x', y') is obtained by updating block Θ_{m_2} from the pair (x^{m_2}, y^{m_2}).

The coupling for updating block Θ_k is defined as follows. Let i and i' be the sites on the boundary of Θ_k. The order of the scan will ensure that at most one of the boundaries is a disagreement in (x^k, y^k), so we only need to define the coupling for boundaries disagreeing on at most one end of Θ_k; suppose without loss of generality that $x^k_{i'} = y^k_{i'} = d$ for some $d \in C$. Firstly, if $x^k_i = y^k_i$ then the set of valid configurations arising from updating Θ_k is the same in both distributions and we use the identity coupling.

Otherwise $x^k_i \neq y^k_i$. Suppose that $k \neq m_2$. If H is not bipartite then Lemma 61 implies the existence of an l_2-edge path between both x^k_i and d and between y^k_i and d. If H is bipartite then x^k_i and y^k_i are in the same colour class but d is in the opposite colour class of H since l_2 is even. Lemma 62 implies the existence of an l_2-edge path between both x^k_i and d and between y^k_i and d. Hence both distributions $D^{(l_2)}_{x^k_i,d}$ and $D^{(l_2)}_{y^k_i,d}$ are non-empty and we obtain (x^{k+1}, y^{k+1}) from $\Psi(D^{(l_2)}_{x^k_i,d}, D^{(l_2)}_{y^k_i,d})$ which is the coupling constructed in Lemma 65. Note that if $k = m_2$ (i.e. the block is the last block which may not be of size l_2) then both distributions remain (trivially) non-empty. For ease of reference we state the following corollary of Lemma 65.

Corollary 66. *For any two sites $v, u \in V$ let $d(v, u)$ denote the edge distance between them. For any block Θ_k let i and i' be the sites on the boundary of Θ_k*

and suppose that $x^k_{i'} = y^k_{i'} = d$ *for some* $d \in C$. *Obtain* (x^{k+1}, y^{k+1}) *from the above coupling. Then for any* $j \in \Theta_k$

$$\Pr(x^{k+1}_j \neq y^{k+1}_j) \leq \begin{cases} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{d(i,j)}{s} \right\rfloor} & \text{if } x^k_i \neq y^k_i \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 67. *For any positive integers* s, k, x

$$\sum_{j=1}^{sk} \left(1 - \frac{1}{x}\right)^{\left\lfloor \frac{j}{s} \right\rfloor} < sx.$$

Proof.

$$\sum_{j=1}^{sk} \left(1 - \frac{1}{x}\right)^{\left\lfloor \frac{j}{s} \right\rfloor} = (s-1) + s\sum_{j=1}^{k-1} \left(1 - \frac{1}{x}\right)^j + \left(1 - \frac{1}{x}\right)^k < s\sum_{j\geq 0} \left(1 - \frac{1}{x}\right)^j < sx.$$

□

The following lemma implies Theorem 24 by Corollary 9 (path coupling).

Lemma 68. *Suppose that* $(x, y) \in S_i^{\sim}$ *and obtain* (x', y') *by one complete scan of* $\mathcal{M}_{\text{FixedOrder}}$. *Then*

$$\mathbf{E}\left[\text{Ham}(x', y')\right] < 1 - \frac{1}{4sq^s + 2}.$$

Proof. First suppose that i is not on the boundary of any block and that Θ_b is the first block containing i. In this case Corollary 66 gives us $\Pr(x^{b+1}_i \neq y^{b+1}_i) = 0$ and so $\text{Ham}(x', y') = 0$.

Now suppose that i is on the boundary of some block Θ_a. Recall the definition of a block

$$\Theta_k = \{k\beta s + 1, \ldots, \min(k\beta s + 2\beta s, n)\}.$$

If i is also contained in a block $\Theta_{a'}$ with $a' < a$ then Corollary 66 gives $\Pr(x^{a'+1}_i \neq y^{a'+1}_i) = 0$ and hence $\text{Ham}(x', y') = 0$.

If site i is not updated before Θ_a then $i = (a+2)\beta s + 1$ as shown in Figure 4.2 and the disagreement percolates through the sites in Θ_a during the update of Θ_a. Using Corollary 66 we have for $j \in \Theta_a$

$$\Pr(x^{a+1}_j \neq y^{a+1}_j) \leq \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{i-j}{s} \right\rfloor} \tag{4.7}$$

in particular, the sites in $\Theta_a \setminus \Theta_{a+1} = \{a\beta s + 1, \ldots (a+1)\beta s\}$ will not get updated

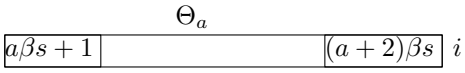

Figure 4.2. Site i is on the boundary of Θ_a and is not contained in any block $\Theta_{a'}$ with $a' < a$.

again during the scan and hence for $j \in \Theta_a \setminus \Theta_{a+1}$

$$\Pr(x'_j \neq y'_j) \leq \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{(a+2)\beta s+1-j}{s} \right\rfloor}. \tag{4.8}$$

Now consider the update of any block Θ_k from the pair of configurations (x^k, y^k) where $k > a$. There cannot be a disagreement at site $(k+2)\beta s+1$ since that site has not been updated (and it was not the initial disagreement) so the only site on the boundary of Θ_k that could be a disagreement in (x^k, y^k) is $k\beta s$. Hence from Corollary 66, for $j \in \{k\beta s+1, \ldots, \min((k+2)\beta s, n)\}$

$$\Pr(x^{k+1}_j \neq y^{k+1}_j \mid x^k_{k\beta s} \neq y^k_{k\beta s}) \leq \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{j-k\beta s}{s} \right\rfloor}. \tag{4.9}$$

We show by induction on k that for $a+1 \leq k \leq m_2$

$$\Pr(x^k_{k\beta s} \neq y^k_{k\beta s}) \leq \left(1 - \frac{1}{q^s}\right)^{\beta(k-a)}. \tag{4.10}$$

The base case, $k = a+1$ follows from (4.7) since $j = k\beta s = (a+1)\beta s = a\beta s + \beta s \in \Theta_a$. Now suppose that (4.10) is true for $k-1$. Then

$$\begin{aligned}
\Pr(x^k_{k\beta s} \neq y^k_{k\beta s}) &= \Pr(x^k_{k\beta s} \neq y^k_{k\beta s} \mid x^{k-1}_{(k-1)\beta s} \neq y^{k-1}_{(k-1)\beta s}) \Pr(x^{k-1}_{(k-1)\beta s} \neq y^{k-1}_{(k-1)\beta s}) \\
&\leq \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{k\beta s-(k-1)\beta s}{s} \right\rfloor} \left(1 - \frac{1}{q^s}\right)^{\beta(k-a-1)} \\
&= \left(1 - \frac{1}{q^s}\right)^{\beta} \left(1 - \frac{1}{q^s}\right)^{\beta(k-a-1)} \\
&= \left(1 - \frac{1}{q^s}\right)^{\beta(k-a)}
\end{aligned}$$

using the inductive hypothesis and (4.9).

Now for each site $j \geq (a+1)\beta s+1$, that is site j is updated at least once following block Θ_a, write $j = k_j\beta s + b_j$ with $1 \leq b_j \leq \beta s$ where k_j denotes is the

index of the block in which j is last updated.

$$\begin{aligned}\Pr(x'_j \neq y'_j) &= \Pr(x_j^{k_j+1} \neq y_j^{k_j+1}) \\ &\leq \Pr(x_j^{k_j+1} \neq y_j^{k_j+1} \mid x_{\beta k_j s}^{k_j} \neq y_{\beta k_j s}^{k_j}) \Pr(x_{\beta k_j s}^{k_j} \neq y_{\beta k_j s}^{k_j}).\end{aligned}$$

We can then apply (4.9) to the first component of the product since $j \in \{k_j\beta s + 1, \ldots, \min(k_j\beta s + 2\beta s, n)\}$ and (4.10) to the second since $a+1 \leq k_j \leq m_2$ to get

$$\Pr(x'_j \neq y'_j) \leq \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{b_j}{s} \right\rfloor} \left(1 - \frac{1}{q^s}\right)^{\beta(k_j - a)}.$$

Then, using linearity of expectation and (4.8), we have

$$\begin{aligned}\mathbf{E}\left[\mathrm{Ham}(x', y')\right] &= \sum_j \Pr(x'_j \neq y'_j) \\ &= \sum_{j \in \Theta_a \setminus \Theta_{a+1}} \Pr(x'_j \neq y'_j) + \sum_{j \in \bigcup_{k \geq a+1} \Theta_k} \Pr(x'_j \neq y'_j) \\ &\leq \sum_{j=as\beta+1}^{(a+1)\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{(a+2)\beta s + 1 - j}{s} \right\rfloor} \\ &\quad + \sum_{k_j=a+1}^{m_2} \sum_{b_j=1}^{\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{b_j}{s} \right\rfloor} \left(1 - \frac{1}{q^s}\right)^{\beta(k_j - a)} \\ &= \sum_{r=1}^{\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{\beta s + r}{s} \right\rfloor} + \sum_{k_j=a+1}^{m_2} \left(1 - \frac{1}{q^s}\right)^{\beta(k_j - a)} \sum_{b_j=1}^{\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{b_j}{s} \right\rfloor} \\ &< \left(1 - \frac{1}{q^s}\right)^{\beta} \sum_{r=1}^{\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{r}{s} \right\rfloor} \\ &\quad + \sum_{t \geq 1} \left(\left(1 - \frac{1}{q^s}\right)^{\beta}\right)^t \sum_{b_j=1}^{\beta s} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{b_j}{s} \right\rfloor} \\ &< \left(1 - \frac{1}{q^s}\right)^{\beta} sq^s + \frac{\left(1 - \frac{1}{q^s}\right)^{\beta} sq^s}{1 - \left(1 - \frac{1}{q^s}\right)^{\beta}}\end{aligned}$$

where the last inequality uses Lemma 67 and the sum of a geometric progression. Substituting the definition of β and using the fact $(1 - 1/x)^x < e^{-1}$ for $x > 0$ we

get

$$\begin{aligned}
\mathbf{E}\left[\mathrm{Ham}(x',y')\right] &< \left(1-\frac{1}{q^s}\right)^{\lceil \log(2sq^s+1)\rceil q^s} sq^s + \frac{\left(1-\frac{1}{q^s}\right)^{\lceil \log(2sq^s+1)\rceil q^s} sq^s}{1-\left(1-\frac{1}{q^s}\right)^{\lceil \log(2sq^s+1)\rceil q^s}} \\
&< \frac{sq^s}{e^{\lceil \log(2sq^s+1)\rceil}} + \frac{sq^s}{e^{\lceil \log(2sq^s+1)\rceil}\left(1-e^{-\lceil \log(2sq^s+1)\rceil}\right)} \\
&= \frac{sq^s}{e^{\lceil \log(2sq^s+1)\rceil}} + \frac{sq^s}{e^{\lceil \log(2sq^s+1)\rceil}-1} \\
&\leq \frac{sq^s}{2sq^s+1} + \frac{sq^s}{2sq^s} \\
&= 1 - \frac{1}{4sq^s+2}
\end{aligned}$$

which completes the proof. □

4.4 H-colourings of the Path with a Random Update Markov Chain

Recall that the random update Markov chain $\mathcal{M}_{\mathrm{RND}}$ on $\Omega_\sim$ is defined as follows. We again let $s = 4q+1$ and define $\gamma = 2q^s+1$. We then define a set of $n+s\gamma-1$ blocks of size at most $s\gamma$ as follows

$$\Theta_k = \begin{cases} \{k,\ldots,\min(k+s\gamma-1,n)\} & \text{when } k \in \{1,\ldots,n\} \\ \{1,\ldots,n+s\gamma-k\} & \text{when } k \in \{n+1,\ldots,n+s\gamma-1\}. \end{cases}$$

By construction of the set of blocks each site is adjacent to at most two blocks and furthermore each site is contained in exactly $s\gamma$ blocks. One step of $\mathcal{M}_{\mathrm{RND}}$ consists of selecting a block uniformly at random and performing a heat-bath update on it. We will prove (using path coupling) Theorem 26 namely that $\mathcal{M}_{\mathrm{RND}}$ mixes in $O(n\log n)$ updates for any H.

