Randomised algorithms for low temperature spin systems

James Stewart

Christ Church

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Abstract

A spin system is a general framework in which the vertices of a graph are assigned spins from a finite set. The local interactions between neighbouring spins give rise to a global weight, which describes the probability that the system is in the given configuration. Spin systems provide a framework for sampling and counting problems in computer science, graph homomorphism problems in combinatorics, and phase transition phenomena in statistical physics. Two natural computational problems associated to a spin system are counting (computing the aggregate weight of all spin configurations) and sampling (constructing a random spin configuration with probability proportional to its weight). In general, the problem of exact counting is computationally hard, and our attention therefore turns towards approximate counting, and the closely related problem of approximate sampling. In computer science, approximate counting and sampling for spin systems, such as the hard-core model on independent sets or the Potts model on colourings, are well-studied problems.

In recent years, on bounded-degree graph classes, an intimate connection has been established between the computational complexity of approximate sampling and counting, and the the existence of phase transitions in corresponding statistical physics models. Roughly speaking, there is a phase transition between a 'high-temperature' regime and a 'low temperature' regime. In the high-temperature regime, interactions between spins are weak and there is disorder in the model. Conversely, in the low temperature regime, interactions between spins are strong and the model is highly ordered – a typical configuration corresponds to one of some number of ground states, which limit the way that spins are assigned. This behaviour is reflected by the algorithmic properties of these models. At high-temperatures, there exists a variety of algorithmic techniques for approximate counting and sampling – examples include the Markov chain

Monte Carlo (MCMC) method and correlation decay. At low temperatures, these approaches no longer succeed. Indeed, for many spin systems on many interesting classes of graphs, the problems of approximate counting and sampling are provably intractable.

Recent research has sought to understand the algorithmic behaviour of spin systems such as the hard-core and Potts models on graphs such as the integer lattice and the random regular graph, the latter of which can be seen as an 'average case' instance of a regular graph. A detailed algorithmic and probabilistic understanding of these models has been developed, and has resulted in efficient low temperature algorithms. This is in contrast to the case of arbitrary bounded-degree graphs, where efficient algorithms can usually only be achieved in the high-temperature regime. These approaches typically use the truncated Taylor series approach of Barvinok to approximate the log partition function of a certain convergent series expansion. As such, they rely on tools from the complex analysis literature, and typically have running times of the form $n^{O(\log \Delta)}$, where Δ is the maximum degree of the input graph.

The work in this thesis can be seen as part of this same program – we develop efficient algorithmic approaches for approximate counting and sampling in low temperature spin systems, in a variety of different settings. We present a Markov chain based approach to approximate counting and sampling from spin systems such as the hard-core and Potts models. Using combinatorial and probabilistic techniques, we prove fast running times for these algorithms. Our results apply to classes of graphs such as the integer lattice, the random regular graph, and bounded-degree expanders, and they apply in a similar range of parameters to existing results. We also generalise these techniques to any spin system in which the interactions between spins are sufficiently strong, and to classes of random graphs of possibly unbounded degree.

Acknowledgements

First and foremost, I would like to thank my supervisors, Leslie Goldberg and Andreas Galanis, for their time, expertise, and guidance. I would also like to thank Zongchen Chen, Will Perkins, and Eric Vigoda, for an enjoyable collaboration. I thank Rahul Santhanam, Heng Guo, and Standa Živný, for serving as examiners for my various DPhil examinations, as well as my family and friends for their support.

Declaration

I declare that the work presented in this thesis is my own, except where otherwise stated. Parts of the work presented in this thesis also appear in the following publications. Chapters 2, 3, and 6 contain work from the following papers:

[22] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. *Random Struct. Algorithms*, 58(2):294–321, 2021. Theorems 5 and 6 from arxiv.org/abs/1901.0665

[21] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik

Chapters 2 and 4 contain work from the following papers:

[40] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. *ACM Transactions on Computation Theory (TOCT)*, 13(4):1–18, 2021

[39] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. In Javier Esparza

and Daniel KráĽ, editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik

Chapters 2 and 5 contain work from the following papers:

[42] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. *Information and Computation*, page 104894, 2022

[41] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. In Mary Wootters and Laura Sanità, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, volume 207 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:13, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik

Contents

List of Figures

List of notation

Chapter 1

$\mathcal{H}(\sigma)$ ..... Hamiltonian of Ising configuration σ p. 1

$Z_{G,\beta}$ ..... Partition function of Ising model p. 1

$\lambda_c(d)$ ..... Uniqueness threshold on infinite d-regular tree p. 3

#BIS ..... Counting independent sets in bipartite graphs p. 6

Chapter 2

$G = (V_G, E_G)$ ..... A graph G p. 13

$\deg_G(v)$ ..... Degree of vertex v in G p. 13

S^c ..... Complement of vertex set S in G p. 13

$\partial_G S$ ..... Vertex boundary of S in G p. 13

S^+ ..... Vertex set S and its boundary in G p. 13

$E_G(S,T)$ ..... Edges from S to T in G p. 13

$e_G(S,T)$ ..... Number of edges from S to T in G p. 13

$G = (V_G^0, V_G^1, E_G)$ ..... A bipartite graph G p. 13

NP ..... Non-deterministic polynomial time p. 14

#P ..... 'Sharp' P p. 14

Chapter 3

Chapter 4

Chapter 5

Chapter 7

Chapter 8

Chapter 1

Introduction

Spin systems are one of the primary objects of study in statistical physics, and are often used to model physical phenomena such as phase transitions. One canonical example of a spin system is the (ferromagnetic) Ising model, which was introduced and studied by Ising and Lenz in the 1920s [73, 63], and intended to model the physical phenomenon of ferromagnetism. The model consists of a random assignment of spins (which can be in one of two states, $+1$ or -1) to the vertices of a graph. Each spin interacts with its neighbours, and neighbouring spins which align have a lower energy than those which do not. The overall system prefers to assume a low energy configuration, and this preference is moderated by the temperature of the system. The Ising model therefore enables the study of phase transitions, where a small fluctuation in some parameter, such as the temperature, can give rise to a dramatic change in the large-scale state of a system, such as its magnetisation. Formally, the ferromagnetic Ising model at inverse temperature $\beta > 0$ on a graph $G = (V_G, E_G)$ is defined as follows. The energy of a spin configuration $\sigma \colon V_G \to \{+1, -1\}$ is given by its Hamiltonian $\mathcal{H}(\sigma) = \sum_{\{u,v\} \in E_G} \sigma_u \sigma_v$. The probability of a spin configuration $\sigma \colon V_G \to \{+1, -1\}$ is given by the Gibbs distribution $\mu_{G,\beta} \propto \exp\{-\beta \mathcal{H}(\sigma)\}$, where the normalising factor $Z_{G,\beta} = \sum_{\sigma \colon V_G \to \{+1,-1\}} \exp\{-\beta \mathcal{H}(\sigma)\}$ is known as the partition function.

In his 1963 paper [50], Glauber proposed a stochastic process for simulating the evolution of the Ising model. Now known as the (heat bath) Glauber dynamics, the process attempts to describe the dynamics by which the system approaches its equilibrium state over time. Starting from any initial configuration, we repeatedly obtain the next, as follows. Choose a spin uniformly

at random, then choose a new value for this spin with probability according to the marginal Gibbs distribution conditioned on the values of neighbouring spins. More generally, this process can be viewed as a Markov chain, whose stationary distribution is the Gibbs distribution of the Ising model. One question of central importance in theoretical computer science, statistical physics, and probability theory, is that of how long it takes for this process to converge towards its stationary distribution.

Markov chains such as the Ising Glauber dynamics can be seen as algorithms for constructing samples from the Gibbs distribution of a spin system. As we will later see, this computational task is closely related to that of computing the partition function. Whilst the Ising model and its associated Glauber dynamics serve to illustrate some of the historical and physical background behind the study of spin systems, we will be interested in more general examples of spin systems, on more general classes of graphs. The primary objective of the work in this thesis is to develop efficient algorithmic approaches for approximately sampling from the Gibbs distributions (and approximating the partition functions) of spin systems.

In the statistical physics literature, the graph underlying a spin system, such as the Ising model, typically represents some physically relevant structure. For example, it could be a lattice, which might describe the atomic structure of a ferromagnet. In computer science, however, it is of greater interest to study spin systems, such as the Ising model, on more general families of graphs. The reason for this is that spin systems capture many interesting combinatorial problems on graphs; for example, the problem of computing the partition function of the Ising model can alternatively be viewed as that of counting its weighted 2-colourings. Several other classical combinatorial counting problems are captured by the spin system framework. For example, counting proper colourings of a graph, counting independent sets of a graph, and, most generally, counting (weighted) graph homomorphisms.

1.1 Bounded-degree graphs

In recent years, several fascinating connections have been uncovered between approximate counting and sampling in spin systems, and phase transition phe-

nomena in statistical physics [91, 83, 82, 46, 43]. Beginning with the hard-core model for weighted independent sets, we will now describe some of the main algorithmic results and techniques. Let $\Delta \geq 3$ be an integer, let $\lambda > 0$ be a real number, and let $G = (V_G, E_G)$ be a graph of maximum degree at most Δ. The hard-core model is defined by the probability distribution $\mu_{G,\lambda}(I) \propto \lambda^{|I|}$, where I is an independent set of G. The normalising factor $Z_{G,\lambda}$ is known as the partition function. Notice that when $\lambda = 1$, the partition function simply counts the number of independent sets of G. In computer science, the problem of counting weighted independent sets is a fundamental problem which has been shown to be #P-hard, even on graphs of maximum degree at most 3 [55]. In light of this computational hardness, a natural question to ask is whether there is an efficient approximation algorithm, specifically an FPRAS, for counting weighted independent sets. For general graphs, it is known that there cannot be an FPRAS for the partition function of the hard-core model unless NP = RP [81]. Attention has therefore focussed on input instances of bounded maximum degree, both in the physics and computer science communities.