We begin by defining the required coupling. For a pair of configurations $(x,y) \in S_i^\sim$ we obtain the pair (x',y') by one step of $\mathcal{M}_{\mathrm{RND}}$. That is we select a block uniformly at random and perform a heat bath move on that block. We can again use Lemma 65 from Section 4.3 to construct the required coupling for updating block Θ_k since the definition of s is the same in both Markov chains. If i is not on the boundary of Θ_k then the sets of valid H-colourings of Θ_k are the same in both distributions and we use the identity coupling. If i is on the boundary of Θ_k then we let the other site on the boundary be coloured d in

both x and y. We then obtain (x', y') from $\Psi(D_{x_i,d}^{(s\gamma)}, D_{y_i,d}^{(s\gamma)})$ which is the coupling constructed in Lemma 65. The disagreement probabilities are summarised in the following corollary of Lemma 65.

Corollary 69. *For any two sites $v, u \in V$ let $d(v, u)$ denote the edge distance between them. Suppose that a block Θ_k has been selected to be updated. For any pair $(x, y) \in S_i^{\sim}$ obtain (x', y') from the above coupling. Then for any $j \in \Theta_k$*

$$\Pr(x'_j \neq y'_j) \leq \begin{cases} \left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{d(i,j)}{s} \right\rfloor} & \text{if } i \text{ is on the boundary of } \Theta_k \\ 0 & \text{otherwise.} \end{cases}$$

The following lemma implies Theorem 26 by Corollary 9 (path coupling).

Lemma 70. *Suppose that $(x, y) \in S_i^{\sim}$ and obtain (x', y') by one step of $\mathcal{M}_{\mathrm{RND}}$. Then*

$$\mathbf{E}\left[\mathrm{Ham}(x', y')\right] < 1 - \frac{s}{n + 2sq^s + s - 1}.$$

Proof. There are $s\gamma$ blocks containing site i and if such a block is selected then $\mathrm{Ham}(x', y') = 0$. There are at most 2 blocks adjacent to site i and if such a block is selected then the discrepancy percolates in the block according to the probabilities stated in Corollary 69. This leaves $n + s\gamma - 1 - s\gamma - 2 = n - 3$ blocks that leave the Hamming distance unchanged. Hence, using Lemma 67, we have

$$\begin{aligned} \mathbf{E}\left[\mathrm{Ham}(x', y')\right] &\leq \frac{2}{n + s\gamma - 1}\left(1 + \sum_{j=1}^{\gamma s}\left(1 - \frac{1}{q^s}\right)^{\left\lfloor \frac{j}{s} \right\rfloor}\right) + \frac{n-3}{n + s\gamma - 1} \\ &< \frac{n-1}{n + s\gamma - 1} + \frac{2sq^s}{n + s\gamma - 1} \\ &= \frac{2sq^s + n - 1}{2sq^s + n - 1 + s} = 1 - \frac{s}{2sq^s + n - 1 + s} \end{aligned}$$

by substituting the definition of γ. □

Chapter 5

Sampling 7-colourings of the Grid

In this chapter we will be concerned with sampling from the uniform distribution on the set of proper 7-colourings of a finite-size rectangular grid using a systematic scan Markov chain. Recall from a previous chapter that proper q-colourings of the grid correspond to the zero-temperature anti-ferromagnetic q-state Potts model on the square lattice, a model of significant importance in statistical physics. The systematic scan Markov chain that we present cycles through blocks consisting of 2×2 sub-grids and performs heat-bath updates on them. We give a computer-assisted proof that this systematic scan Markov chain mixes in $O(\log n)$ scans, where n is the size of the rectangular sub-grid. We make use of a heuristic to compute required couplings for the updates of the 2×2 sub-grids. This is the first time the mixing time of a systematic scan Markov chain on the grid has been shown to mix for less than 8 colours, a result which is implied by Theorem 16. Finally we also give partial results that underline the challenges of proving rapid mixing of a systematic scan Markov chain for sampling 6-colourings of the grid by considering 2×3 and 3×3 sub-grids. We give lower bounds on the appropriate influence parameter that imply that the proof technique we employ does not imply rapid mixing of systematic scan for 6-colourings of the grid when using 2×2, 2×3 and 3×3 sub-grids.

5.1 Preliminaries

We present a computer-assisted proof that a systematic scan Markov chain mixes rapidly when considering 7-colourings of the grid. We will bound the influence on a site, which we have previously shown implies rapid mixing of systematic scan, by using a heuristic to mechanically construct sufficiently good couplings of proper colourings of a 2×2 sub-grid. We will hence use a heuristic based computation

in order to establish a rigorous result about the mixing time of a systematic scan Markov chain. Throughout this chapter we let the weights assigned to the sites of the underlying graph be uniform and hence omit them.

Let $G = (V, E)$ be a finite rectangular grid with toroidal boundary conditions. Working on the torus is common practice as it avoids treating several technicalities regarding the sites on the boundary of a finite grid as special cases and hence lets us present the proof in a more "clean" way. We point out however that these technicalities are straightforward to deal with. Let Ω be the set of all proper 7-colourings of G and π be the uniform distribution on Ω. Recall that if $x \in \Omega^+$ is a configuration and $j \in V$ is a site then x_j denotes the colour assigned to site j in configuration x. Furthermore, for a subset of sites $\Lambda \subseteq V$ and a configuration $x \in \Omega^+$ we let x_Λ denote the colouring of the sites in Λ under x. Recall the definition of the Markov chain $\mathcal{M}_{\text{grid}}$. Let $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ be a set of m blocks such that each block $\Theta_k \subseteq V$ is a 2×2 sub-grid that covers V. For each block Θ_k, $P^{[k]}$ is the transition matrix for performing a heat-bath move on Θ_k. Then $\mathcal{M}_{\text{grid}}$ is the systematic scan Markov chain with state space Ω and transition matrix $P_{\text{grid}} = \prod_{k=1}^{m} P^{[k]}$. We will prove Theorem 28 which is the following bound on the mixing time of $\mathcal{M}_{\text{grid}}$.

Theorem 28. *Let G be a finite and rectangular piece of the infinite square lattice. Consider the systematic scan Markov chain $\mathcal{M}_{\text{grid}}$ on Ω. The mixing time of $\mathcal{M}_{\text{grid}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{grid}}, \varepsilon) \leq 63 \log(n\varepsilon^{-1})$$

scans of the grid. This corresponds to $O(n \log n)$ block updates since each block is of size 4.

As usual we extend the state space of the chain to Ω^+ in order to make use of Theorem 14 in the analysis. Lemma 8 implies that the derived bound on the mixing time of the extended chain is also an upper bound on the mixing time of $\mathcal{M}_{\text{grid}}$. Recall from Chapter 2 that $\partial\Theta_k$ is the *boundary* of Θ_k, namely the set of sites adjacent to, but not included in, Θ_k. Note from our previous definitions that $x_{\partial\Theta_k}$ denotes the colouring of the boundary of Θ_k under a configuration $x \in \Omega^+$. We will refer to $x_{\partial\Theta_k}$ as a *boundary colouring*. Finally we say that a 7-colouring of the 2×2 sub-grid Θ_k *agrees* with a boundary colouring $x_{\partial\Theta_k}$ if (1) no adjacent sites in Θ_k are assigned the same colour and (2) each site $j \in \Theta_k$ is assigned a colour that is different to the colours of all boundary sites adjacent to j.

5.2 Bounding the Mixing Time of Systematic Scan

This section contains a proof of Theorem 28 although the proof of a crucial lemma, which requires computer-assistance, is deferred to Section 5.3. We will bound the mixing time of $\mathcal{M}_{\text{grid}}$ by bounding the maximum influence on a site, which as usual we denote by α. If α is sufficiently small then Theorem 14 implies that $\mathcal{M}_{\text{grid}}$ mixes in $O(\log n)$ scans regardless of the order of the blocks.

In order to upper bound α we are required to upper bound the probability of a discrepancy at each site $j \in \Theta_k$ under a coupling $\Psi_k(x, y)$ of the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ for any pair of configurations $(x, y) \in S_i$ that only differ at the colour of site i. Our main task is hence to specify a coupling $\Psi_k(x, y)$ of $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ for each pair of configurations $(x, y) \in S_i$ and upper bound the probability of assigning a different colour to each site in a pair of colourings drawn from that coupling.

Consider any block Θ_k and any pair of colourings $(x, y) \in S_i$ that differ only on the colour assigned to some site i. Observe that the distribution on valid configurations for Θ_k induced by $P^{[k]}(x, \cdot)$ only depends on the boundary colouring $x_{\partial\Theta_k}$. If $i \notin \partial\Theta_k$ then the distributions on the configurations for Θ_k induced by $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ respectively, are the same and we let $\Psi_k(x, y)$ be the coupling in which any pair of configurations drawn from $\Psi_k(x, y)$ agree on Θ_k. That is, if the pair (x', y') of configurations are drawn from $\Psi_k(x, y)$ then $x' = x$ off Θ_k, $y' = y$ off Θ_k and $x' = y'$ on Θ_k. This gives $\rho^k_{i,j} = 0$ for any $i \notin \partial\Theta_k$ and $j \in \Theta_k$.

We now need to construct $\Psi_k(x, y)$ for the case when $i \in \partial\Theta_k$. For ease of reference we let $p_j(\Psi_k(x, y)) = \Pr_{(x',y')\in\Psi_k(x,y)}(x'_j \neq y'_j)$ denote the probability of a disagreement at site j in a pair of configurations drawn from $\Psi_k(x, y)$. Note that $\rho^k_{i,j} = \max_{(x,y)\in S_i}\{p_j(\Psi_k(x, y))\}$. For each $j \in \Theta_k$ we need $p_j(\Psi_k(x, y))$ to be sufficiently small in order to avoid $\rho^k_{i,j}$ being too big. If the $\rho^k_{i,j}$-values are too big the parameter α will be too big (that is greater than one) and we cannot make use of Theorem 14 to show rapid mixing. Constructing $\Psi_k(x, y)$ by hand such that $p_j(\Psi_k(x, y))$ is sufficiently small is a difficult task. It is, however, straight forward to mechanically determine which configurations have positive measure in the distributions $P^{[k]}(x, \cdot)$ and $P^{[k]}(y, \cdot)$ for a given pair of boundary colourings $x_{\partial\Theta_k}$ and $y_{\partial\Theta_k}$. It is important to observe from the definition of $\rho^k_{i,j}$ that $\Psi_k(x, y)$ is a function of x and y (and hence also of i), but that the coupling construction cannot depend on site j (see Section 3.5 of Chapter 3 for a more detailed discussion

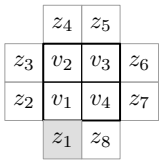

Figure 5.1. General labeling of the sites in a 2×2-block Θ_k and the sites $\partial\Theta_k$ on the boundary of the block.

of this). By considering each case separately we can hence "tune" the coupling to work best for each individual case, which is a main difference from the hand-proofs of the previous chapters where we generally needed to consider a worst-case scenario in the coupling construction. From the distributions $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$ we can hence use some suitable heuristic to construct a coupling that is good enough for our purposes. We hence need to construct a specific coupling for each individual pair of configurations differing only at the colour assigned to a single site, which is done via the following lemma whose proof requires computer-assistance and is deferred to Section 5.3.