1.1.1 High temperature algorithms

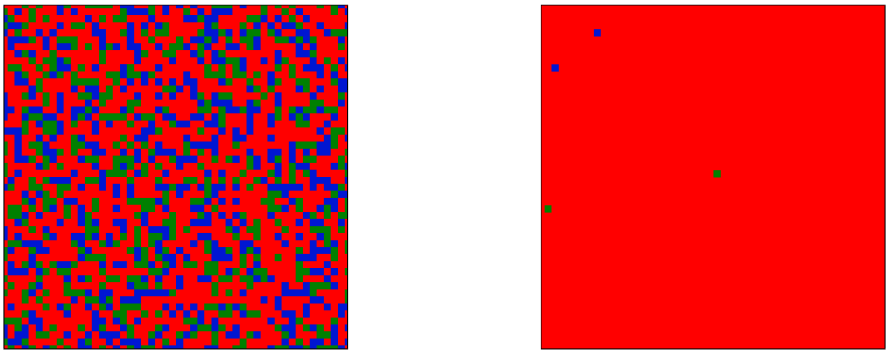
For the hard-core model on the infinite d-regular tree, there exists a critical threshold $\lambda_c(d) = (d-1)^{(d-1)}/(d-2)^d$ at which a range of interesting behaviour emerges. The critical value $\lambda_c(d)$ is known as the uniqueness threshold, below which there is a unique Gibbs measure on the infinite d-regular tree, and above which there are two – one corresponding to the 'even' phase of the model and the other corresponding to the 'odd' phase. In addition to this phase transition at the uniqueness threshold, another interesting phenomenon emerges, termed weak spatial mixing. Weak spatial mixing essentially says that if we fix a set of spins far away from the root of the tree, the marginal probability in the Gibbs distribution that the root is occupied, conditioned on these fixed spins, decays exponentially in the distance from the root to the fixed spins. In the uniqueness regime, there is weak spatial mixing, and the effect of the fixed set of spins disappears as the distance tends to infinity. Conversely, in the non-uniqueness regime, weak spatial mixing does not hold and the boundary condition can continue to have an effect on the marginal probability that the root is occupied, even as the distance tends to infinity.

In a breakthrough paper [91], Weitz exploited the idea of weak spatial mixing in order to give the first efficient algorithm for sampling from the hard-core model (and approximating its partition function) which works for all values of λ below the uniqueness threshold. More formally, he gave an FPTAS for the partition function of the hard-core model that runs in polynomial time given any graph of maximum degree Δ, when $\lambda < \lambda_c(\Delta)$. Prior to this work, efficient algorithms were known only for smaller values of λ, and did not cover the entire uniqueness regime. One example is the result of Vigoda [90], which proved rapid mixing of the Glauber dynamics on graphs of maximum degree Δ, for all values of λ strictly less than $2/(\Delta - 2)$. Much more recently, using a technique known as spectral independence [23, 1, 20], the Glauber dynamics Markov chain has been shown to mix rapidly on graphs of maximum degree Δ, for all values of $\lambda < \lambda_c(\Delta)$.

For a variety of reasons such as those described above, efficient algorithms for sampling from the Gibbs distributions of spin systems typically apply when interactions between spins are sufficiently weak. The low fugacity regime of the hard-core model, where $\lambda < \lambda_c(\Delta)$, can roughly be thought of as the analogue of the high temperature regime of the Ising model, where interactions between spins are weak. In a slight abuse of terminology, for an arbitrary spin system, we will often refer to the regime in which spin interactions are weak as the 'high temperature' regime; similarly, we will often refer to the regime in which spin interactions are strong as the 'low temperature' regime.

As a second example, consider the q-colour ferromagnetic Potts model (this is the q-spin generalisation of the Ising model). Here, efficient algorithms exist for the high temperature regime; that is, when neighbouring spins have a weaker preference to align. One example is the work of Bordewich, Greenhill, and Patel [8], who prove that the Glauber dynamics for the q-colour ferromagnetic Potts model mixes rapidly on all graphs of maximum degree Δ, for β sufficiently small as a function of q and Δ. For spin systems in general, it is often the case that single update Markov chains such as the Glauber dynamics mix rapidly when interactions between spins are weak. Intuitively, this is because all configurations are weighted approximately equally, and the Glauber dynamics can move freely over all states. Figure 1.1 demonstrates this intuition – it shows a simulation of the Glauber dynamics for the 3-colour Potts model on

the grid, initialised from the all-red configuration. At high temperatures, there is a weaker preference for neighbouring spins to align and there is significant disorder in the model. This is in contrast to the low temperature regime, where there is a strong tendency for adjacent spins to align, and the Glauber dynamics struggles to move away from the all-red configuration.



(a) High temperature regime: $\beta < \beta_c$ (b) low temperature regime: $\beta > \beta_c$

Figure 1.1: A simulation of the Glauber dynamics for the 3-colour ferromagnetic Potts model on a grid, started from the all-red configuration.[1]

1.1.2 Low temperature bottlenecks

When interactions between spins are strong, algorithmic approaches such as the Glauber dynamics and correlation decay (the method introduced by Weitz) no longer succeed. For example, for the hard-core model on graphs of maximum degree Δ, where Weitz gave an efficient sampling algorithm for all $\lambda < \lambda_c(\Delta)$, it was shown by Mossel, Weitz, and Wormald [77] that the Glauber dynamics mixes exponentially slowly for $\lambda > \lambda_c(\Delta)$. It had long been conjectured that the critical point $\lambda_c(\Delta)$ gave a hardness threshold for sampling from the hard-core model on bounded-degree graphs, but it was not until the breakthrough results of Sly [83], Sly and Sun [84], and Galanis, Štefankovič, and Vigoda [43], that this was confirmed. Their results showed that for all $\lambda > \lambda_c(\Delta)$, there can

[1]The code used to produce this simulation is available at `https://github.com/j-iss/potts-glauber`.

be no FPRAS for the partition function of the hard-core model on bounded-degree graphs, unless NP = RP. This complements the result of Weitz, giving a complexity dichotomy for the problem of sampling from the hard-core model on bounded-degree graphs.

As a second example, consider the q-colour ferromagnetic Potts model. In contrast to their rapid mixing result at high temperature, Bordewich, Greenhill, and Patel [8] show that at sufficiently low temperatures, the Glauber dynamics mixes exponentially slowly on almost all regular graphs. Whilst this establishes the slow mixing of the Glauber dynamics, a stronger result was proven by Galanis, Štefankovič, Vigoda, and Yang [46], who showed that it is in fact #BIS-hard to approximate the partition function of the q-colour ferromagnetic Potts model on graphs of maximum degree Δ, when β is sufficiently large as a function of q and Δ. #BIS is the problem of counting independent sets in bipartite graphs, and its complexity is a major open problem in the area of approximate counting – it is conjectured not to admit an FPRAS.

1.2 Efficient low temperature algorithms

In order to make progress towards resolving the complexity of #BIS, it is interesting to consider relaxations of #BIS-hard problems and attempt to classify their computational complexity. With this in mind, recent work has considered the class of bounded-degree expander graphs. A second motivation to consider this class of graphs is that a random regular graph is, with high probability, a bounded-degree expander, both in the bipartite and non-bipartite case. Studying the complexity of hard problems on random instances helps us understand the average case complexity of a problem. It is often the case that although a problem might be hard to solve in the worst case, we can develop efficient algorithms that work for *most* input instances.

Recently, Jenssen, Keevash, and Perkins [68] gave efficient low temperature algorithms for a variety of #BIS-hard problems, restricted to instances with expansion properties. For the hard-core model on bounded-degree bipartite expanders, they give an FPTAS and an efficient sampling algorithm, when λ is sufficiently large with respect to the maximum degree of the input graph. Fascinatingly, this is in contrast to the traditional setting of bounded-degree

graphs, where all known efficient algorithms apply for sufficiently small values of λ. Although their results do not apply all the way down to the uniqueness threshold $\lambda_c(\Delta)$, so as to give efficient algorithms for all values of λ, they do cover the case where $\lambda = 1$. This therefore gives, for sufficiently large Δ, an efficient algorithm for counting independent sets of a random Δ-regular bipartite graph. Efficient approximation algorithms are also given for the low temperature Potts model, and counting proper colourings. They give an FPTAS for the ferromagnetic Potts model when β is a sufficiently large function of q and Δ, and an FPTAS for counting proper colourings when Δ is a sufficiently large function of q. These results apply on bounded-degree expander graphs which are not necessarily bipartite. This is in contrast to their results for the hard-core model, which only apply to bipartite graphs. For the same reasons as we discussed earlier, these results all apply to random regular (bipartite) graphs, with high probability, as a corollary. Independently, Liao, Lin, Lu, and Mao obtained similar results [74].

Other classes of graphs for which similar low temperature algorithms have been developed are finite subsets of the integer lattice $\mathbb{Z}^d$ (with boundary conditions) and the torus $(\mathbb{Z}/n\mathbb{Z})^d$. Although not expanders, using the isoperimetric properties of, for example, $\mathbb{Z}^d$, similar methods to those of Jenssen, Keevash, and Perkins [68] can be applied to yield efficient low temperature algorithms for the Potts and hard-core models. For more information, see a paper of Helmuth, Perkins, and Regts [59].

1.2.1 Polymer models

The main tool employed by the above works is the polymer model framework. The polymer model framework [71, 56] is a classical tool from statistical mechanics. Defined by Kotecký and Preiss [71], they provide an important tool for studying the equilibrium phases of statistical physics models on lattices. For an example of such work see [72, 13], and for a more detailed history of their use in statistical physics, see [9]. Recently, polymer models have been used to obtain efficient approximation algorithms for analysing spin systems (such as the ferromagnetic Potts model) in parameter regimes where standard algorithmic approaches are provably inefficient on general graphs. These algorithms

apply to certain classes of graphs that typically have sufficiently strong expansion properties relative to their local growth rates; for example, on the integer lattice [59] and random regular graphs [68, 74].

Polymer models allow us to re-write the partition function of a low temperature spin system in terms to the deviations that allowed configurations take from some ground state. This is done by defining appropriate polymers – connected components of vertices that are not assigned the ground state spin. These polymers have weights, and they form a graph whose independent sets are allowed polymer configurations. The allowed polymer configurations represent the possible deviations that a spin configuration can take from the ground state. The model induces a partition function of its own, whose value should closely approximate that of the underlying spin system. A polymer model is essentially an instance of the (unbounded degree) multivariate hard-core model, where polymers correspond to vertices and their weights correspond to fugacities. What is particularly interesting, is that the obtained polymer model has much in common with an instance of the 'high temperature' hard-core model. As we have seen through several earlier examples, algorithms for sampling from spin systems typically apply in this regime. This therefore offers some hope that existing algorithmic techniques might again become useful, by using them to sample from the high temperature representation (the polymer model) of the underlying low temperature spin system.