Lemma 71. *Let $v_1,\dots,v_4$ be the four sites in a 2×2-block and $z_1,\dots,z_8$ be the boundary sites of the block. Let the labeling be as in Figure 5.1. Let Z and Z' be any two 7-colourings (not necessarily proper) of the boundary sites such that Z and Z' agree on each site except on z_1. Let π_Z and $\pi_{Z'}$ be the uniform distributions on proper 7-colourings of the block that agree with Z and Z', respectively. For $i=1,\dots,4$ let $p_{v_i}(\Psi)$ denote the probability that the colour of site v_i differs in a pair of colourings drawn from a coupling Ψ of π_Z and $\pi_{Z'}$. Then there exists a coupling Ψ such that $p_{v_1}(\Psi)<0.283$, $p_{v_2}(\Psi)<0.079$, $p_{v_3}(\Psi)<0.051$ and $p_{v_4}(\Psi)<0.079$.*

We use the coupling Ψ from Lemma 71 to construct $\Psi_k(x,y)$ in the $\partial\Theta_k$ case as follows. The colouring of Θ_k is drawn from the coupling Ψ of π_Z and $\pi_{Z'}$ where Z is the boundary colouring obtained from $x_{\partial\Theta_k}$ and Z' is obtained from $y_{\partial\Theta_k}$. The colour of the remaining sites, $V\backslash\Theta_k$, are unchanged. That is, if the pair (x',y') of configurations are drawn from $\Psi_k(x,y)$ then $x'=x$ off Θ_k, $y'=y$ off Θ_k and the colourings of Θ_k in x' and y' are drawn from the coupling Ψ in Lemma 71 (see the proof for details on how to construct Ψ). It is straightforward to verify that this is indeed a coupling of $P^{[k]}(x,\cdot)$ and $P^{[k]}(y,\cdot)$. Note that due to the symmetry of the 2×2-block, with respect to rotation and mirroring, we can always label the sites of Θ_k and $\partial\Theta_k$ such that label z_1 in Figure 5.1 represents site i on the boundary. Hence we can make use of Lemma 71 to compute upper bounds on the parameters $\rho^k_{i,j}$. We summarise the $\rho^k_{i,j}$-values in the following

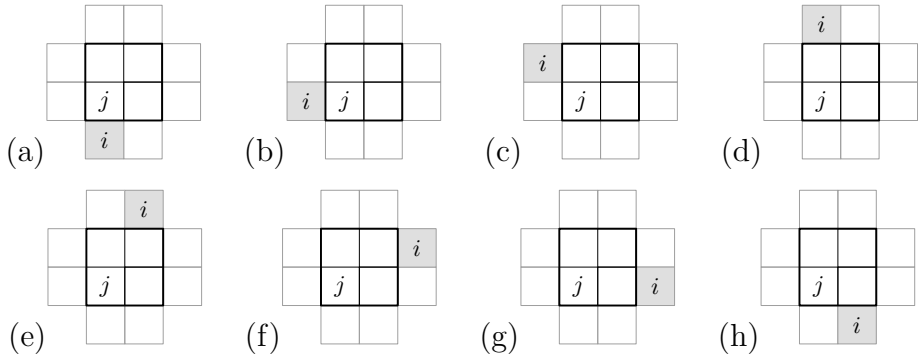

Figure 5.2. A 2×2-block Θ_k showing all eight positions of a site $i \in \partial\Theta_k$ on the boundary of the block in relation to a site $j \in \Theta_k$ in the block.

Corollary of Lemma 71. Due to the symmetry of the block we can assume that site $j \in \Theta_k$ in the corollary is located in the bottom left corner, as Figure 5.2 shows.

Corollary 72. *Let Θ_k be any 2×2-block, let $j \in \Theta_k$ be any site in the block and let $i \in \partial\Theta_k$ be a site on the boundary of the block. Then*

$$\rho_{i,j}^k < \begin{cases} 0.283, & \text{if } i \text{ and } j \text{ are positioned as in Figure 5.2(a) or (b),} \\ 0.079, & \text{if } i \text{ and } j \text{ are positioned as in Figure 5.2(c) or (h),} \\ 0.051, & \text{if } i \text{ and } j \text{ are positioned as in Figure 5.2(e) or (f),} \\ 0.079, & \text{if } i \text{ and } j \text{ are positioned as in Figure 5.2(d) or (g).} \end{cases}$$

If $i \notin \partial\Theta_k$ is not on the boundary of the block then $\rho_{i,j}^k = 0$.

Remark. Lemma 71 is stated such that, in the proof, we only need to consider boundary colourings which is an advantage in the representation of the computer-assisted proof. Corollary 72 provides the link between the boundary colourings of Lemma 71 and the set of all configurations. This link is required for the proof of Theorem 28.

Theorem 28. *Let G be a finite and rectangular piece of the infinite square lattice. Consider the systematic scan Markov chain $\mathcal{M}_{\text{grid}}$ on Ω. The mixing time of $\mathcal{M}_{\text{grid}}$ is*

$$\text{Mix}(\mathcal{M}_{\text{grid}}, \varepsilon) \leq 63 \log(n\varepsilon^{-1})$$

scans of the grid. This corresponds to $O(n \log n)$ block updates since each block is of size 4.

Proof. Let $\alpha_{k,j} = \sum_i \rho^k_{i,j}$ be the influence on j under Θ_k. We need to show that $\alpha_{k,j} < 1$ for each block Θ_k and site $j \in \Theta_k$ in order to ensure that $\alpha = \max_k \max_{j\in\Theta_k} \alpha_{k,j} < 1$. Fix any block Θ_k and any site $j \in \Theta_k$. A site $i \in \partial\Theta_k$ on the boundary of the block can occupy eight different positions on the boundary in relation to j as showed in Figure 5.2(a)–(h). Thus, using the bounds from Corollary 72 we have

$$\alpha_{k,j} = \sum_i \rho^k_{i,j} < 2(0.283 + 0.079 + 0.051 + 0.079) = 0.984.$$

Then $\alpha = \max_k \max_{j\in\Theta_k} \alpha_{k,j} < \max_k 0.984 = 0.984 < 1$ and we obtain the stated bound on the mixing time of $\mathcal{M}_{\text{grid}}$ by Theorem 14. □

Of course we have yet to establish a proof of Lemma 71, which is what the subsequent section will be concerned with. Our computational proof uses some ideas described by Goldberg et al. [33] which have been further explored by Goldberg, Jalsenius, Martin and Paterson [31]. In particular, we will be focusing on minimising the probability of assigning different colours to site v_1 in the couplings constructed by our programs. We will however be required to construct a coupling on the 2×2 sub-grid, rather than establishing bounds on the disagreement probability of a site adjacent to the initial discrepancy and then extending this to a coupling on the whole block recursively. Our approach is similar to the one Achlioptas et al. [1] take, however we do not have the option of constructing an "optimal" coupling using a suitable linear program (even when feasible) since our probabilities will be maximised over all boundary colourings. The crucial difference between the approaches is that Achlioptas et al. [1] are using path coupling as a proof technique which requires them to bound the expected Hamming distance between a pair of colourings dawn from a coupling. This in turn enables them to specify an "optimal" coupling which minimises Hamming distance for a given boundary colouring. We are, however, required to bound the influence of i on j for each boundary colouring and sum over the maximum of these influences. The reason for this is the inherit maximisation over boundary colourings in the definition of $\rho^k_{i,j}$.

Remark. It is worth mentioning that providing bounds on the expected Hamming distance is similar to showing that the influence of a site is small. Recall that this condition is known to imply rapid mixing of a random update Markov chain (see for example Weitz [55]). In a single-site setting the condition "the influence of a site is small" also implies rapid mixing of systematic scan (Dyer

et al. [18]) however in a block setting it is not sufficient to give rapid mixing of systematic scan as we discussed in Section 3.5 of Chapter 3 which is why we need to bound the influence *on* a site.

5.3 Constructing the Coupling by Machine

In order to prove Lemma 71 we will construct a coupling Ψ of π_Z and $\pi_{Z'}$ for all pairs of boundary colourings Z and Z' that are identical on all sites except for site z_1. Recall that π_Z and $\pi_{Z'}$ are the uniform distributions on proper 7-colourings of the block that agree with Z and Z' respectively. For each coupling constructed we verify that the probabilities $p_{v_i}(\Psi)$, $i = 1, \ldots, 4$, are within the bounds of the lemma. The method is well suited to be carried out with computer-assistance and we have implemented a C-program to do so. For details of the program see `http://www.csc.liv.ac.uk/~kasper/grid_scan/`. Before stating the proof of Lemma 71 we will discuss how a coupling can be represented by an edge-weighted complete bipartite graph. We make use of this representation of Ψ in the proof of the lemma.

5.3.1 Representing a Coupling as a Bipartite Graph

Let U be a set of objects and let W be a set of $|U|$ pairs (s, ω_s) such that $s \in U$ and $\omega_s \geq 0$ is a non-negative value representing the weight of s. Each element $s \in U$ is contained exactly once in W. If the value ω_s is an integer (which it is in our case) it can be regarded as the multiplicity of s in a multiset. The set W is referred to as a *weighted set of* U. Let $\pi_{U,W}$ be the distribution on U such that the probability of s is proportional to ω_s, where $(s, \omega_s) \in W$. More precisely, the probability of s in $\pi_{U,W}$ is $\Pr_{\pi_{U,W}}(s) = \omega_s / \sum_{(t,\omega_t)\in W} \omega_t$. For example, let W be a weighted set of U and let $U' \subseteq U$ be a subset of U. Assume the weight $\omega_s = 0$ if $s \in U \backslash U'$ and $\omega_s = k$ if $s \in U'$, where $k > 0$ is a positive constant. Then $\pi_{U,W}$ is the uniform distribution on U'.

The reason for introducing the notion of a weighted set is that it can be used when specifying a coupling of two distributions. Let U be a set and let W and W' be two weighted sets of U such that the sum of the weights in W equals the sum of the weights in W'. Let ω_{total} denote this sum. That is, $\omega_{\text{total}} = \sum_{(s,\omega_s)\in W} \omega_s = \sum_{(s',\omega'_{s'})\in W'} \omega'_{s'}$. The two weighted sets W and W' define two distributions $\pi_{U,W}$ and $\pi_{U,W'}$ on U. We want to specify a coupling Ψ of $\pi_{U,W}$ and $\pi_{U,W'}$. Let $K_{|U|,|U|}$ be an edge-weighted complete bipartite graph with vertex sets W and W'. That

is, for each pair $(s, \omega_s) \in W$ there is an edge to every pair $(s', \omega'_{s'}) \in W'$. Every edge e of $K_{|U|,|U|}$ has a weight $\omega_e \geq 0$ such that the following condition holds. Let (s, ω_s) be any pair in $W \cup W'$ and let E be the set of all $|U|$ edges incident to (s, ω_s). Then $\sum_{e \in E} \omega_e = \omega_s$. It follows that the sum of the edge weights of all $|U|^2$ edges in $K_{|U|,|U|}$ is ω_{total}. The idea is that $K_{|U|,|U|}$ represents a coupling Ψ of $\pi_{U,W}$ and $\pi_{U,W'}$. In order to draw a pair of elements from Ψ we randomly select an edge e in $K_{|U|,|U|}$ proportional to its weight. The endpoints of e represent the elements in U drawn from $\pi_{U,W}$ and $\pi_{U,W'}$. More precisely, the probability of choosing edge e in $K_{|U|,|U|}$ with weight ω_e is $\omega_e/\omega_{\text{total}}$. If edge $e = ((s, \omega_s), (s', \omega'_{s'}))$ is chosen it means that we have drawn s from $\pi_{U,W}$ and s' from $\pi_{U,W'}$, the marginal distributions of Ψ.

The bipartite graph representation of a coupling will be used when we construct couplings of colourings of 2×2-blocks in the proof of Lemma 71.

5.3.2 Proof of Lemma 71

Lemma 71. *Let $v_1, \ldots, v_4$ be the four sites in a 2×2-block and $z_1, \ldots, z_8$ be the boundary sites of the block. Let the labeling be as in Figure 5.1. Let Z and Z' be any two 7-colourings (not necessarily proper) of the boundary sites such that Z and Z' agree on each site except on z_1. Let π_Z and $\pi_{Z'}$ be the uniform distributions on proper 7-colourings of the block that agree with Z and Z', respectively. For $i = 1, \ldots, 4$ let $p_{v_i}(\Psi)$ denote the probability that the colour of site v_i differs in a pair of colourings drawn from a coupling Ψ of π_Z and $\pi_{Z'}$. Then there exists a coupling Ψ such that $p_{v_1}(\Psi) < 0.283$, $p_{v_2}(\Psi) < 0.079$, $p_{v_3}(\Psi) < 0.051$ and $p_{v_4}(\Psi) < 0.079$.*

Proof. Fix two boundary colourings Z and Z' that differ on site z_1. Let c be the colour of site z_1 in Z and let $c' \neq c$ be the colour of z_1 in Z'. Let Q_Z and $Q_{Z'}$ be the two sets of proper 7-colourings of the block that agree with Z and Z', respectively. Let Q be the set of all proper 7-colourings of the block without taking a boundary colouring into account. Let W_Z and $W_{Z'}$ be two weighted sets of Q. The weights are assigned as follows.