1.2.2 Truncated Taylor series

Although a polymer model resembles a high temperature instance of the multivariate hard-core model, we cannot simply apply existing high temperature methods in order to sample from its Gibbs distribution. One reason is that the maximum degree of the graph of polymers is typically not bounded – a condition required to apply most existing methods. In order to sample from the Gibbs distribution of a polymer model, we must use some of the additional structure that the polymer inherit from the underlying instance of a spin system. Subject to certain conditions on the weights of the polymers, the combined work of Helmuth, Perkins, and Regts [59] and Jenssen, Keevash, and Perkins [68], gives an FPTAS for the partition function of a polymer model. Roughly speaking, this condition is satisfied by ensuring that the underlying spin system is

sufficiently far into the low temperature regime. Their approach follow that of Barvinok [2], and they write the log partition function of a polymer model as a Taylor series, which they prove does not vanish outside of some disc in the complex plane. They then prove that a certain truncation of this log partition function approximates that of the polymer model, and how to efficiently compute the coefficients of this truncated Taylor series. Although this approach gives an efficient algorithm, it should be noted that its analysis relies on extensive tools from the complex analysis literature, and its running time is typically of the form $n^{O(\log \Delta)}$, where Δ is the maximum degree of the underlying instance of the spin system. Such running times are much slower than those typically obtained via Markov chains.

1.3 Our results

We conclude this introductory section by giving a high-level overview of the results presented in this thesis. The theme of these results is low temperature algorithms for spin systems, and we present several different algorithmic approaches that apply in a variety of different settings. In a similar spirit to existing work in the area of low temperature algorithms, our results appeal to and build upon the polymer model framework of [59] and [68]; that is, we will design and analyse algorithmic approaches for sampling from the Gibbs distribution of a polymer model. Our algorithms are mostly based on Markov chains[2], which provides several advantages over existing approaches. As algorithms, Markov chains are inherently simple (both to describe and implement) when compared with other kinds of algorithms such as the truncated Taylor series approach taken in [68]. In addition to this, Markov chains typically result in algorithms with running times of the form $O(n \log n)$, which is a marked improvement over those of existing techniques. Finally, by analysing Markov chains, we can more easily relate our results to canonical approaches such as the Glauber dynamics. We apply our new algorithmic framework in a variety of different settings.

[2]All of the algorithms that we present in this thesis are randomised. This is in contrast to those of previous work (e.g. [59, 68, 74]), which are deterministic.

1.3.1 Markov chains for polymer models

In Chapter 3, we present a Markov chain based algorithm for sampling from the Gibbs distribution of an abstract polymer model, subject to only a natural condition on the weights of the polymers. We call this Markov chain the polymer dynamics, and prove its rapid mixing using standard path coupling techniques. By using the same polymer models presented in [68] and [74], we obtain efficient algorithms for the low temperature Potts model on bounded-degree expanders (see Theorem 3.2.4), and the high-fugacity hard-core model on bounded-degree bipartite expanders (see Theorem 3.2.5). The algorithms that we obtain apply in a similar range of parameters to those presented in [68] and [74], yet provide a simplified approach. We also improve the running times of these algorithms; for example, our sampling algorithm produces an ε-sample approximation from the low temperature Potts model in time $O(n \log(n/\varepsilon) \log(1/\varepsilon))$.

1.3.2 General spin systems

The results of both Jenssen, Keevash, and Perkins [68] and Section 3.3 use the polymer model framework to give efficient algorithms for the Potts and hard-core models. In Chapter 4, we develop a polymer model for general spin systems on regular bipartite expander graphs. This generalises existing results, and gives us efficient approximation algorithms for *any* spin system, when the interactions between spins are sufficiently strong. In order to prove the correctness of our algorithms, we show that the so-called 'maximal bicliques' of the interaction matrix of a spin system correspond to ground state configurations. We then propose a natural polymer model which represents the deviations that spin configurations take from these maximal bicliques. Using the framework proposed in Section 3.3, we present an algorithm for sampling from the Gibbs distribution of any general spin system, when the degree and the spectral gap of the graph are sufficiently large (see Theorem 4.2.1).

1.3.3 Unbounded degree graphs

So far, the polymer model framework only applies to classes of graphs of bounded maximum degree. The reason for this restriction is that the analysis of algorithms for sampling from polymer models requires very careful control over the

growth rate of the number of polymers, and a simple way to achieve this is to ensure that the maximum degree of the underlying graph is bounded. This restriction greatly limits the applicability of the polymer model framework, and rules out low temperature algorithms for classes of graphs such as sparse random graphs, which typically have strong expansion properties and bounded average degree. This perhaps does not reflect intuition, which suggests that the presence of high-degree vertices should only intensify the ground state phenomenon that enables the application of polymer models.

In order to overcome this limitation, In Chapter 5 we revise the polymer model framework presented in Section 3.3, by measuring the size of a polymer by its total degree, as opposed to its number of vertices. Through this simple modification, and subject to new conditions on the weights of polymers, we are able to obtain a rapidly mixing Markov chain for sampling from the Gibbs distribution of a polymer model. The essential reason is that we are instead able to bound the growth rate of the number of polymers, in terms of the total degree of these polymers, which is mitigated by the expansion properties of the underlying graph. In order to demonstrate the applicability of our revised method, we give an efficient algorithm for the low temperature Potts model on random graphs with a given degree sequence, in which some vertices have unbounded degrees (see Theorem 5.2.3). This is an application not possible through previous versions of the polymer model framework; for example, of Section 3.3 or the work of [68].

1.3.4 Restricted Glauber dynamics

The Potts model perhaps best serves to illustrate the intuition behind the slow mixing of the Glauber dynamics in the low temperature regime. At low temperatures, on bounded-degree expanders, a typical sample from the Gibbs distribution of the Potts model is mostly monochromatic, with the exception of perhaps some small regions that might go against the dominant colour. This ensures that the state space is effectively partitioned into q equally weighted regions, each corresponding to one of q colours, and separated by exponentially small bottlenecks. Single-update Markov chains such as the Glauber dynamics struggle to cross these bottlenecks, and as such, cannot efficiently sample from the full Gibbs measure. In spite of this, a natural idea, which was suggested in

both [59] and [68], is to fix a colour (let's call this colour red) and simulate the Glauber dynamics within the portion of the state space space that is mostly red.

Inspired by this suggestion, we show in Chapter 6 that when the Glauber dynamics for the low temperature Potts model is restricted to remain within the red portion of the state space, it rapidly converges to the Gibbs distribution restricted to red-dominated colourings. We prove this by using standard Markov chain comparison techniques, by comparing the restricted Glauber dynamics to the polymer dynamics Markov chain we introduce in Section 3.3. The result is an efficient algorithm for sampling from the low temperature Potts model on bounded-degree expanders (see Theorem 6.1.1). This establishes that, in spite of its slow convergence to the full Gibbs distribution, canonical Markov chains such as the Glauber dynamics can still be used as the basis for efficient low temperature algorithms.

1.3.5 Markov chains for $\mathbb{Z}^d$

Helmuth, Perkins, and Regts [59] give an FPTAS for the partition function of the low temperature Potts model on a finite region of $\mathbb{Z}^d$ with padded monochromatic boundary conditions. In a similar spirit to the work in Section 3.3, and with similar motivation in mind, In Chapter 7 we develop a Markov chain based algorithm for sampling from the Gibbs distribution of the low temperature Potts model on $\mathbb{Z}^d$, with boundary conditions (see Theorem 7.2.4). More generally, we give an efficient algorithm for sampling from the Gibbs distribution of a contour model. Contour models, similar to polymer models, can be used to represent low temperature spin systems on the the integer lattice, and are the main tool used to obtain the algorithms of [59]. Our approach combines the polymer dynamics chain of Section 3.3 with an additional 'recolouring' procedure, which allows us to prove the rapid mixing of a certain Markov chain whose stationary distribution is that of the contour model.

Chapter 2

Preliminaries

In this chapter, we provide some of the background, definitions, and results that we will frequently refer to throughout this thesis. Much of the content of this chapter consists of standard definitions and results from, but not limited to, the probability and combinatorics literature.

2.1 Graph notation

We use $G = (V_G, E_G)$ to denote the graph G with vertex set V_G and edge set $E_G \subseteq V_G \times V_G$. For a vertex $v \in V_G$, we let $\deg_G(v)$ denote the degree of v in G, and for a vertex subset $S \subseteq V_G$, we let $\deg_G(S) = \sum_{v \in S} \deg_G(v)$ denote the total degree of S in G. For a vertex subset $S \subseteq V_G$, we let $S^c = V_G \setminus S$ denote the complement of S, we let $\partial_G S = \{v \in S^c \mid \exists u \in S,\ \{u, v\} \in E_G\}$ denote the vertex boundary of S, and we let $S^+ = S \cup \partial_G S$. For vertex subsets $S, T \subseteq V_G$, we let $E_G(S, T)$ denote the set of edges that have one endpoint in S and the other in T, and we let $e_G(S, T) = |E_G(S, T)|$; when $S = T$, we let $E_G(S)$ and $e_G(S)$ denote $E_G(S, S)$ and $e_G(S, S)$, respectively. We will often use n to refer to the number of vertices of a graph, and m to refer to its number of edges. We use $G = (V_G^0, V_G^1, E_G)$ to denote the bipartite graph G with bipartition (V_G^0, V_G^1) and edge set $E_G \subseteq V_G^0 \times V_G^1$. We use $V_G = V_G^0 \cup V_G^1$ to denote the vertex set of G.

2.2 Approximate counting and sampling

A counting problem is a natural generalisation of a decision problem. In many settings, we are interested in knowing the number of solutions to a problem, and not just whether one exists. Just as the complexity class NP captures the difficulty of decision, the complexity class #P, which can be viewed as a generalisation of NP, captures the complexity of counting.

Definition 2.2.1. A counting problem $f\colon \{0,1\}^* \to \mathbb{N}$ is in #P if there exists a polynomial-time non-determinsitc Turing machine M such that for all $x \in \{0,1\}^*$, the number of accepting paths of M on input x, is $f(x)$.