- For the pair $(x, \omega_x) \in W_Z$ let the weight $\omega_x = |Q_{Z'}|$ if $x \in Q_Z$, otherwise let $\omega_x = 0$.

- For the pair $(x, \omega_x) \in W_{Z'}$ let the weight $\omega_x = |Q_Z|$ if $x \in Q_{Z'}$, otherwise let $\omega_x = 0$.

It follows from the assignment of the weights that the distribution π_{Q,W_Z} is the uniform distribution on Q_Z. That is, $\pi_{Q,W_Z} = \pi_Z$. Similarly, $\pi_{Q,W_{Z'}}$ is the uniform distribution $\pi_{Z'}$ on $Q_{Z'}$. Note that the sum of the weights is $|Q_Z||Q_{Z'}|$ in both W_Z and $W_{Z'}$. Then a coupling Ψ of π_{Q,W_Z} and $\pi_{Q,W_{Z'}}$ can be specified with an edge-weighted complete bipartite graph $K = K_{|Q|,|Q|}$. For a given valid assignment to the weights of the edges of K, making K represent a coupling Ψ, we can compute the probability of assigning different colours to a site v_i within the block in two configurations drawn from Ψ. Let E_K be the set of all edges $e = ((x, \omega_x), (x', \omega'_{x'}))$ in K such that x and x' differ on site v_i. Then $p_{v_i}(\Psi) = \sum_{e \in E_K} \omega_e / |Q_Z||Q_{Z'}|$.

In order to obtain sufficiently small upper bounds on $p_{v_i}(\Psi)$ for the four sites $v_1, \ldots, v_4$ in the block we would like to assign weights to the edges of K such that much weight is assigned to edges between colourings that agree on many sites in the block. In general it is not clear exactly how to assign weights to the edges. For instance, if we assign too much weight to edges between colourings that are identical on site v_2 we might not be able to assign as much weight as we would like to on edges between colourings that are identical on site v_4. Thus, the probability of assigning different colours to site v_4 would increase. Intuitively a good strategy would be to assign as much weight as possible to edges between colourings that are identical on the whole block. This implies that we try to assign as much weight as possible to edges between colourings that are identical on site v_1, the site adjacent to the discrepancy site z_1 on the boundary. If site v_1 is assigned different colours it should be a good idea to assign as much weight as possible to edges between colourings that are identical on the whole block apart from site v_1. This idea leads to a heuristic in which the assignment of the edge weights is divided into three phases. The exact procedure is described as follows.

In phase one we match identical colourings. For all colourings $x \in Q$ of the block the edge $e = ((x, \omega_x), (x, \omega'_x))$ in K will be given weight $\omega_e = \min(\omega_x, \omega'_x)$. That is, we maximise the probability of drawing the same colouring x from both π_{Q,W_Z} and $\pi_{Q,W_{Z'}}$.

For the following two phases we define an ordering of the colourings in Q. We order the colourings lexicographically with respect to the site order v_3, v_2, v_4, v_1. That is, if the seven colours are $1, \ldots, 7$ the colouring of v_3, v_2, v_4, v_1 will start with 1, 1, 1, 1, respectively. The next colouring will be 1, 1, 1, 2, and so on. This ordering of colourings in Q carries over to an ordering of the pairs in W_Z and $W_{Z'}$. That is, we order the pairs (x, ω_x) in W_Z with respect to the lexicographical ordering of x. Similarly we order the pairs in $W_{Z'}$. This ordering of the pairs will be important in the next two phases. It provides some control of how colourings

are being paired up in terms of the assignment of the weights on edges between pairs. Edges will be considered with respect to this ordering because choosing an arbitrary ordering of the edges would not necessarily result in probabilities $p_{v_i}(\Psi)$ that would be within the bounds of the lemma.

In the second phase we ignore the colour of site v_1 and match colourings that are identical on all of the remaining three sites v_2, v_3 and v_4. More precisely, for each pair $(x, \omega_x) \in W_Z$, considered in the ordering explained above, we consider the edges $e = ((x, \omega_x), (x', \omega'_{x'}))$ where $x \in Q$ and $x' \in Q$ are identical on all sites but v_1. The edges are considered in the ordering of the second component $(x', \omega'_{x'}) \in W_{Z'}$. We assign as much weight as possible to e such that the total weight on edges incident to $(x, \omega_x) \in W_Z$ does not exceed ω_x and such that the total weight on edges incident to $(x', \omega'_{x'}) \in W_{Z'}$ does not exceed $\omega'_{x'}$. Note that in the lexicographical ordering of the colourings, site v_1 is the least significant site and therefore the ordering provides some level of control of pairing up colourings that are similar on the remaining three sites. It turns out that the resulting coupling is sufficiently good for proving the lemma.

In the third and last phase we assign the remaining weights on the edges. As in phase two, for each pair $(x, \omega_x) \in W_Z$ we consider the edges $e = ((x, \omega_x), (x', \omega'_{x'}))$. The pairs and edges are considered in accordance with the ordering explained above. The difference between the second and third phase is that now we do not have any restrictions on the colourings x and x'. We assign as much weight as possible to e such that the total weight on edges incident to $(x, \omega_x) \in W_Z$ does not exceed ω_x and such that the total weight on edges incident to $(x', \omega'_{x'}) \in W_{Z'}$ does not exceed $\omega'_{x'}$. After phase three we have assigned all weights to the edges of K and hence K represents a coupling Ψ of π_Z and $\pi_{Z'}$.

From K we compute the probabilities $p_{v_1}(\Psi)$, $p_{v_2}(\Psi)$, $p_{v_3}(\Psi)$ and $p_{v_4}(\Psi)$ as described above. We have written a C-program which loops through all (non-symmetric) colourings Z and Z' of the boundary of the block and constructs the bipartite graph K as described above. For each boundary the probabilities $p_{v_1}(\Psi)$, $p_{v_2}(\Psi)$, $p_{v_3}(\Psi)$ and $p_{v_4}(\Psi)$ are successfully verified to be within the bounds of the lemma. For details on the C-program, see `http://www.csc.liv.ac.uk/~kasper/grid_scan/`. □

5.4 Partial Results for 6-colourings of the Grid

As we have seen, a systematic scan on the grid using 2×2-blocks and seven colours mixes rapidly. An immediate question is whether we can do better and show rapid

mixing with six colours which is possible in the random update case. This matter will be discussed in this section and we will show that, even with bigger block sizes (up to 3×3), it is not possible to show rapid mixing using the technique of this paper. More precisely, we will establish lower bounds on the parameter α for 2×2-blocks, 2×3-blocks and 3×3-blocks. All three lower bounds are greater than one and hence we cannot make use of Theorem 14 to show rapid mixing.

5.4.1 Establishing Lower Bounds for 2×2 Blocks

We start by examining the 2×2-block again but this time with six colours. Lemma 71 provides upper bounds (under any colourings of the boundary) on the probabilities of having discrepancies at each of the four sites of the block when two 7-colourings are drawn from the specified coupling. For six colours we will show lower bounds on these probabilities under any coupling and a specified pair of boundary colourings. Once again, let $v_1, \ldots, v_4$ be the four sites in a 2×2-block and let $z_1, \ldots, z_8$ be the boundary sites of the block and let the labeling be as in Figure 5.1. Let Z and Z' be any two 6-colourings of the boundary sites that assign the same colour to each site except for z_1. Let π_Z and $\pi_{Z'}$ be the uniform distributions on the sets of proper 6-colourings of the block that agree with Z and Z', respectively. Let $\Psi_{v_k}^{\min}(Z, Z')$ be a coupling of π_Z and $\pi_{Z'}$ that minimises $p_{v_k}(\Psi)$. That is, $p_{v_k}(\Psi) \geq p_{v_k}(\Psi_{v_k}^{\min}(Z, Z'))$ for all couplings Ψ of π_Z and $\pi_{Z'}$. Also let $p_{v_k}^{\text{low}} = \max_{Z,Z'} p_{v_k}(\Psi_{v_k}^{\min}(Z, Z'))$. We can hence say that there exist two 6-colourings Z and Z' of the boundary of a 2×2 block, that assign the same colour to each site except for z_1, such that $p_{v_k}(\Psi) \geq p_{v_i}^{\text{low}}$ for any coupling Ψ of π_Z and $\pi_{Z'}$. We have the following lemma, which is proved by computation.

Lemma 73. *Consider* 6*-colourings of the* 2×2*-block in Figure 5.1. Then* $p_{v_1}^{\text{low}} \geq 0.379$*,* $p_{v_2}^{\text{low}} \geq 0.107$*,* $p_{v_3}^{\text{low}} \geq 0.050$ *and* $p_{v_4}^{\text{low}} \geq 0.107$*.*

Proof. Fix one site v_k in the block and fix two colourings Z and Z' of the boundary of the block that differ only on the colour of site z_1. Let Q_Z and $Q_{Z'}$ be the two sets of proper 6-colourings of the block that agree with Z and Z', respectively. For $c = 1, \ldots, 6$ let n_c be the number of colourings in C_Z in which site v_k is assigned colour c. Similarly let n'_c be the number of colourings in $Q_{Z'}$ in which site v_k is assigned colour c. It is clear that the probability that v_k is assigned colour c in a colouring x' drawn from π_Z is $\Pr_{\pi_Z}(x'_{v_k} = c) = n_c/|Q_Z|$. For $c = 1, \ldots, 6$ define $m_c = n_c|Q_{Z'}|$, $m'_c = n'_c|Q_Z|$ and $M = |Q_Z||Q_{Z'}|$. It follows that $\Pr_{\pi_Z}(x'_{v_k} = c) = m_c/M$ and $\Pr_{\pi_{Z'}}(y'_{v_k} = c) = m'_c/M$, where x' and y' are

colourings drawn from π_Z and $\pi_{Z'}$, respectively. Observe that the quantities m_c, m'_c and M can be easily computed for a given pair of boundary colourings.

Now let Ψ be any coupling of π_Z and $\pi_{Z'}$. The probability that site v_k is coloured c in both colourings drawn from Ψ is be at most $\min(m_c, m'_c)/M$. Therefore, the probability of drawing two colourings from Ψ such that the colour of site v_k is the same in both colourings is at most $\sum_{c=1,\ldots,6} \min(m_c, m'_c)/M$, and the probability of assigning different colours to site v_k is at least $p_{v_k}(\Psi) \geq 1 - \sum_{c=1,\ldots,6} \min(m_c, m'_c)/M$. We have successfully verified the bounds in the statement of the lemma by maximising the lower bound on $p_{v_k}(\Psi)$ over all boundary colourings Z and Z' for each site v_k in the block. The computations are carried out with the help of a computer program written in C. For details on the program, see `http://www.csc.liv.ac.uk/~kasper/grid_scan/`. □

For seven colours, Corollary 72 makes use of Lemma 71 to establish upper bounds on the influence parameters $\rho^k_{i,j}$. These parameters are used in the proof of Theorem 28 to obtain an upper bound on the parameter α. The upper bound on α is shown to be less than one which implies rapid mixing for seven colours when applying Theorem 14. We can use Lemma 73 to obtain lower bounds on the influence parameters $\rho^k_{i,j}$ by completing the coupling in a way analogous to the coupling in Corollary 72. This in turn will result in a lower bound on the parameter α that is greater than one. That is, following the proof of Theorem 28 and making use of Lemma 73, a lower bound on α will be

$$\alpha \geq 2(0.379 + 0.107 + 0.050 + 0.107) = 1.286 > 1.$$

Hence we fail to show rapid mixing of systematic scan with six colours using 2×2-blocks using this approach.