This class was introduced by Valiant [88], who proved the #P-completeness of computing the permanent of a 0-1 matrix in his seminal work which initiated the study of the complexity of counting. Since then, many other[1] counting problems have been classified as #P-hard. Examples include counting satisfying assignments to a CNF formula, counting satisfying assignments to a DNF formula, counting proper k-colourings of a graph, and many more.

Clearly, counting is at least as hard as decision. In fact, it is known by Toda's theorem [87] that any problem in the polynomial hierarchy can be solved given an oracle for #P. In light of the apparent computational barrier that the #P-hardness of many natural counting problems suggests, our attention instead turns towards approximation.

We say that $\hat{x}$ is an ε-approximation to x if $(1-\varepsilon)x \leq \hat{x} \leq (1+\varepsilon)x$. Furthermore, we say that a random variable X is an (ε, δ)-approximation to x if X is an ε-approximation to x with probability at least $1-\delta$. We will primarily be interested in obtaining the following kind of algorithm.

Definition 2.2.2. A fully-polynomial randomised approximation scheme (FPRAS) for a counting problem $f\colon \{0,1\}^* \to \mathbb{R}$ is randomised algorithm that takes as input an instance $x \in \Sigma^*$ and an accuracy parameter $\varepsilon > 0$ and, outputs a random variable that is an $(\varepsilon, \frac{1}{4})$-approximation to $f(x)$ in time polynomial in both $|x|$ and ε^{-1}.

[1] This includes problems that correspond to decision problems in P. As an example, computing the permanent of a 0-1 matrix can be viewed as counting the number of perfect matchings in a bipartite graph, which is known to be in P.

The above definition is fairly robust. The value $1/4$ can be replaced by any $\delta < 1/2$ by taking the median of $O(\log \delta^{-1})$ trials, resulting in the same definition. Similarly, if we instead say that $\hat{x}$ is an ε-approximation to x if $e^{-\varepsilon}x \leq \hat{x} \leq e^{\varepsilon}x$, this also results in the same definition. Finally, for many natural f, we can boost a constant factor approximation to an ε-approximation at very little computational expense[2]. We will occasionally encounter the deterministic analogue of an FPRAS, which returns an ε-approximation with probability 1, and is known as a fully-polynomial time approximation scheme (FPTAS).

The problem of approximate counting is closely related to the problem of approximate sampling. We can associate a solution set $S(x)$ to each input x to a counting problem in #P. We will often be interested in sampling a solution at random according to some probability distribution (perhaps this is the uniform distribution on the solution set $S(x)$). We will mostly be concerned with approximate sampling, where we are interested in constructing a sample that is distributed within a small total variation distance of the target distribution.

Definition 2.2.3. The total variation distance between two probability distributions π and π' over a countable sample space Ω is given by[3]

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

If a sample is distributed within total variation distance ε of π, we call this an ε-sample from π.

We will be interested in obtaining efficient approximate sampling algorithms. One possible notion of such as algorithm is as follows.

Definition 2.2.4. A fully-polynomial almost uniform sampler (FPAUS) for a counting problem $f\colon \{0,1\}^* \to \mathbb{N}$ is an algorithm that takes as input an instance $x \in \{0,1\}^*$ and an accuracy parameter $\varepsilon > 0$, and outputs an ε-sample[4] from the uniform distribution on $S(x)$ in time polynomial in both $|x|$ and $\log \varepsilon^{-1}$.

[2]As an example, suppose that we have access to an oracle which returns a constant factor approximation to the number of proper colourings of a graph G. We can use this oracle to compute a constant factor approximation to the number of proper colourings of k disjoint copies of G. Taking the k^{th} root of this quantity yields an ε-approximation to the number of proper colourings of G, for a certain value of $k = O(1/\varepsilon)$

[3]For a proof of the second equality, see Lemma 12.1 of [76].

[4]For MCMC, it is sufficient to return a sample distributed within total variation distance $1/4$ of the uniform distribution, as this can be boosted into an ε-sample in $O(\log \varepsilon^{-1})$ time. See Theorem 12.16 of [76] for more details.

A polynomial-time equivalence between almost uniform sampling and approximate counting was demonstrated in the seminal work of Jerrum, Valiant, and Vazirani [69], for the fairly general class of so-called self-reducible counting problems. We will also be interested in algorithms of the above kind where the uniform distribution is replaced with some other distribution.

2.3 Spin systems

Spin systems are general frameworks that encompass sampling and counting problems in computer science, graph homomorphisms in combinatorics, and phase transition phenomena in statistical physics. We begin by giving two important examples of spin systems, which will appear frequently throughout this thesis.

2.3.1 The Potts model

Let $q \geq 2$ be an integer and let $[q] = \{1, \ldots, q\}$. Let $\beta > 0$ be a real number. For a graph $G = (V_G, E_G)$, let $\Omega_{G,q}$ denote the set of all (not necessarily proper) colourings of V_G. For $\sigma \in \Omega_{G,q}$, let $b_G(\sigma)$ denote the number of bichromatic edges of E_G induced by σ. The partition function of the q-colour ferromagnetic Potts model at inverse temperature β, is defined by

$$Z_{G,q,\beta} = \sum_{\sigma \in \Omega_{G,q}} e^{-\beta b_G(\sigma)}.$$

The Gibbs distribution of the model is the probability distribution $\mu_{G,q,\beta}$ over $\Omega_{G,q}$, defined by

$$\mu_{G,q,\beta}(\sigma) = \frac{e^{-\beta b_G(\sigma)}}{Z_{G,q,\beta}}.$$

In the case where $q = 2$, the model is known as the Ising model.

2.3.2 The hard-core model

Let $\lambda > 0$ be a real number. For a graph $G = (V_G, E_G)$, let $\mathcal{I}_G$ denote the set of all independent sets of G. The partition function of the hard-core model at fugacity λ, is defined by

$$Z_{G,\lambda} = \sum_{I \in \mathcal{I}_G} \lambda^{|I|}.$$

The Gibbs distribution of the model is the probability distribution $\mu_{G,\lambda}$ over $\mathcal{I}_G$, defined by

$$\mu_{G,\lambda}(I) = \frac{\lambda^{|I|}}{Z_{G,\lambda}}.$$

In the case where $\lambda = 1$, the partition function of the hard-core model is equal to $|\mathcal{I}_G|$, and the Gibbs distribution is the uniform distribution over $\mathcal{I}_G$.

2.3.3 General spin systems

The Potts model is one example of what is known as a general spin system. Let $q \geq 2$ be an integer. A q-spin system is specified by the set of spins $[q]$ and a symmetric interaction matrix $H \in \mathbb{R}_{\geq 0}^{q \times q}$. Given a graph $G = (V_G, E_G)$, a spin configuration is an assignment $\sigma : V_G \to [q]$ and the weight of σ is given by $w_{G,H}(\sigma) = \prod_{\{u,v\} \in E_G} H_{\sigma(u),\sigma(v)}$. Let $\Sigma_{G,H}$ denote the set of all possible spin configurations. The partition function of the spin system is defined by

$$Z_{G,H} = \sum_{\sigma \in \Sigma_{G,H}} w_{G,H}(\sigma).$$

The Gibbs distribution of the system is the probability distribution $\mu_{G,H}$ over $\Sigma_{G,H}$ defined by

$$\mu_{G,H}(\sigma) = \frac{w_{G,H}(\sigma)}{Z_{G,H}}.$$

As mentioned earlier, well-known examples of spin systems are the Ising/Potts models, where the matrix H has all diagonal entries equal to 1 and off-diagonal entries equal to $e^{-\beta}$, where β is the inverse temperature; the case $q = 2$ is the Ising model, and $q > 2$ is the Potts model (see Figure 2.1 (a)). Proper colourings also fit naturally into this framework (see Figure 2.1 (b)), as do graph homomorphisms, whenever H has 0-1 entries. The hard-core model of Section 2.3.2 is not strictly captured by the spin system framework; however it is if we extend the general spin system framework to include vertex weights (which is often referred to as an external field).

$$\begin{bmatrix} 1 & e^{-\beta} & e^{-\beta} & e^{-\beta} \\ e^{-\beta} & 1 & e^{-\beta} & e^{-\beta} \\ e^{-\beta} & e^{-\beta} & 1 & e^{-\beta} \\ e^{-\beta} & e^{-\beta} & e^{-\beta} & 1 \end{bmatrix}$$

(a) The 4-colour Potts model

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

(b) Proper 5-colourings

Figure 2.1: Examples captured by the spin system framework.

2.4 Markov chains and mixing times

In this section, we present some the key background and definitions related to Markov chains. Markov chains provide a simple framework for generating random samples from a probability distribution, and find a range of applications in areas such as Bayesian statistics, information theory, physics, and statistical mechanics. In the latter of these settings, Markov chains provide a model for simulating the evolution of physically relevant spin systems, such as the Ising model. In computer science, they provide an algorithmic technique for sampling from the Gibbs distribution of a spin system — a fundamental idea that many of the results in this thesis will exploit.

2.4.1 Glauber dynamics

Let $q \geq 2$ be an integer, let $H \in \mathbb{R}_{\geq 0}^{q \times q}$ be a symmetric interaction matrix, and let $G = (V_G, E_G)$ be a graph. The Glauber dynamics for the spin system is a local Markov chain that chooses a vertex of G uniformly at random, and updates its spin according to the Gibbs distribution of the spin system, conditioned on the spins of its neighbours. More formally, for an integer $t \geq 0$, the heat-bath Glauber dynamics is the Markov chain that transitions from a spin configuration $\sigma_t \in \Sigma_{G,H}$ to a spin configuration $\sigma_{t+1} \in \Sigma_{G,H}$ according to the following rules.

1. Choose a vertex $v \in V_G$ uniformly at random.
2. Let $\sigma_{t+1}(u) = \sigma_t(u)$ for all $u \in V_G$ such that $u \neq v$.
3. Choose $i \in [q]$ with probability $\frac{\prod_{u \in \partial_G\{v\}} H_{i,\sigma_{t+1}(u)}}{\sum_{j \in [q]} \prod_{u \in \partial_G\{v\}} H_{j,\sigma_{t+1}(u)}}$ and let $\sigma_{t+1}(v) = i$.