5.4.2 Bigger Blocks

We failed to show rapid mixing of systematic scan with six colours and 2×2-blocks and we will now show that increasing the block size to both 2×3 and 3×3 will not be sufficient either when using the technique from Theorem 14. Lemma 74 below considers 2×3-blocks and is analogous to Lemma 73. We make use of the same notation as for Lemma 73, only the block is bigger and the labeling of the sites is different (see Figure 5.3(a)). Lemma 74 is proved by computation in the same way as Lemma 73. For details on the C-program used in the proof, see `http://www.csc.liv.ac.uk/~kasper/grid_scan/`.

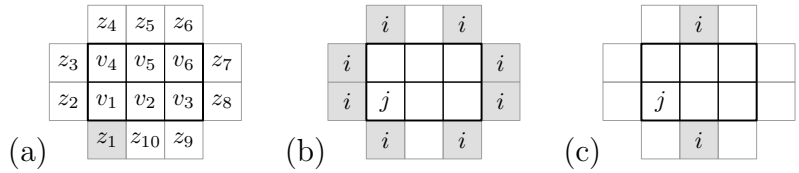

Figure 5.3. (a) General labeling of the sites in a 2×3-block Θ_k and the sites $\partial\Theta_k$ on the boundary of the block. (b)–(c) All ten positions of a site $i \in \partial\Theta_k$ on the boundary of the block in relation to a site $j \in \Theta_k$ in the corner of the block.

Lemma 74. *Consider* 6*-colourings of the* 2×3*-block in Figure 5.3*(a)*. Then* $p_{v_1}^{\text{low}} \geq 0.3671$, $p_{v_3}^{\text{low}} \geq 0.0298$, $p_{v_4}^{\text{low}} \geq 0.0997$ *and* $p_{v_6}^{\text{low}} \geq 0.0174$.

We will now use Lemma 74 to show that $\alpha > 1$ for 2×3 blocks. Let Θ_k be any 2×3-block and let $j \in \Theta_k$ be a site in a corner of the block. A site $i \in \partial\Theta_k$ on the boundary of the block can occupy ten different positions on the boundary in relation to j. See Figure 5.3(b) and (c). We can again determine lower bounds on the influences $\rho_{i,j}^k$ of i on j under Θ_k from Lemma 74. However, Lemma 74 provides lower bounds on $\rho_{i,j}^k$ only when $i \in \partial\Theta_k$ is adjacent to a corner site of the block, as in Figure 5.3(b). If i is located as in Figure 5.3(c) we do not know more than that $\rho_{i,j}^k$ is bounded from below by zero. Nevertheless, the lower bound on α exceeds one. Let $\alpha_{k,j} = \sum_i \rho_{i,j}^k$ be the influence on j under Θ_k. Following the proof of Theorem 28 and using the lower bounds in Lemma 74 we have

$$\begin{aligned} \alpha_{k,j} &= \sum_{i \text{ in Fig. 5.3(b)}} \rho_{i,j}^k + \sum_{i \text{ in Fig. 5.3(c)}} \rho_{i,j}^k \\ &\geq 2(0.3671 + 0.0298 + 0.0997 + 0.0174) = 1.028, \end{aligned}$$

where we set the lower bound on the second sum to zero. Now,

$$\alpha = \max_k \max_{j \in \Theta_k} \alpha_{k,j} \geq 1.028 > 1.$$

Hence we cannot use Theorem 14 to show rapid mixing of systematic scan with six colours and 2×3-blocks. It is interesting to note that considering 2×3-blocks was sufficient for Achlioptas et al. [1] to prove mixing of a random update Markov chain for sampling 6-colourings of the grid.

Lastly, we increase the block size to 3×3 and show that a lower bound on α is still greater than one. We have the following lemma which is proved by computation in the same way as Lemmas 73 and 74.

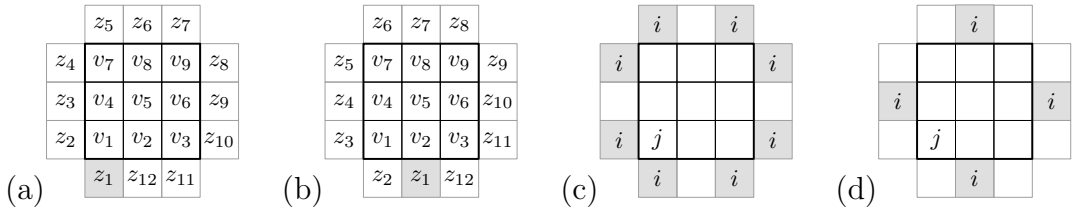

Figure 5.4. (a)–(b) General labeling of the sites in a 3×3-block Θ_k and two different labellings of the sites $\partial\Theta_k$ on the boundary of the block. The discrepancy site on the boundary has label z_1. (b)–(c) All twelve positions of a site $i \in \partial\Theta_k$ on the boundary of the block in relation to a site $j \in \Theta_k$ in the corner of the block.

Lemma 75. *For* 6*-colourings of the* 3×3*-block with sites labeled as in Figure 5.4*(a) *we have* $p_{v_1}^{\text{low}} \geq 0.3537$, $p_{v_3}^{\text{low}} \geq 0.0245$, $p_{v_7}^{\text{low}} \geq 0.0245$ *and* $p_{v_9}^{\text{low}} \geq 0.0071$. *Furthermore, for* 6*-colourings of the* 3×3*-block in Figure 5.4*(b) *we have* $p_{v_1}^{\text{low}} \geq 0.0838$, $p_{v_3}^{\text{low}} \geq 0.0838$, $p_{v_7}^{\text{low}} \geq 0.0138$ *and* $p_{v_9}^{\text{low}} \geq 0.0138$.

Note that Lemma 75 provides lower bounds on the probabilities of having a mismatch on a corner site of the block when the discrepancy site on the boundary (labeled z_1) is adjacent to a corner site (Figure 5.4(a)) and adjacent to a middle site (Figure 5.4(b)). Let Θ_k be any 3×3-block and let $j \in \Theta_k$ be a site in a corner of the block. A site $i \in \partial\Theta_k$ on the boundary of the block can occupy twelve different positions on the boundary in relation to j. See Figure 5.4(c) and (d). Analogous to Corollary 72 lower bounds on the influences $\rho_{i,j}^k$ of i on j under Θ_k can be determined from Lemma 75. Let $\alpha_{k,j} = \sum_i \rho_{i,j}^k$ be the influence on j under Θ_k. Following the proof of Theorem 28 and using the lower bounds in Lemma 75 we have

$$\begin{aligned}
\alpha_{k,j} &= \sum_{i \text{ in Fig. 5.4(c)}} \rho_{i,j}^k + \sum_{i \text{ in Fig. 5.4(d)}} \rho_{i,j}^k \\
&\geq 2(0.3537 + 0.0245 + 0.0245 + 0.0071) \\
&\quad +(0.0838 + 0.0838 + 0.0138 + 0.0138) \\
&= 1.0148.
\end{aligned}$$

Thus, $\alpha = \max_k \max_{j \in \Theta_k} \alpha_{k,j} \geq 1.0148 > 1$. Hence, we cannot use Theorem 14 to show rapid mixing of systematic scan with six colours and 3×3-blocks.

A natural question is whether we can show rapid mixing using even bigger blocks. It seems possible to do this although the computations rapidly become intractable as the block size increases. Already with a 3×3-block the number of boundary colourings we need to consider (after removing isomorphisms) is in

excess of 10^6 and for each boundary colouring there are more than 10^7 colourings of the block to consider. In addition to simply generating the distributions on colourings of the block, the time it would take to actually construct the required couplings, as we did in the proof of Lemma 71, would also increase. Finally when using a larger block size, different positions of site j in the block need to be considered whereas we could make use of to the symmetry of the 2×2-block to only consider one position of site j in the block. If different positions of j have to be considered this has to be captured in the construction of the coupling and would likely require more computations.

The above discussion suggests that in order to show rapid mixing for six and fewer colours of systematic scan on the grid one may need to rely on a different proof technique than Dobrushin uniqueness in the form of Theorem 14. Furthermore, the fact that path coupling can be used to show rapid mixing of a random update Markov chain for 6-colourings of the grid seems to support this view. It is also possible that the condition in Theorem 14 is currently too strong. Other possible conditions were discussed in Section 3.5 of Chapter 3, but it remains on open question to see if a weaker condition would be sufficient to establish a proof of Theorem 14.

Chapter 6

Single-site Systematic Scan for Bipartite Graphs

In this chapter we study the mixing time of a systematic scan that makes single-site updates. We take advantage of the fact that the underlying graph is bipartite by fixing the scan order such that each site in the first colour class is updated before updating the sites in the other colour class.

6.1 Preliminaries

Let $G = (V, E)$ be any bipartite graph with maximum degree Δ. We denote the colour classes of G by $L(V)$ and $R(V)$. Let $C = \{1, \cdots, q\}$ be the set of colours and Ω be the set of proper q-colourings of G. Recall from Chapter 2 that $\mathcal{M}_{\text{LR}}$ is the systematic scan Markov chain which makes the following transitions

1. for each $v \in L(V)$ make a Metropolis move on site v

2. for each $v \in R(V)$ make a Metropolis move on site v.

Recall from Example 11 that a single-site Metropolis move on site v (and given a configuration x) is made by selecting a colour c uniformly at random from C and recolouring site v with colour c. Let x' be the configuration obtained from x by recolouring site v to c. If no edge containing v is monochromatic in x' then the resulting configuration of the Metropolis move is x', otherwise output of the Metropolis move is configuration x. Finally we remind the reader that each site in $L(V)$ is assigned weight $\omega_l = q^3 - 4$ and each site in $R(V)$ is assigned weight $\omega_r = 2\omega_l - 4$. We will prove Theorem 30 namely

Theorem 30. *Let G be any bipartite graph with maximum vertex-degree $\Delta \geq 3$. Consider the systematic scan Markov chain $\mathcal{M}_{\mathrm{LR}}$ on the state space Ω. Let $\gamma = \omega_r\left(1+\frac{1}{q^3}\right) - \frac{\Delta\omega_l}{q} - \frac{\Delta\omega_r}{q} - \frac{\Delta^2\omega_r}{q^2}$ where $\omega_l = q^3 - 4$ and $\omega_r = 2\omega_l - 4$. If $q \geq 2\Delta$ then $\gamma > 0$ and the mixing time of $\mathcal{M}_{\mathrm{LR}}$ is*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{LR}}, \varepsilon) \leq \frac{\omega_r \log(n\omega_r\varepsilon^{-1})}{\gamma}$$

scans.

As a final piece of notation we let x^j denote the configuration obtained by one partial scan of $\mathcal{M}_{\mathrm{LR}}$ (starting from configuration x) where site j is the next site to be updated. For a configuration x^j and a colour c let $(x^j \uparrow c)$ be the configuration obtained from the following two step process. Let σ be the configuration obtained from x^j by assigning colour c to site j. If no edge containing site j is monochromatic in σ then $(x^j \uparrow c) = \sigma$, otherwise $(x^j \uparrow c) = x^j$. The reason for introducing this notation is that a Metropolis move on site j can now be formulated as follows. Select a colour $c \in C$ uniformly at random and let $x^{j+1} = (x^j \uparrow c)$.

Our method of proof will be path coupling using weighted Hamming distance as the metric.

6.2 Definition of the Coupling

We begin by defining the coupling that we will use in the proof. We define the coupling for pairs of configurations $(x, y) \in S_i$ which differ only on the colour assigned to site i. We consider the update of a site j.

When it is time to update site j it is possible that more than one site is coloured differently in x^j and y^j due to previous updates that have been made in the scan. Suppose that j has k neighbour sites which are assigned different colours in x^j and y^j. Let these sites be denoted by $j^1, \ldots, j^k$. Note that if $k = 0$ then we can couple the configuration $(x^j \uparrow c)$ with $(y^j \uparrow c)$ for each $c \in C$ which ensures that $x'_j = y'_j$. Similarly if $x^j_j \neq y^j_j$ (which is only the case when $i = j$) we also couple the choice $(x^j \uparrow c)$ with $(y^j \uparrow c)$ for each $c \in C$ which may cause site i to become a discrepancy. Otherwise $x^j_j = y^j_j$ and we construct the coupling as follows. For each site $j^{k'}$ where $k' \in \{1, \ldots, k\}$ make the following choices:

- If

$$x^j_{j^{k'}} \notin \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

and

$$y^j_{j^{k'}} \notin \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

then couple the choice of $(x^j \uparrow x^j_{j^{k'}}) = x^j$ with the choice $(y^j \uparrow y^j_{j^{k'}}) = y^j$ in order to ensure that site j is assigned the same colour in both x' and y'. Also couple the choice $(x^j \uparrow y^j_{j^{k'}})$ with the choice $(y^j \uparrow x^j_{j^{k'}})$ which may cause site j to be coloured differently in x' and y'.