The unique stationary distribution of the Glauber dynamics is the Gibbs distribution of the spin system, which it is straightforward to verify using detailed balance[5].

As it will feature as a running example throughout, but is not strictly captured by the general spin system framework of Section 2.3.3, we define the Glauber dynamics for the hard-core model at fugacity λ as follows. For an integer $t \geq 0$, we transition from an independent set $I_t \in \mathcal{I}_G$ to an independent set $I_{t+1} \in \mathcal{I}_G$ according to the following rules.

1. Choose a vertex $v \in V_G$ uniformly at random.
2. Mutually exclusively:
   (a) let $I' = I_t \cup \{v\}$ with probability $\frac{\lambda}{1+\lambda}$ and
   (b) let $I' = I_t \setminus \{v\}$ with probability $\frac{1}{1+\lambda}$.
3. Let $I_{t+1} = I'$ if $I' \in \mathcal{I}_G$; otherwise, let $I_{t+1} = I_t$.

2.4.2 Mixing times

In order to use a Markov chain as a tool for obtaining efficient sampling algorithms, we will often be interested in its mixing time. Intuitively, the mixing time tells us how long we need to simulate a Markov chain for, until its distribution is close (with respect to the total variation distance) to stationary. We will typically be interested in the number of steps required (as a function of ε) for a Markov chain to reach within total variation distance (see Definition 2.2.3) ε of its stationary distribution.

Definition 2.4.1. Let $\mathcal{M}$ be an ergodic Markov chain on a countable state space Ω, with transition matrix P and unique stationary distribution π. For an integer $t \geq 0$, let p_x^t denote the distribution of $\mathcal{M}$ after t steps, when starting

[5] A Markov chain $\mathcal{M}$ on a finite state space Ω with transition matrix P is said to be time-reversible with respect to a distribution π on Ω if all states $i, j \in \Omega$ satisfy $\pi_i P_{i,j} = \pi_j P_{j,i}$. If $\mathcal{M}$ is time-reversible with respect to π then this is a stationary distribution of $\mathcal{M}$, and if $\mathcal{M}$ is ergodic then this is the unique stationary distribution of $\mathcal{M}$. For further details, see Section 7.3 of [76].

from state $x \in \Omega$. For $\varepsilon > 0$, the mixing time of $\mathcal{M}$ with initial state x is given by

$$T_x(\varepsilon) = \min\{t > 0 \mid \text{for all } t' \geq t, \|p_x^{t'} - \pi\|_{\mathsf{TV}} \leq \varepsilon\}.$$

The mixing time of $\mathcal{M}$ is given by $T_{\text{mix}}(\varepsilon) = \max_{x \in \Omega} T_x(\varepsilon)$, and we will write $T_{\text{mix}}(\mathcal{M}, \varepsilon)$ if the Markov chain is not clear from the context.

2.4.3 Coupling of Markov chains

Coupling is a standard technique from probability theory that allows us to compare two random variables by 'forcing' them to behave in the same way. When applied to Markov chains, coupling provides a powerful tool for bounding the mixing time, and is a central technique to many of the results in this thesis.

Definition 2.4.2. A coupling of a Markov chain (M_t) on a finite state space Ω, is a Markov chain (X_t, Y_t) with state space $\Omega \times \Omega$ such that

$$\Pr[X_{t+1} = x' \mid (X_t, Y_t) = (x, y)] = \Pr[M_{t+1} = x' \mid M_t = x]$$

and

$$\Pr[Y_{t+1} = y' \mid (X_t, Y_t) = (x, y)] = \Pr[M_{t+1} = y' \mid M_t = y].$$

The above definition says that, in a coupling (X_t, Y_t) of the Markov chain (M_t), each of the Markov chains (X_t) and (Y_t) are faithful copies of (M_t). The chains (X_t) and (Y_t) are not necessarily in the same state, but their transition probabilities should be the same as those of (M_t). In order to bound the mixing time of (M_t), we will be interested in finding couplings that bring (X_t) and (Y_t) together, by forcing their behaviour to depend on the same random choices. Intuitively, if we imagine that one chain in the coupling is at stationarity, the time until coupling will tell us the mixing time of the underlying Markov chain (M_t). The following lemma is known as the Coupling Lemma, and formalises this idea.

Lemma 2.4.3. *[76, Lemma 12.2] Let (X_t, Y_t) be a coupling of a Markov chain $\mathcal{M}$ on a finite state space Ω and with a unique stationary distribution. Suppose that there is $t\colon (0,1) \to \mathbb{Z}_{\geq 0}$ such that for all $x, y \in \Omega$ and all $\varepsilon \in (0,1)$, we have that*

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

then $T_{mix}(\mathcal{M}, \varepsilon) \leq t(\varepsilon)$.

Instead of directly considering how X_t and Y_t differ, we can use a metric to monitor the convergence of a coupling. The following result is known as the Coupling Contraction Lemma [15].

Lemma 2.4.4. *Let $\mathcal{M}$ be a Markov chain on a finite state space Ω and with a unique stationary distribution. Let $d\colon \Omega \times \Omega \to \mathbb{Z}_{\geq 0}$ be a metric on Ω and let $D = \max_{x,y\in\Omega} d(x,y)$. Let $\beta \in (0,1)$. Suppose that (X_t, Y_t) is a coupling for $\mathcal{M}$ such that for all $x, y \in \Omega$ and all $t \geq 0$ we have that*

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

then $T_{mix}(\mathcal{M}, \varepsilon) \leq \lceil \ln(D/\varepsilon)/\ln(1/\beta) \rceil$.

2.4.4 Path coupling

Finding couplings that converge quickly is often a very creative process. This process can be simplified using a framework known as path coupling. Instead of having to establish contraction of a metric (in the sense of Lemma 2.4.4) for all pairs of states, path coupling only requires us to establish contraction for a certain subset of $\Omega \times \Omega$, in order to deduce a mixing time bound. An edge weighted graph (H, d) for Ω consists of a connected graph H with vertex set Ω along with a distance $d\colon \Omega \times \Omega \to \mathbb{Z}_{\geq 0}$. An edge weighted graph (H, d) is minimal if $d(x,y)$ is the length of the shortest path from x to y in H, for all pairs of states $x, y \in \Omega$. The following lemma is known as the Path Coupling Lemma, and is due to Bubley and Dyer [15]

Lemma 2.4.5. *[15, Theorem 1] Let $\mathcal{M}$ be a Markov chain on a finite state space Ω and with a unique stationary distribution. Let (H, d) be a minimal edge weighted graph for Ω and let $D = \max_{x,y\in\Omega} d(x,y)$. Let $\beta \in (0,1)$. Suppose that (X_t, Y_t) is a coupling for $\mathcal{M}$ such that for all edges (x,y) of H and all $t \geq 0$ we have that*

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

then $T_{mix}(\mathcal{M}, \varepsilon) \leq \lceil \ln(D/\varepsilon)/\ln(1/\beta) \rceil$.

2.5 Random graphs

Random graphs are often studied in order to understand the properties of *typical* graphs. Indeed, in algorithms and complexity, studying random inputs allows us to better understand the average case complexity of hard computational problems. We will typically be interested in the asymptotic properties of random graphs, as the number of vertices grows. In this section, we introduce some graph-theoretic notions and simple models for random graphs. As we will see, these models exhibit some very interesting properties that we will later be able to exploit algorithmically.

2.5.1 Random graph models

We will consider models for random graphs with a given degree sequence $\vec{x}$. We write $G \sim \mathcal{G}(n, \vec{x})$ to indicate that G is a graph chosen uniformly at random from the set of all simple n-vertex graphs with degree sequence $\vec{x}$. We say that G satisfies a property with high probability (whp) if the probability that G satisfies the property is $1 - o(1)$ as a function of n, uniformly over $\vec{x}$.

To work with $G \sim \mathcal{G}(n, \vec{x})$, we consider the standard configuration model, where a random multigraph $H = (V_H, E_H)$ with the given degree sequence $\vec{x}$ is sampled by the following process. For each $i \in [n]$, we attach x_i half-edges to the vertex i. We then sample a uniformly random perfect matching on the half-edges to give E_H. This uniformly random perfect matching can be sampled by performing the following until no half-edges remain: choose any remaining half-edge, choose another remaining half-edge uniformly at random, then pair these two half-edges and remove them from the set of remaining half-edges. We write that $H \sim \mathsf{CM}(n, \vec{x})$. Note, for two vertices $i, j \in V_H$ such that $i \neq j$, the probability that a half edge attached to i and a half edge attached to j are paired is

$$p_{\{i,j\}} = \frac{x_i x_j}{2m - 1},$$

where $m = \frac{1}{2}\sum_{k=1}^{n} x_k$. Similarly, the probability that two half-edges of i are connected is

$$p_{\{i,i\}} = \frac{x_i(x_i-1)}{2m(2m-1)}.$$

We will often first prove results about $\mathsf{CM}(n,\vec{x})$, since it is more convenient to work with, and asymptotic properties of $\mathsf{CM}(n,\vec{x})$ can easily be transferred back to $\mathcal{G}(n,\vec{x})$ using the following straightforward consequence of [64, Theorem 1.1].