- If

$$x^j_{j^{k'}} \notin \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

and

$$y^j_{j^{k'}} \in \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

then couple the choice $(x^j \uparrow x^j_{j^{k'}})$ with the choice $(y^j \uparrow x^j_{j^{k'}})$. This may cause site j to be coloured differently in x' and y'.

- If

$$x^j_{j^{k'}} \in \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

and

$$y^j_{j^{k'}} \notin \{x^j_{j^1}, \ldots, x^j_{j^{k'-1}}, y^j_{j^1}, \ldots, y^j_{j^{k'-1}}\}$$

then couple the choice $(x^j \uparrow y^j_{j^{k'}})$ with the choice $(j^j \uparrow y^j_{j^{k'}})$. This may cause site j to be coloured differently in x' and y'.

For any remaining colours $c \in C \setminus \{x^j_{j^1}, \ldots, x^j_{j^k}, y^j_{j^1}, \ldots, y^j_{j^k}\}$, couple the choice $(x^j \uparrow c)$ with the choice $(y^j \uparrow c)$ which ensures that the same colour is assigned to site j in x' and y'. This completes the coupling construction since each colour $c \in C$ has been used exactly once.

By construction of the coupling, the marginal distribution is correct since each colour is used exactly once in both $(x \uparrow \cdot)$ and $(y \uparrow \cdot)$. We now state and prove an upper bound on the probability of a site which is coloured the same in x and y receiving different colours in x' and y' obtained from one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from $(x, y) \in S_i$.

Lemma 76. *Suppose that $(x, y) \in S_i$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). Let $b(j)$ be the number of sites adjacent to site j that are coloured differently in x^j and y^j. Then for any $j \neq i$*

$$\Pr(x'_j \neq y'_j \mid b(j) = k) \leq \frac{k}{q}.$$

Proof. In the construction of the coupling each site which is coloured differently in x^j and y^j is considered exactly once and will match one of the three stated cases. Each of these cases will produce at most one assignment of a colour to j in each copy such that $x'_j \neq y'_j$. There are k such sites and thus at at most k such choices will exist in the joint distribution, each being selected with probability $1/q$. Hence the probability of site j being coloured differently in x' and y' is at most $\frac{k}{q}$. □

6.3 Proof of Mixing

We first consider the case when the original discrepancy is in the left colour class of G, and hence the site containing the original discrepancy is updated before it can percolate to any of its neighbour sites.

Lemma 77. *Suppose that $(x, y) \in S_i$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). If $i \in L(V)$ then*

$$\mathbf{E}\left[\mathrm{Ham}(x', y')\right] \leq \left(1 - \frac{\beta}{\omega_l}\right) \mathrm{Ham}(x, y)$$

where

$$\beta = \omega_l - \frac{\Delta}{q}\left(\omega_l + \frac{\Delta\omega_r}{q}\right).$$

In particular, when $q \geq 2\Delta$ then $\beta > 0$.

Proof. We begin by showing that if $i \in L(V)$ then $\mathbf{E}\left[\mathrm{Ham}(x', y')\right] \leq \frac{\Delta}{q}\left(\omega_l + \frac{\Delta\omega_r}{q}\right)$. Since all sites in $L(V)$ are updated before $R(V)$, site i will be updated before any of its neighbours and hence $(x^i, y^i) \in S_i$. Since site i has at most Δ neighbours it will be coloured differently in x' and y' with probability at most Δ/q and contribute ω_l to the weighted Hamming distance.

Suppose that site i is coloured differently in each copy when the sites in $R(V)$ are being updated. Then each of i's neighbour sites will be coloured differently in x' and y' with probability at most $1/q$ by Lemma 76 and each will contribute with weight ω_r to the weighted Hamming distance. Adding it up we get the stated bound on the expectation since site i has at most Δ neighbours.

The statement of the lemma now follows since $i \in L(V)$ implies $\mathrm{Ham}(x, y) = \omega_l$ and using the assumption $q \geq 2\Delta$ gives

$$\beta \geq \omega_l - \frac{\Delta}{2\Delta}\left(\omega_l + \frac{\Delta\omega_r}{2\Delta}\right) = 1 > 0$$

by substituting the definition of ω_r, which completes the proof. □

We now consider the case when the initial discrepancy is in the colour class $R(V)$, and hence the discrepancy can percolate to the neighbours of site i before i is updated.

Lemma 78. *Suppose that $(x, y) \in S_i$ and that $i \in R(V)$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). If $j \neq i \in R(V)$ and d is the number of sites in $L(V)$ adjacent to both i and j then*

$$\Pr(x'_j \neq y'_j) \leq \frac{d}{q^2}.$$

Proof. Let $A(j)$ be the random variable denoting the number of sites adjacent to j that are coloured differently in x^j and y^j. Note from the statement of the lemma that it most hold that $A(j) \leq d$. From the definition of conditional probability we have

$$\Pr(x'_j \neq y'_j) = \sum_{k=0}^{d} \Pr\left(A(j) = k\right) \Pr(x'_j \neq y'_j \mid A(j) = k).$$

From Lemma 76 we have $\Pr(x'_j \neq y'_j \mid A(j) = k) \leq \frac{k}{q}$ so

$$\Pr(x'_j \neq y'_j) \leq \sum_{k=0}^{d} \Pr\left(A(j) = k\right) \frac{k}{q} = \frac{1}{q} \mathbf{E}\left[A(j)\right].$$

For each $l \in \{1, \ldots, d\}$ let I_l be the indicator random variable denoting the event: $x^j_l \neq y^j_l$ and $p_l = \Pr(x^j_l \neq y^j_l)$ be the probability of that event occurring. Using linearity of expectation

$$\Pr(x'_j \neq y'_j) \leq \frac{1}{q} \mathbf{E}\left[\sum_{l=1}^{d} I_l\right] = \frac{1}{q} \sum_{l=1}^{d} \mathbf{E}\left[I_l\right] = \frac{1}{q} \sum_{l=1}^{d} p_l.$$

From Lemma 76 we obtain $p_l \leq \frac{1}{q}$ for $l \in \{1, \ldots, d\}$ since site i is the only site adjacent to l that is coloured differently in x^l and y^l. Thus,

$$\Pr(x'_j \neq y'_j) \leq \frac{1}{q} \sum_{l=1}^{d} \frac{1}{q} = \frac{d}{q^2}$$

which completes the proof. □

Lemma 79. *Suppose that $(x, y) \in S_i$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). If $i \in R(V)$ then*

$$\mathbf{E}\left[\mathrm{Ham}(x', y')\right] - \omega_r \Pr(x'_i \neq y'_i) \leq \frac{\Delta\omega_l}{q} + \frac{\Delta(\Delta-1)\omega_r}{q^2}.$$

Proof. From Lemma 76 we know that the expected number of additional discrepancies in $L(V)$ is at most Δ/q since site i has at most Δ neighbour sites, each of which will be coloured differently in each copy with probability at most $1/q$. Each of those sites has weight ω_l.

To upper bound the expected number of additional discrepancies in $R(V)$ we need to upper bound the number of sites in $L(V)$ adjacent to both i and some $j \neq i \in R(V)$. We let $d(v, u)$ denote the minimum distance (number of edges) between site u and v in G, and $u \rightharpoonup v$ the existence of an edge between u and v. The sum over all $j \neq i \in R(V)$ of the number of sites adjacent to both i and j is thus

$$\begin{aligned}\sum_{\substack{j\in V\\ d(i,j)=2}} \sum_{\substack{k\in V\\ k\rightharpoonup j\\ k\rightharpoonup i}} 1 &= \sum_{\substack{k\in V\\ k\rightharpoonup i}} \sum_{\substack{j\in V\\ j\neq i\\ j\rightharpoonup k}} 1 \\ &\leq \sum_{\substack{k\in V\\ k\rightharpoonup i}} (\Delta - 1) \\ &\leq \Delta(\Delta-1).\end{aligned}$$

Combining this bound with Lemma 78 we have, by linearity of expectation, that the expected number of additional disagreements in $R(V)$ is at most $\frac{(\Delta-1)\Delta}{q^2}$ each of which has weight ω_r. □

We now need to upper bound the probability of site i being coloured differently in x' and y'. To that end we introduce the following terminology.

Definition 80 (Colour compatibility)**.** *Let $N(v)$ be the set of sites adjacent to a site v, and let*

$$C(v) = C \setminus \bigcup_{v' \in N(v)} \{x_{v'}, y_{v'}\}$$

be the set of colours not adjacent to v. Two distinct sites k and l are said to be 'colour compatible' if $C(k) \cap C(l) \neq \emptyset$.

Lemma 81. *Suppose $(x, y) \in S_i$ for some $i \in R(V)$. Let $N(v)$ be the set of sites adjacent to a site v. If $\deg i = \Delta$ and $q \geq \Delta + 3$ then there are two distinct sites $v_k \in N(i)$ and $v_l \in N(i)$ which are colour compatible.*

Proof. For ease of reference, let $N(i) = \{v_1, \ldots, v_\Delta\}$ and also let $c(v)$ be the size of the set $C(v)$. Each site $v \in N(i)$ has at most Δ neighbours. Since site i is the only site that contributes two colours to $C \setminus C(v)$ it holds that

$$c(v) \geq q - (\Delta - 1) - 2 = q - \Delta - 1 \tag{6.1}$$

for every $v \in N(i)$. We need to show the existence of two distinct sites v_k and v_l that are colour compatible. We will do this by contradiction. Suppose that no two sites in $N(i)$ are colour compatible. Then

$$C(v_k) \subseteq C \setminus \bigcup_{\substack{v_l \in N(i) \\ l<k}} C(v_l) \tag{6.2}$$

for all $k \in \{1, \ldots, \Delta\}$ since otherwise some site $v_l \in \{v_1, \ldots, v_{k-1}\}$ would be colour compatible with site v_k. By (6.2), $C(v_k)$ cannot contain any of the colours in

$$\bigcup_{\substack{v_l \in N(i) \\ l<k}} C(v_l).$$

Also, it cannot contain x_i or y_i so

$$c(v_k) = q - \sum_{0<l<k} c(v_l) - 2 \leq q - (k-1)(q - \Delta - 1) - 2$$

by (6.1). Hence

$$q - \Delta - 1 \leq c(v_\Delta) \leq q - (\Delta - 1)(q - \Delta - 1) - 2$$

where the lower bound is from (6.1).