Lemma 2.5.1. *Let d be a positive real number. For every positive integer n, let $\mathcal{E}_n$ be a set of n-vertex multigraphs. If, for some n-vertex degree sequence $\vec{x} = \{x_1, x_2, \cdots, x_n\}$ with $\sum_{i\in[n]} x_i^2 \leq dn$, we have $G \sim \mathcal{G}(n,\vec{x})$ and $H \sim \mathsf{CM}(n,\vec{x})$ then the following is true. If $H \in \mathcal{E}_n$ with high probability, then $G \in \mathcal{E}_n$ with high probability.*

Proof. Suppose that H' is drawn from $\mathsf{CM}(n,\vec{x})$ conditioned on being simple. It is well-known (for example, see [89, Proposition 7.15]) that H' is a uniformly-random graph with degree sequence $\vec{x}$. Thus, it follows that

$$\begin{aligned}\Pr(G \notin \mathcal{E}_n) &= \Pr(H' \notin \mathcal{E}_n)\\ &= \frac{\Pr(H \notin \mathcal{E}_n, H \text{ is simple})}{\Pr(H \text{ is simple})}\\ &\leq \frac{\Pr(H \notin \mathcal{E}_n)}{\Pr(H \text{ is simple})}. \end{aligned} \tag{2.1}$$

By assumption, we have that $\Pr(H \notin \mathcal{E}_n) = o(1)$. Applying Theorem 1.1 of [64], whose conditions are satisfied by the fact that $\sum_{i\in[n]} x_i^2 \leq dn$, it follows that there is a positive p such that when n is sufficiently large $\Pr(H \text{ is simple}) > p$. By (2.1), we therefore have that $\Pr(G \notin \mathcal{E}_n) = o(1)$, and the result follows. □

2.5.2 Expander graphs

Expander graphs are graphs that are both sparse and highly connected. There are many different ways in which one can describe the expansion properties of a graph; for example, they can be interpreted both in terms of spectral (based on the eigenvalues of the adjacency matrix of the graph) and combinatorial properties. Here, we give the definitions that will be most relevant to the results

we later present. For a detailed introduction to expander graphs, see [60]. We begin with a simple combinatorial notion of expansion.

Definition 2.5.2. Let $\alpha > 0$ be a real number. A graph $G = (V_G, E_G)$ is an α-expander if for all vertex subsets $S \subseteq V_G$ with $|S| \leq |V_G|/2$, we have that $e_G(S, S^c) \geq \alpha|S|$.

We will also require the following combinatorial notion of expansion in bipartite graphs.

Definition 2.5.3. Let $\alpha > 0$ be a real number. A bipartite graph $G = (V_G^0, V_G^1, E_G)$ is a bipartite α-expander if for all $i \in \{0, 1\}$ and all $S \subseteq V_G^i$ with $S \leq |V_G^i|/2$, we have that $|\partial_G S| \geq (1 + \alpha)|S|$.

As mentioned above, we can also describe the expansion of a graph using its spectral properties. Let G be an n-vertex Δ-regular bipartite graph. Let $\lambda_1(G) \geq \lambda_2(G) \geq \cdots \geq \lambda_n(G)$ denote the eigenvalues of the adjacency matrix of G. It is a well-known fact (see, e.g. [60]) that $\lambda_1(G) = \Delta$ and $\lambda_n(G) = -\lambda_1(G)$. We define $\lambda(G) = \lambda_2(G)$.

Definition 2.5.4. Let $\Delta \geq 3$ be an integer and let $\lambda < \Delta$ be a real number. We let $\mathcal{G}_{\Delta,\lambda}^{\text{bip}}$ denote the set of all connected Δ-regular bipartite graphs G, for which $\lambda(G) \leq \lambda$.

We will consider the class $\mathcal{G}_{\Delta,\lambda}^{\text{bip}}$, when λ is a small constant. The reason for this is that the smaller λ is, the better the expansion properties of graphs in $\mathcal{G}_{\Delta,\lambda}^{\text{bip}}$. The following result makes this idea rigorous, and relates the spectrum of a regular bipartite graph to its edge-expansion properties. This result was first proven in [57, Theorem 5.1], though the version we state below is taken from [24].

Lemma 2.5.5. *[24, Lemma 8] Let $G = (V_G^0, V_G^1, E_G) \in \mathcal{G}_{\Delta,\lambda}^{bip}$ with $n = |V_G^0| = |V_G^1|$. Then, for sets $S_0 \subseteq V_G^0, S_1 \subseteq V_G^1$, we have that*

$$\left| e_G(S_0, S_1) - \frac{\Delta |S_0||S_1|}{n} \right| \leq \lambda \sqrt{|S_0||S_1| \left(1 - \frac{|S_0|}{n}\right)\left(1 - \frac{|S_1|}{n}\right)}.$$

The following simple consequence of the above result gives a lower bound on the edge expansion of $G \in \mathcal{G}_{\Delta,\lambda}^{\text{bip}}$ when λ is sufficiently small.

Corollary 2.5.6. *Let $G = (V_G^0, V_G^1, E_G) \in \mathcal{G}_{\Delta,\lambda}^{bip}$ with $n = |V_G^0| = |V_G^1|$. Then, for sets $S_0 \subseteq V_G^0, S_1 \subseteq V_G^1$ such that $\lambda \leq \frac{\Delta}{2n}\sqrt{|S_0||S_1|}$, it holds that $e_G(S_0, S_1) \geq \frac{\Delta}{2n}|S_0||S_1|$.*

Proof. Lemma 2.5.5 implies that

$$\left|e_G(S_0, S_1) - \frac{\Delta|S_0||S_1|}{n}\right| \leq \frac{\Delta|S_0||S_1|}{2n}, \text{ therefore } e_G(S_0, S_1) \geq \frac{\Delta|S_0||S_1|}{2n}.$$

□

A second combinatorial notion of expansion is vertex expansion. A well-known result from Tanner [86] relates the spectrum of a graph to its vertex expansion properties (see also [70] for a more refined estimate). Here we state a version from [60, Theorem 4.15]; there, the result is stated and proved for non-bipartite graphs, but a minor adaptation of the proof in [60], which we give for completeness here, also applies to bipartite graphs.

Lemma 2.5.7. *Let $G = (V_G^0, V_G^1, E_G) \in \mathcal{G}_{\Delta,\lambda}^{bip}$, $\rho > 0$ be a real number and $i \in \{0, 1\}$. Then, for all $S \subseteq V_G^i$ with $|S| \leq \rho|V_G^i|$, it holds that $|\partial_G S| \geq |S|/\big(\rho + \frac{\lambda^2}{\Delta^2}(1-\rho)\big)$.*

Proof. We prove the result for $i = 0$; the case $i = 1$ is symmetric. Let $S \subseteq V_G^0$ be such that $|S| \leq \rho n$ where $n = |V_G^0| = |V_G^1|$. We will show that $|\partial_G S| \geq |S|/(\rho + \frac{\lambda^2}{\Delta^2}(1-\rho))$.

Let $A = A_G$ denote the adjacency matrix of G, and B be the bi-adjacency matrix of G, so that $A = \left[\begin{smallmatrix} 0 & B \\ B^T & 0 \end{smallmatrix}\right]$. Then, we have that μ is an eigenvalue of A iff μ^2 is an eigenvalue of BB^T. Let $v_1, \ldots, v_n$ be orthonormal eigenvectors of BB^T, with eigenvalues $\lambda_1^2, \ldots, \lambda_n^2$, where $\lambda_1, \ldots, \lambda_n$ are the non-negative eigenvalues of A in decreasing order. Note that $\lambda_1^2 = \Delta^2$, $\lambda_2^2 = \lambda^2$ and since G is Δ-regular, we have that $v_1 = \frac{1}{\sqrt{n}}e_1$, where e_1 is the n-dimensional vector with all ones.

From the spectral theorem, we have the decomposition

$$BB^T = \sum_{i\in[n]} \lambda_i^2 v_i v_i^T.$$

Let 1_S be the n-dimensional vector whose i^{th} entry is equal to 1 if $i \in S$ and 0 otherwise. Since $v_1, \ldots, v_n$ is a basis of $\mathbb{R}^n$, we can write

$$1_S = \sum_{i\in[n]} \alpha_i v_i \text{ for some real numbers } \alpha_1, \ldots, \alpha_n. \tag{2.2}$$

Using the orthonormality of $v_1, \dots, v_n$, we obtain by multiplying (2.2) with v_1^T that $\alpha_1 = |S|/\sqrt{n} \le \rho\sqrt{n}$ and by considering the norm of 1_S that $\sum_{i\in[n]} \alpha_i^2 = |S|$. Hence, we have that

$$\begin{aligned}\big\|1_S^T B\big\|^2 &= 1_S^T B B^T 1_S \\ &= \sum_{i\in[n]} \lambda_i^2 \alpha_i^2 \\ &\le \Delta^2 \alpha_1^2 + \lambda^2(|S| - \alpha_1^2) \\ &\le |S|\big(\Delta^2\rho + \lambda^2(1-\rho)\big). \end{aligned} \tag{2.3}$$

For $j \in [n]$, let u_j be the j^{th} entry in $1_S^T B$; observe that u_j is the number of neighbours in S of the j^{th} vertex in V_G^1 and hence there are exactly $|\partial_G S|$ non-zero entries in $1_S^T B$. Moreover, since G is Δ-regular, we have that $\sum_{j\in\partial_G S} u_j = \sum_{j\in[n]} u_j = 1_S^T B e_1 = \Delta|S|$. From the Cauchy–Schwarz inequality, it follows that

$$\begin{aligned}\big\|1_S^T B\big\|^2 &= \sum_{j\in[n]} u_j^2 \\ &= \sum_{j\in\partial S} u_j^2 \\ &\ge \frac{1}{|\partial_G S|}\Big(\sum_{j\in\partial S} u_j\Big)^2 \\ &= \frac{1}{|\partial_G S|}\Big(\sum_{j\in[n]} u_j\Big)^2 \\ &= \Delta^2|S|^2/|\partial_G S|. \end{aligned} \tag{2.4}$$

Combining (2.3) and (2.4) yields the desired inequality. □

2.5.3 Expansion properties of random graphs

In Chapter 5, we will prove several interesting asymptotic properties (local tree-likeness, expansion, etc.) of random graphs with a given degree sequence satisfying certain mild properties. For the more specialised case of random regular graphs, we state the following classical result, which motivates our interest in random graphs.

Theorem 2.5.8. *[36, Theorem 1.1] Let ε be a strictly positive real number and let Δ be a positive even integer. With high probability over the choice of a uniformly random n-vertex Δ-regular simple graph G, we have that $\lambda(G) \leq 2\sqrt{\Delta - 1} + \varepsilon$.*

2.6 Polymer models

Our presentation of polymer models follows mostly [68], but is slightly modified so that a polymer model is a function of both an underlying graph G and a host graph J_G. This will be for convenience, since it allows algorithms that operate on polymer models to have access to both the underlying graph and the host graph (since it is possible that, in some polymer models, certain information about the structure of G is lost when constructing J_G). Our polymer models are subset polymers as defined by Gruber and Kunz [56].