When $\Delta \geq 3$ it follows that $q \leq \Delta + 2$ which contradicts our assumption that $q \geq \Delta + 3$. Hence there must be a pair of colour compatible sites in $N(i)$. □

Lemma 82. *Suppose $(x, y) \in S_i$ for some $i \in R(V)$. Let $N(i)$ be the set of sites adjacent to site i. If $\deg i = \Delta$ and $q \geq \Delta + 3$ then there are two sites $k \in N(i)$ and $l \in N(i)$ such that*

$$\Pr\left(x_k^i = x_l^i = y_k^i = y_l^i\right) \geq \frac{1}{q^2}.$$

Proof. By Lemma 81 there exist two distinct sites k and l in $N(i)$ that are colour compatible. Since k and l are colour compatible there is at least one colour c that will be accepted when updating both sites k and l. With probability at least $1/q$, colour c will be selected and accepted in the recolouring of site k and independently with probability at least $1/q$ in the recolouring of site l. □

Lemma 83. *Suppose that $(x, y) \in S_i$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). If $i \in R(V)$ then*

$$\Pr(x'_i \neq y'_i) \leq \frac{\Delta(q+1)}{q^2} - \frac{1}{q^3}.$$

Proof. Let $A(i)$ be the random variable denoting the number of sites in $N(i)$ that are assigned different colours in configuration x^i and y^i. Note from the statement of the lemma that it most hold that $A(i) \leq \Delta$. From the definition of conditional probability we have

$$\Pr(x'_i \neq y'_i) = \sum_{k=0}^{\Delta} \Pr(A(i) = k) \Pr(x'_i \neq y'_i \mid A(i) = k).$$

We consider the two cases $\deg i = \Delta$ and $\deg i \leq \Delta - 1$ separately. First suppose $\deg i \leq \Delta - 1$. If there are k sites adjacent to i that are assigned different colours x^i and y^i then there can be at most $\Delta - 1 + k$ different colours adjacent to site i. Hence $\Pr(x'_i \neq y'_i \mid A(i) = k) \leq \frac{\Delta - 1 + k}{q}$ which gives

$$\begin{aligned}
\Pr(x'_i \neq y'_i) &\leq \sum_{k=0}^{\Delta-1} \Pr(A(i) = k) \frac{\Delta - 1 + k}{q} \\
&= \frac{\Delta - 1}{q} \sum_{k=0}^{\Delta-1} \Pr(A(i) = k) + \frac{1}{q} \mathbf{E}\left[A(i)\right] \\
&= \frac{\Delta - 1}{q} + \frac{1}{q} \mathbf{E}\left[A(i)\right]
\end{aligned}$$

by definition of probability spaces.

Now for each $l \in \{1, \ldots, \deg i\}$ let I_l be the indicator random variable denoting the event: $x^i_l \neq y^i_l$. Also let $p_l = \Pr(x^i_l \neq y^i_l)$ be the probability that the event

occurs. Using linearity of expectation we have

$$\begin{aligned}\Pr(x'_i \neq y'_i) &\leq \frac{\Delta - 1}{q} + \frac{1}{q}\mathbf{E}\left[\sum_{l=1}^{\deg i} I_l\right] \\ &= \frac{\Delta - 1}{q} + \frac{1}{q}\sum_{l=1}^{\deg i}\mathbf{E}\left[I_l\right] \\ &= \frac{\Delta - 1}{q} + \frac{1}{q}\sum_{l=1}^{\deg i} p_l.\end{aligned}$$

From Lemma 76 we have $p_l \leq 1/q$ for $l \in \{1, \ldots, \deg i\}$ since site i is the only site adjacent to l in that can be coloured differently in x^l and y^l. Thus,

$$\begin{aligned}\Pr(x'_i \neq y'_i) &\leq \frac{\Delta - 1}{q} + \frac{1}{q}\sum_{l=1}^{\deg i}\frac{1}{q} \\ &= \frac{\Delta - 1}{q} + \frac{\deg i}{q^2} \\ &\leq \frac{(\Delta - 1)(q+1)}{q^2}.\end{aligned}$$

Now consider the case when $\deg i = \Delta$. As before, define $N(i)$ as the set of sites adjacent to i. Let E_i be shorthand for the following event: There exists two distinct sites $a \in N(i)$ and $b \in N(i)$ such that $x^i_a = x^i_b = y^i_a = y^i_b$. If there are k sites adjacent to i that are assigned different colours in x^i and y^i then there can be at most $\Delta + k$ different colours adjacent to site i. However, if E_i holds there can be at most $\Delta + k - 1$ different colours adjacent to i since two sites are known to have the same colour. Hence

$$\begin{aligned}\Pr(x'_i \neq y'_i \mid A(i) = k) \leq{}& \Pr\left(E_i \mid A(i) = k\right)\frac{\Delta + k - 1}{q} \\ &+ \left(1 - \Pr\left(E_i \mid A(i) = k\right)\right)\frac{\Delta + k}{q} \\ ={}& \frac{\Delta + k}{q} - \Pr\left(E_i \mid A(i) = k\right)\frac{1}{q}.\end{aligned}$$

Proceeding as in the previous case

$$\begin{aligned}
\Pr(x'_i \neq y'_i) &\leq \sum_{k=0}^{\Delta} \Pr(A(i)=k)\left(\frac{\Delta+k}{q} - \Pr(E_i \mid A(i)=k)\frac{1}{q}\right) \\
&= \sum_{k=0}^{\Delta} \Pr(A(i)=k)\frac{\Delta+k}{q} - \sum_{k=0}^{\Delta} \Pr(A(i)=k)\Pr(E_i \mid A(i)=k)\frac{1}{q} \\
&= \frac{\Delta}{q}\sum_{k=0}^{\Delta} \Pr(A(i)=k) + \frac{1}{q}\mathbf{E}\left[A(i)\right] \\
&\quad - \frac{1}{q}\sum_{k=0}^{\Delta} \Pr(A(i)=k)\Pr(E_i \mid A(i)=k) \\
&= \frac{\Delta}{q} + \frac{1}{q}\mathbf{E}\left[A(i)\right] - \frac{1}{q}\Pr(E_i)
\end{aligned}$$

by definition of probability spaces and observing that $0 \leq A(i) \leq \Delta$.

Again let I_l be the indicator random variable denoting the event: $x^i_l \neq y^i_l$ defined for each $l \in \{1, \ldots, \Delta\}$. Also let $p_l = \Pr(x^i_l \neq y^i_l)$ be the probability that the event occurs. Using linearity of expectation

$$\begin{aligned}
\Pr(x'_i \neq y'_i) &\leq \frac{\Delta}{q} + \frac{1}{q}\mathbf{E}\left[\sum_{l=1}^{\Delta} I_l\right] - \frac{1}{q}\Pr(E_i) \\
&= \frac{\Delta}{q} + \frac{1}{q}\sum_{l=1}^{\Delta} \mathbf{E}\left[I_l\right] - \frac{1}{q}\Pr(E_i) \\
&= \frac{\Delta}{q} + \frac{1}{q}\sum_{l=1}^{\Delta} p_l - \frac{1}{q}\Pr(E_i).
\end{aligned}$$

From Lemma 76 we have $p_l \leq 1/q$ for $l \in \{1, \ldots, \Delta\}$ since site i is the only site adjacent to l in that can be coloured differently in x^l and y^l. Thus,

$$\begin{aligned}
\Pr(x'_i \neq y'_i) &\leq \frac{\Delta}{q} + \frac{1}{q}\sum_{l=1}^{\Delta} \frac{1}{q} - \frac{1}{q}\Pr(E_i) \\
&= \frac{\Delta}{q} + \frac{\Delta}{q^2} - \frac{1}{q}\Pr(E_i) \\
&= \frac{\Delta(q+1)}{q^2} - \frac{1}{q}\Pr(E_i)
\end{aligned}$$

and so

$$\begin{aligned}\Pr(x'_i \neq y'_i) &\leq \max\left(\frac{(\Delta-1)(q+1)}{q^2}, \frac{\Delta(q+1)}{q^2} - \frac{1}{q}\Pr(E_i)\right)\\ &= \frac{\Delta(q+1)}{q^2} - \frac{1}{q}\Pr(E_i)\end{aligned}$$

which can be verified by letting $\Pr(E_i) \leq 1$. Finally we have $\Pr(E_i) \geq 1/q^2$ from Lemma 82 which completes the proof. □

Lemma 84. *Suppose that $(x, y) \in S_i$ and $\Delta \geq 3$. Obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). If $i \in R(V)$ then*

$$\mathbf{E}\left[\mathrm{Ham}(x', y')\right] \leq \left(1 - \frac{\gamma}{\omega_r}\right)\mathrm{Ham}(x, y)$$

where

$$\gamma = \omega_r\left(1 + \frac{1}{q^3}\right) - \frac{\Delta\omega_l}{q} - \frac{\Delta\omega_r}{q} - \frac{\Delta^2\omega_r}{q^2}.$$

In particular, when $q \geq 2\Delta$ then $\gamma \geq 1 - \frac{12}{q^3} > 0$.

Proof. From Lemmas 79 and 83 we have

$$\begin{aligned}\mathbf{E}\left[\mathrm{Ham}(x', y')\right] &\leq \frac{\Delta\omega_l}{q} + \frac{\omega_r(\Delta-1)\Delta}{q^2} + \omega_r\left(\frac{\Delta(q+1)}{q^2} - \frac{1}{q^3}\right)\\ &= \frac{\Delta\omega_l}{q} + \frac{\Delta\omega_r}{q} + \frac{\Delta^2\omega_r}{q^2} - \frac{\omega_r}{q^3}\end{aligned}$$

by expanding and simplifying.

Given that $i \in R(V)$ we have $\mathrm{Ham}(x, y) = \omega_r$ which implies the definition of γ in the statement of the lemma. Finally use the assumption $q \geq 2\Delta$ to verify

$$\begin{aligned}\gamma &\geq \omega_r\left(1 + \frac{1}{q^3}\right) - \frac{\omega_l}{2} - \frac{\omega_r}{2} - \frac{\omega_r}{4}\\ &= \frac{2\omega_l - q^3 - 4}{q^3}\\ &= 1 - \frac{12}{q^3} > 0\end{aligned}$$

using $\omega_r = 2\omega_l - 4$ and $\omega_l = q^3 - 4$ and the fact that $q \geq 6$. □

This enables us to prove Theorem 30.

Proof of Theorem 30. Suppose that $(x, y) \in S_i$ and obtain a pair of configurations (x', y') by one complete scan of $\mathcal{M}_{\mathrm{LR}}$ starting from (x, y). From Lemmas 77 and 84

we have

$$\begin{aligned}\mathbf{E}\left[\mathrm{Ham}(x',y')\right] &\leq \max\left(\mathrm{Ham}(x,y)\left(1-\frac{\gamma}{\omega_r}\right),\mathrm{Ham}(x,y)\left(1-\frac{\beta}{\omega_l}\right)\right)\\ &= \mathrm{Ham}(x,y)\left(1-\min\left(\frac{\gamma}{\omega_r},\frac{\beta}{\omega_l}\right)\right)\\ &= \mathrm{Ham}(x,y)\left(1-\frac{\gamma}{\omega_r}\right) < \mathrm{Ham}(x,y)\end{aligned}$$

since $\gamma > 0$ from Lemma 84. The claim that $\min\left(\frac{\gamma}{\omega_r},\frac{\beta}{\omega_l}\right) = \frac{\gamma}{\omega_r}$ can be verified since $3 \leq \Delta \leq \frac{q}{2}$. This bound on the expected value of $\mathrm{Ham}(x',y')$ implies Theorem 30 by Corollary 9 (path coupling). □

Chapter 7

Conclusion

This thesis has been concerned with analysing the mixing time of systematic scan Markov chains. We conclude by summarising the contributions this thesis has made to the field of computer science as well as highlighting some open problems it poses.

Summary of Contributions

The main contribution of this thesis has been the introduction of a new technique for bounding the mixing time of systematic scan Markov chains using block dynamics. This technique involves providing sufficiently good upper bounds on an influence parameter which represents the maximum influence *on* a site of the underlying graph. These bounds are derived through the construction of a coupling for the update of each block starting from two configurations that are identical except for on the spin assigned to a single site. This influence parameter is often simple to compute for a given coupling. If the maximum influence on a site is sufficiently small then the systematic scan Markov chain mixes in $O(\log n)$ scans. This new technique has the immediate advantage that one is not required to keep track of intermediate states of the chain during the analysis. A further advantage that this proof technique has is that it implies rapid mixing of systematic scan for *any order* of the given set of blocks whereas path coupling is specific to the stated order. The new proof technique is based on a known technique called Dobrushin uniqueness and the main result (Theorem 14) was proved in Chapter 3. The condition on the influence parameter that needs to be satisfied when applying this technique is a generalisation of a similar condition that applies to single-site dynamics as we discussed in Section 3.5 of Chapter 3.

We have presented several applications of this technique in this thesis and

they all reduce the gap between the parameters that imply mixing for random update and systematic scan. In summary these applications were

- systematic scan for sampling from the uniform distribution on proper q-colourings of general graphs with maximum vertex-degree Δ mixes in $O(\log n)$ scans whenever $q \geq 2\Delta$ (Theorem 16 which was proved in Chapter 3),
- systematic scan for sampling from the uniform distribution on proper q-colourings of a height-H tree mixes in $O(H)$ scans whenever $q > \Delta + 2\sqrt{\Delta - 1}$ in the single-site case and in fewer colours using some suitable block dynamics (Theorems 18 and 20 which were proved in Chapter 3),
- systematic scan for sampling from the uniform distribution on proper 7-colourings the grid mixes in $O(\log n)$ scans (Theorem 28 which was proved in Chapter 5), and
- systematic scan for sampling from the uniform distribution on H-colourings of the n-vertex path mixes in $O(\log n)$ scans whenever H has a 2-edge path between all vertices (Theorem 22 which was proved in Chapter 4).

We have also used path coupling in some cases when the underlying graph of the spin system could help to facilitate an analysis. In summary these results were

- a systematic scan Markov chain for sampling for the uniform distribution of H-colourings of the n-vertex path mixes in $O(\log n)$ scans for any fixed H (Theorem 24 proved in Chapter 4), and
- a single-site systematic scan Markov chain for sampling from the uniform distribution of proper q-colourings of a general bipartite graphs with maximum vertex-degree Δ mixes in $O(\log n)$ scans whenever $q \geq 2\Delta$ (Theorem 30 proved in Chapter 6).

A determining factor that helped significantly to facilitate the coupling analysis of the two systematic scan Markov chains mentioned above was the structure of the underlying graph. In the case of the systematic scan for sampling H-colourings of the path, the fact that the underlying graph is a path clearly makes it more feasible to keep track of any discrepancies that percolate during each individual scan. In the case of proper q-colourings of bipartite graphs we were able to scan each colour class of the underlying graph separately which significantly limited the set of sites that could potentially have become discrepancies during one scan.

Finally the results for sampling H-colourings of the path using systematic scan created a temporary gap between the parameters required for mixing of systematic scan and random update. This gap was closed by the following result about a random update Markov chain which was included for completeness

- a random update Markov chain for sampling for the uniform distribution of H-colourings of the n-vertex path mixes in $O(n \log n)$ block updates for any fixed H (Theorem 26 proved in Chapter 4).

Open Problems

Despite the improvements in the parameters that imply mixing of systematic scan for various spin systems presented in this thesis, the gap between the parameters sufficient for mixing of systematic scan and random update still persists (although in many cases the gap is now somewhat reduced). For example, in the case when the spin system correspond to proper q-colourings of a general graph with maximum vertex-degree Δ then the condition $q \geq (11/6)\Delta$ is sufficient for rapid mixing of a random update Markov chain (Vigoda [53]) whereas the corresponding condition required for rapid mixing of systematic scan is $q \geq 2\Delta$ (Theorem 16). Similar gaps also exist for special graphs such as trees or the grid and it is of general interest to either close those gaps or to show that systematic scan does not mix under the same conditions as random update. The possibility of the latter, namely that systematic scan does not mix under the same conditions as random update, is however unlikely. Currently the only types of examples where there is a genuine difference between the mixing properties of systematic scan and random update is the relatively uninteresting case when the spin system corresponds to proper colourings of a graph with no edges (where random update requires $\Omega(n \log n)$ updates but systematic scan mixes in one scan) or contrived examples such as the spin system in Observation 53 (where random update mixes rapidly but systematic scan does not mix at all).

Another open problem that arises from the work presented in this thesis is whether the condition required for mixing in Theorem 14 is too strong. The possibility of using other conditions was explored to some depth in Section 3.5 of Chapter 3, however it remains possible that a weaker condition on the influence on a site could be sufficient to prove rapid mixing of systematic scan. Note that it may be possible to develop conditions that hold for certain spin systems such as proper q-colourings but not for general spin systems, and such conditions

would also be of interest. A final open problem related to Theorem 16, which was also raised in Chapter 3, is whether it is possible to find a general method for obtaining a set of weights that would make the influence *on* a site sufficiently small provided that the influence *of* a site is small. This would be a generalisation of the matrix balancing in the single-site case as we have previously discussed. Note that we do rule out the possibility of finding such a set of weights when using a natural definition of "the influence *of* a site" that is similar to the path coupling condition. None the less, it remains possible that a stronger definition of "the influence *of* a site" would make it possible to find such a set of weights (see Observation 54 and the remark following it).

Bibliography

[1] Dimitris Achlioptas, Mike Molloy, Cristopher Moore, and Frank van Bussel. Sampling grid colourings with fewer colours. In *LATIN*, pages 80–89, 2004.

[2] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de probabilités XVII*, pages 243–297. Springer-Verlag, 1983.

[3] David Aldous and James Fill. Reversible Markov chains and random walks on graphs. http://www.stat.berkeley.edu/users/aldous/RWG/book.html.

[4] Magnus Bordewich, Martin Dyer, and Marek Karpinski. Stopping times, metrics and approximate counting. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP*, volume 4051 of *Lecture Notes in Computer Science*, pages 108–119. Springer, 2006.

[5] Russ Bubley and Martin Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *FOCS*, pages 223–231. IEEE Computer Society, 1997.

[6] Russ Bubley, Martin Dyer, and Catherine S. Greenhill. Beating the 2Δ bound for approximately counting colourings: A computer-assisted proof of rapid mixing. In *SODA*, pages 355–363. ACM/SIAM, 1998.

[7] Robert Burton and Jeffrey Steif. Nonuniqueness of measures of maximal entropy for subshifts of finite type. *Ergodic Theory and Dynamical Systems*, 14(2):213–236, 1994.

[8] Colin Cooper, Martin Dyer, and Alan Frieze. On Markov chains for randomly H-colouring a graph. *Journal of Algorithms*, 39(1):117–134, 2001.

[9] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of The American Statistical Association*, 91(434):883–904, 1996.

[10] Mary Kathryn Cowles, Gareth O. Roberts, and Jeffrey S. Rosenthal. Possible biases induced by MCMC convergence diagnostics. *Journal of Statistical Computation and Simulation*, 64:87–104, 1999.

[11] Persi Diaconis and Arun Ram. Analysis of systematic scan Metropolis algorithms using Iwahoti-Hecke algebra techniques. *Michigan Mathematical Journal*, 48:157–190, 2000.

[12] Roland Lvovich Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability and Its Applications*, 15:458–486, 1970.

[13] Roland Lvovich Dobrushin and Senya B. Shlosman. Constructive criterion for the uniqueness of Gibbs field. In Jozsef Fritz, Arthur Jaffe, and Domokos Szasz, editors, *Statistical mechanics and dynamical systems*, volume 10 of *Progress in Physics*, pages 371–403. Birkhäuser, Boston, 1985.

[14] Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM Journal on Computing*, 31(5):1527–1541, 2002.

[15] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. On the relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2003.

[16] Martin Dyer, Leslie Ann Goldberg, Catherine S. Greenhill, Mark Jerrum, and Michael Mitzenmacher. An extension of path coupling and its application to the glauber dynamics for graph colourings. *SIAM Journal on Computing*, 30(6):1962–1975, 2001.

[17] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Counting and sampling H-colourings. *Information and Computation*, 189:1–16, 2004.

[18] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Dobrushin conditions and systematic scan. In Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, editors, *APPROX-RANDOM*, volume 4110 of *Lecture Notes in Computer Science*, pages 327–338. Springer, 2006.

[19] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. arXiv:math.PR/0702744 (submitted), 2006.

[20] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Systematic scan and sampling colourings. *Annals of Applied Probability*, 16(1):185–230, 2006.

[21] Martin Dyer, Leslie Ann Goldberg, Mark Jerrum, and Russell Martin. Markov chain comparison. *Probability Surveys*, 3:89–111, 2006.

[22] Martin Dyer and Catherine Greenhill. A more rapidly mixing Markov chain for graph colourings. *Random Structures and Algorithms*, 13:285–317, 1998.

[23] Martin Dyer and Catherine S. Greenhill. Random walks on combinatorial objects. In J. D. Lamb and D. A. Preece, editors, *Surveys in Combinatorics*, volume 267 of *London Mathematical Society Lecture Notes Series*, pages 101–136. Cambridge University Press, 1999.

[24] Martin Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms. *Random Structures and Algorithms*, 17:260–289, 2000.

[25] Martin Dyer and Catherine S. Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.

[26] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures and Algorithms*, 24(4):461–479, 2004.

[27] George S. Fishman. Coordinate selection rules for gibbs sampling. *The Annals of Applied Probability*, 6(2):444–465, 1996.

[28] Hans Föllmer. A covariance estimate for Gibbs measures. *Journal of Functional Analysis*, 46(3):387–395, 1982.

[29] Anna Galluccio, Pavol Hell, and Jaroslav Nešetřil. The complexity of H-colouring of bounded degree graphs. *Discrete Mathematics*, 222:101–109, 2000.

[30] Hans-Otto Georgii. *Gibbs Measures And Phase Transitions*. de Gruyter Studies in Mathematics 9. Walter de Gruyter & Co, 1998.

[31] Leslie Ann Goldberg, Markus Jalsenius, Russell Martin, and Mike Paterson. Improved mixing bounds for the anti-ferromagnetic potts model on $\mathbb{Z}^2$. *LMS Journal of Computation and Mathematics*, 9:1–20, 2006.

[32] Leslie Ann Goldberg, Steven Kelk, and Mike Paterson. The complexity of choosing an H-colouring (nearly) uniformly at random. *SIAM Journal on Computing*, 33(2):416–432, 2004.

[33] Leslie Ann Goldberg, Russell Martin, and Mike Paterson. Strong spatial mixing for lattice graphs with fewer colours. *SIAM Journal on Computing*, 35(2):486–517, 2005.

[34] Leslie Ann Goldberg, Russell Martin, and Mike Paterson. Random sampling of 3-colourings in $\mathbb{Z}^2$. *Random Structures and Algorithms*, 24(3):279–302, 2004.

[35] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, 3rd edition, 2001.

[36] Thomas P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *FOCS*, pages 39–46. IEEE Computer Society, 2006.

[37] Pavol Hell and Jaroslav Nešetřil. On the complexity of H-colouring. *Journal of Combinatorial Theory, Series B*, 48:92–110, 1990.

[38] Markus Jalsenius and Kasper Pedersen. A systematic scan for 7-colourings of the grid. arXiv:0704.1625 (submitted), 2007.

[39] Mark 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.

[40] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Birkhäuser, 2003.

[41] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[42] Frank P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 47:370–395, 1985.

[43] Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. In *FOCS*, pages 568–578. IEEE Computer Society, 2001.

[44] Michael Luby, Dana Randall, and Alistair Sinclair. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing*, 31:167–192, 2001.

[45] Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets: Part I. *Random Structures and Algorithms*, 15(3–4):229–241, 1999.

[46] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Glauber dynamics on trees: Boundary conditions and mixing time. *Communications in Mathematical Physics*, 250(2):301–334, 2004.

[47] Kasper Pedersen. Dobrushin conditions for systematic scan with block dynamics. In Luděk Kučera and Antonín Kučera, editors, *MFCS*, volume 4708 of *Lecture Notes in Computer Science*, pages 264–275. Springer, Berlin, 2007.

[48] Kasper Pedersen. On systematic scan for sampling H-colourings of the path. arXiv:0706.3794 (submitted), 2007.

[49] Gareth O. Roberts and Sujit K. Sahu. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):291–317, 1997.

[50] Jesus Salas and Alan D. Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86:551–579, 1997.

[51] Barry Simon. *The Statistical Mechanics of Lattice Gases*, volume I. Princeton University Press, 1993.

[52] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[53] Eric Vigoda. Improved bounds for sampling colourings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.

[54] Dror Weitz. *Mixing in Time and Space for Discrete Spin Systems*. PhD thesis, University of California, Berkley, 2004.

[55] Dror Weitz. Combinatorial criteria for uniqueness of Gibbs measures. *Random Structures and Algorithms*, 27(4):445–475, 2005.

[56] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149. IEEE Computer Society, 2006.

[57] Benjamin Widom and John S. Rowlinson. New model for the study of liquid-vapour phase transition. *The Journal of Chemical Physics*, 52(4):1670–1684, 1970.