Let $\mathcal{G}$ be a class of graphs and let $q \geq 2$ be an integer. Given an underlying graph $G \in \mathcal{G}$ and a set of spins $[q]$, we construct a host graph J_G[6]. We assign to each vertex $v \in V_{J_G}$ a set of 'ground state' spins $g_v \subseteq [q]$. A polymer is a pair $\gamma = (V_\gamma, \sigma_\gamma)$ consisting of a J_G-connected set of vertices V_γ and an assignment $\sigma_\gamma \colon V_\gamma \to [q]$ such that $\sigma_\gamma(v) \in [q] \setminus g_v$ for all $v \in V_\gamma$. Let $\mathcal{P}_G$ be the set of all polymers. A polymer model for an underlying graph G and host graph J_G is defined by a set of allowed polymers $\mathcal{C}_G \subseteq \mathcal{P}_G$, and a weight function $w_G \colon \mathcal{C}_G \to \mathbb{R}_{\geq 0}$. For polymers $\gamma, \gamma' \in \mathcal{P}_G$, we write $\gamma \sim \gamma'$ to denote that γ, γ' are compatible, i.e., if $d_{J_G}(V_\gamma, V_{\gamma'}) > 1$ where $d_{J_G}(\cdot, \cdot)$ denotes the graph distance in J_G and for $S, T \subseteq V_{J_G}$ we let $d_{J_G}(S, T) = \min_{u \in S, v \in T} d_{J_G}(u, v)$. We define $\Omega_G = \{\Gamma \subseteq \mathcal{C}_G \mid \forall \gamma, \gamma' \in \Gamma, \gamma \sim \gamma'\}$ to be the set of all sets of mutually compatible polymers of $\mathcal{C}_G$; elements of Ω_G are called polymer configurations. The polymer model induces the partition function

$$Z_G = \sum_{\Gamma \in \Omega_G} \prod_{\gamma \in \Gamma} w_G(\gamma),$$

where $\prod_{\gamma \in \emptyset} w_G(\gamma) = 1$. The Gibbs distribution of the polymer model is the probability distribution μ_G on Ω_G, which we define as follows. For all $\Gamma \in \Omega_G$,

$$\mu_G(\Gamma) = \frac{\prod_{\gamma \in \Gamma} w_G(\gamma)}{Z_G},$$

[6]In many cases we will set $J_G = G$; however, other choices are of course possible.

We use $(\mathcal{C}_G, w_G, J_G)$ to denote the polymer model and we use $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ to denote the family of polymer models corresponding to the graph class $\mathcal{G}$. We say that this family of polymer models has degree bound Δ if the maximum degree of both G and J_G is at most Δ, for all $G \in \mathcal{G}$.

The main computational tasks associated with polymer models are approximately counting Z_G and approximately sampling from μ_G. We will consider two conditions on the weights of the polymers of a model, and their consequences for the aforementioned two tasks.

Definition 2.6.1. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}_\mathcal{G}$ satisfies the polymer mixing condition with constant $\theta \in (0, 1)$ if

$$\sum_{\gamma' \nsim \gamma} |V_{\gamma'}| \cdot w_G(\gamma') \leq \theta |V_\gamma|$$

for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{C}_G$.

In many cases, the above condition will be sufficient to prove the rapid mixing of certain Markov chains for sampling from the Gibbs distribution of a polymer model. In order to use these Markov chains as the basis for efficient approximate counting and sampling algorithms, we will require two further conditions.

Definition 2.6.2. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}_\mathcal{G}$ is computationally feasible if for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{P}_G$, it is possible to decide whether $\gamma \in \mathcal{C}_G$ and to compute $w_G(\gamma)$, if it is, in time polynomial in $|V_\gamma|$.

The following condition requires that the weights of a polymer decays exponentially in its size. Here, we measure the 'size' of a polymer by its number of vertices. We will also consider other notions of size, such as the total degree of a polymer (see Chapter 5 for more details).

Definition 2.6.3. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}$ satisfies the polymer sampling condition with constant $\tau > 0$ if

$$w_G(\gamma) \leq e^{-\tau |V_\gamma|},$$

for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{C}_G$.

Subject to the above two conditions, we will give efficient algorithms for approximating the partition function of a polymer model, and approximately sampling from its Gibbs distribution. See Chapter 3 for more details.

To see how the partition function of a spin system can be re-written as that of an appropriately defined polymer model, we present the following (rather trivial) example.

Example 1. One instance of a polymer model is the hard-core model itself. The underlying graph is $G = (V_G, E_G)$, as is the host graph; that is, $J_G = G$. Polymers are single vertices of the graph $G = (V_G, E_G)$, labeled with '1' (for occupied) against a ground state '0' (for unoccupied). Each polymer (vertex) $v \in V_G$ comes with the weight function $w_G(v) = \lambda$. Then the set of allowable polymer configurations is exactly the set of independent sets of G, and so the polymer model partition function is exactly the partition function of the hard-core model on G.

Chapter 3

Low temperature algorithms via Markov chains

The work in this chapter is based on Sections 1 to 4 of the following paper:

[22] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. *Random Struct. Algorithms*, 58(2):294–321, 2021. Theorems 5 and 6 from arxiv.org/abs/1901.0665

An extended abstract also appeared at:

[21] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik

Efficient algorithms for approximate counting and sampling in spin systems typically apply in the so-called high-temperature regime, where the interaction between neighboring spins is 'weak'. Instead, recent work of Jenssen, Keevash and Perkins yields polynomial-time algorithms in the low temperature regime on bounded-degree (bipartite) expander graphs using polymer models (as introduced in Section 2.6) and the cluster expansion. In this chapter, we present

a Markov chain for polymer models and show that it is rapidly mixing under exponential decay of polymer weights. This yields, for example, an $O(n \log n)$ time sampling algorithm for the low temperature ferromagnetic Potts model on bounded-degree expander graphs.

3.1 Introduction

While classical statistical physics is most interested in studying the hard-core model on the integer lattice $\mathbb{Z}^d$, the perspective of computer science is to consider wider families of graphs, such as the set of all graphs, all graphs of maximum degree Δ, or all bipartite graphs of maximum degree Δ.

Almost all proven efficient algorithms for approximate counting and sampling from the hard-core model work for low fugacities (the weak interaction regime, akin to the low temperature regime of the Potts model). In the high temperature regime there are at least three distinct algorithmic approaches to approximate counting and sampling: Markov chains, correlation decay, and polynomial interpolation. One striking advantage of the Markov chain approach is that the algorithms are much faster and simpler than the algorithms from the other approaches. In particular, it is common for a Markov chain sampling algorithm to run in time $O(n \log n)$, e.g., see [32, 33], while typical running times for algorithms based on correlation decay [91, 75] and polynomial interpolation [2] are of the form $n^{O(\log \Delta)}$ where Δ is the maximum degree of the graph.

In general there are no known efficient algorithms at low temperatures (high fugacities), but recently efficient algorithms have been developed for some special classes of graphs including subsets of $\mathbb{Z}^d$ [59], random regular bipartite graphs, and bipartite expander graphs in general [68, 74]. What these settings have in common is that for large enough λ, typical independent sets drawn from $\mu_{G,\lambda}$ align closely with one side or the other of the bipartition (the two ground states). This phenomenon is related to the phase transition phenomenon in infinite graphs, and implies the exponentially slow mixing time of local Markov chains [10, 47, 77]. The algorithms introduced in [59] exploit this phenomenon by expressing the partition function $Z_{G,\lambda}$ in terms of deviations from the two ground states, and then using a truncation of a convergent series expansion (the Taylor series of the cluster expansion) to approximate the log partition function.

In statistical physics this is called a perturbative approach, and while in general it does not work in the largest possible range of parameters, when it does work it gives a very detailed probabilistic understanding of the model [79, 13, 27].

To apply the so-called perturbative approach at low temperatures, one rewrites the original spin model as a new model in which single spin interactions are replaced by the interaction of connected components representing deviations from a chosen ground state. Such models are known in general as abstract polymer models, as introduced in Section 2.6, and have long been used in statistical physics to understand phase transitions. In this chapter, we show that once a low temperature spin model has been transformed into a polymer model, Markov chains once again become an effective algorithmic tool. Using this approach, we obtain nearly linear and quadratic time sampling algorithms for low temperature models on expander graphs in cases where only $n^{O(\log \Delta)}$ time algorithms were previously known.

3.2 Main results

Recall from Section 2.6 the definitions relating to the abstract polymer model framework. There are two main computational problems associated to a polymer model: approximately sampling from its Gibbs distribution μ_G, and approximating its partition function Z_G. To accomplish the former of these tasks, we first present a Markov chain (which we call the polymer dynamics) for sampling from the Gibbs distribution of a polymer model, and prove that it mixes rapidly subject to the polymer mixing condition (Definition 2.6.1).

Theorem 3.2.1. *Let $\mathcal{G}$ be a class of graphs and let $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of polymer models that satisfies the polymer mixing condition 2.6.1. For all $G \in \mathcal{G}$ and all $\varepsilon > 0$, the mixing time of the polymer dynamics is $T_{mix}(\varepsilon) = O(n \log(n/\varepsilon))$.*

Subject to further conditions on the weights of polymers, we show that this Markov chain can be used as the basis an efficient algorithm for approximately sampling from the Gibbs distribution of a polymer model.

Theorem 3.2.2. *Let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\mathcal{F}_{\mathcal{G}}$ be a family of computationally feasible (Definition 2.6.2) polymer*

models with degree bound Δ, that satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$.

There is an algorithm which takes as input a graph $G \in \mathcal{G}$ and an accuracy parameter $\varepsilon > 0$, and outputs an ε-sample from μ_G in $O(n\log(n/\varepsilon)\log(1/\varepsilon))$ time.

Finally, we can use the sampling algorithm to give an FPRAS for computing the partition function of a polymer model.

Theorem 3.2.3. *Let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\mathcal{F}_\mathcal{G}$ be a family of computationally feasible (Definition 2.6.2) polymer models with degree bound Δ, that satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$.*

There is an algorithm which takes as input a graph $G \in \mathcal{G}$ and an accuracy parameter $\varepsilon > 0$, and outputs an ε-approximation to Z_G, with running time $O((n/\varepsilon)^2\log^3(n/\varepsilon))$ and success probability at least $3/4$.

We apply our results for polymer model framework to two specific examples: the ferromagnetic Potts model and the hard-core model on expander graphs. For the low temperature Potts model on expanders, Jenssen, Keevash, and Perkins [68] gave an FPTAS and polynomial-time sampling algorithm, with an algorithm based on the cluster expansion and Barvinok's method of polynomial interpolation. Under essentially the same conditions on the parameters we give a Markov chain based sampling algorithm with near linear running time. Let $\mathcal{G}_{\Delta,\alpha}$ be the class of α-expander graphs (see Definition 2.5.2) of maximum degree at most Δ.

Theorem 3.2.4. *Let $q \geq 3$ be an integer, let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. There is an algorithm which takes as input a graph $G \in \mathcal{G}_{\Delta,\alpha}$ and an accuracy parameter $qe^{-n} \leq \varepsilon < 1$, and outputs an ε-sample from $\mu_{G,q,\beta}$ in $O(n\log(n/\varepsilon)\log(1/\varepsilon))$ time. There is an algorithm which takes as input a graph $G \in \mathcal{G}_{\Delta,\alpha}$ and an accuracy parameter $qe^{-n} \leq \varepsilon < 1$, and outputs an ε-approximation to $Z_{G,q,\beta}$, with running time $O((n/\varepsilon)^2\log^3(n/\varepsilon))$ and success probability at least $3/4$.*

Similarly, we give a Markov chain based algorithm for sampling from the high fugacity hard-core model for essentially the same range of parameters for which an FPTAS is given in [68]. Let $\mathcal{G}^{\text{bip}}_{\Delta,\alpha}$ be the set of all bipartite α-expander graphs (see Defintion 2.5.3) of maximum degree at most Δ.

Theorem 3.2.5. *Let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (3\Delta)^{6/\alpha}$ be a real number. There is an algorithm which takes as input a graph $G = (V_G^0, V_G^1, E_G) \in \mathcal{G}^{bip}_{\Delta,\alpha}$ and an accuracy parameter $4e^{-n} \leq \varepsilon < 1$, and outputs an ε-sample from $\mu_{G,\lambda}$. There is an algorithm which takes as input a graph $G = (V_G^0, V_G^1, E_G) \in \mathcal{G}^{bip}_{\Delta,\alpha}$ and an accuracy parameter $4e^{-n} \leq \varepsilon < 1$, and outputs an ε-approximation to $Z_{G,\lambda}$, with probability at least $1 - \varepsilon$. Both algorithms run in $O((n/\varepsilon)^2 \log^3(n/\varepsilon) \log(1/\varepsilon))$ time.*

The algorithms of Jenssen, Keevash, and Perkins [68] also apply to the low temperature Potts model and the high fugacity hard-core model on bounded degree expanders. For the Potts model, their algorithms apply when $\beta \geq \frac{4+2\log(q\Delta}{\alpha}$, whereas a lower bound of $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ is required by Theorem 3.2.4. For the hard-core model, their algorithms apply when $\lambda > C\Delta^{4/\alpha}$ (where C is an absolute constant), whereas a lower bound of $\lambda \geq (3\Delta)^{6/\alpha}$ is required by Theorem 3.2.5. At the expense of the marginally restricted range of parameters in which our algorithms apply, we improve the running time from the $n^{O(\log \Delta)}$ required by the algorithms presented in [68]. We also note that our algorithms are randomised, whereas those of [68] are deterministic. Subsequently to our work, Blanca, Cannon, and Perkins [6] improved the range of parameters and gave a randomised $O(n \log n)$ time algorithm for constructing an approximate sample from the ferromagnetic Potts model on bounded-degree expanders when $\beta \geq \frac{1+\log(\frac{\Delta+1}{e\Delta}+1)+\log((q-1)\Delta)}{\alpha}$.

3.3 Markov chains for polymer models

Recently, polymer models (see Section 2.6) have been used to develop efficient deterministic algorithms for sampling and approximating the partition functions of statistical physics models on lattices [59] and expander graphs [68, 74] at low temperatures, the regime in which Markov chains like the Glauber dynamics are known to mix slowly. In this section, we present some Markov chain based

algorithms for sampling from the Gibbs distribution of a polymer model, and approximating its partition function.

To begin, we compare various conditions on the weight functions of a polymer model, namely the Kotecký–Preiss [71] condition and the polymer sampling condition, and show that the latter implies the former. Then, we define the polymer Markov chain which we use to prove Theorems 3.2.1 and 3.2.2.

3.3.1 A comparison of the conditions on the weights

In what follows, let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and consider a family of q-spin polymer models $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ with degree bound Δ. Here we show that the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5+3\log((q-1)\Delta)$ implies the well-known Kotecký–Preiss [71] condition:

$$\sum_{\gamma' \nsim \gamma} e^{|V_{\gamma'}|} \cdot w_G(\gamma') \leq |V_\gamma|.$$

In previous work [59, 68], the above condition is used to prove the convergence of the cluster expansion – a certain power series which gives a close approximation to the log partition function of a polymer model.

To see the implication of the above condition, we use a lemma of Borgs, Chayes, Kahn, and Lovász.

Lemma 3.3.1. *[12, Lemma 2.1] Let $\Delta \geq 3$ be an integer, let $G = (V_G, E_G)$ be a graph of maximum degree at most Δ, and let $v \in V_G$. The number of connected induced subgraphs of G of size k containing v is at most $(e\Delta)^{k-1}$.*

Fix a polymer $\gamma \in \mathcal{C}_G$. We have that

$$\sum_{\gamma' \nsim \gamma} e^{|V_{\gamma'}|} \cdot w_G(\gamma') \leq \sum_{v \in V_\gamma \cup \partial_G V_\gamma} \sum_{k \geq 1} \sum_{\substack{\gamma' \in \mathcal{C}_G, \\ |V_{\gamma'}| = k, \\ v \in V_{\gamma'}}} e^k e^{-\tau k}.$$

In order to account for all of the polymers that we sum over in the above, we consider the connected induced subgraphs of J_G of size k that contain v, and

the assignments to them of $q-1$ colours. Using Lemma 3.3.1, we therefore obtain that

$$\begin{aligned}\sum_{\gamma' \nsim \gamma} e^{|V_{\gamma'}|} \cdot w_G(\gamma') &\leq |V_\gamma|(\Delta+1)\sum_{k\geq 1}(e\Delta)^{k-1}(q-1)^k e^k e^{-\tau k} \\ &= \frac{|V_\gamma|(\Delta+1)}{e\Delta}\sum_{k\geq 1}(e\Delta)^k(q-1)^k e^k e^{-\tau k} \\ &\leq \frac{|V_\gamma|(\Delta+1)}{e\Delta}\sum_{k\geq 1} e^{-3k} \\ &\leq |V_\gamma|,\end{aligned}$$

so the Kotecký–Preiss condition is satisfied.

The Kotecký–Preiss condition, in turn, implies the polymer mixing condition (2.6.1) with $\theta = 1/e$, since $ex \leq e^x$ for $x \geq 1$. For the same reason, it is easy to see that the polymer mixing condition is weaker than the Kotecký–Preiss condition.

3.3.2 The polymer Markov chain

For each $v \in V_G$, let $\mathcal{A}(v) = \{\gamma \in \mathcal{C}_G : v \in V_\gamma\}$ denote the collection of all polymers containing v, and let $a(v) = \sum_{\gamma\in\mathcal{A}(v)} w_G(\gamma)$. By applying the polymer mixing condition (Definition 2.6.1) to the smallest γ containing v we have $a(v) \leq \theta < 1$ for all $v \in V_G$. Define the probability distribution ν_v on $\mathcal{A}(v) \cup \{\emptyset\}$ by $\nu_v(\gamma) = w_G(\gamma)$ for $\gamma \in \mathcal{A}(v)$ and $\nu_v(\emptyset) = 1 - a(v)$.

The polymer dynamics on Ω_G are defined by the following transition rule from a configuration Γ_t to a configuration Γ_{t+1}:

Algorithm 1: Polymer dynamics

1. Choose $v \in V_G$ uniformly at random. Let $\gamma_v \in \Gamma_t \cap \mathcal{A}(v)$ if this is non-empty; otherwise, let $\gamma_v = \emptyset$. Note that γ_v is well defined since Γ_t can have at most one polymer containing v.
2. Mutually exclusively do the following:
   (a) With probability $\frac{1}{2}$, let $\Gamma_{t+1} = \Gamma_t \setminus \{\gamma_v\}$.
   (b) With probability $\frac{1}{2}$, sample $\boldsymbol{\gamma}$ from ν_v, let $\Gamma_{t+1} = \Gamma_t \cup \{\boldsymbol{\gamma}\}$ if this is in Ω; otherwise, let $\Gamma_{t+1} = \Gamma_t$.

Note that the polymer dynamics are aperiodic, since there are self-loops, and irreducible since we can transition from any $\Gamma \in \Omega_G$ to any $\Gamma' \in \Omega_G$ (e.g., via the empty set). Since the polymer dynamics are finite, irreducible, and aperiodic, they are also ergodic. Next, we observe that the stationary distribution of the polymer dynamics is μ_G by checking detailed balance. Note that each transition of the dynamics changes a configuration Γ by at most one polymer γ, therefore we let $\Gamma' = \Gamma \cup \{\gamma\}$. We have that

$$\frac{\mu_G(\Gamma')}{\mu_G(\Gamma)} = \frac{\prod_{\gamma' \in \Gamma'} w_G(\gamma')}{\prod_{\gamma' \in \Gamma} w_G(\gamma')} = w_G(\gamma) = \frac{\frac{|V_\gamma|}{n} \cdot \frac{1}{2} \cdot w_G(\gamma)}{\frac{|V_\gamma|}{n} \cdot \frac{1}{2}} = \frac{\Pr(\Gamma \to \Gamma')}{\Pr(\Gamma' \to \Gamma)},$$

It therefore follows, by detailed balance, that μ_G is the unique stationary distribution of the polymer dynamics.

3.3.3 Rapid mixing of the polymer dynamics

Theorem 3.2.1. *Let $\mathcal{G}$ be a class of graphs and let $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of polymer models that satisfies the polymer mixing condition 2.6.1. For all $G \in \mathcal{G}$ and all $\varepsilon > 0$, the mixing time of the polymer dynamics is $T_{mix}(\varepsilon) = O(n \log(n/\varepsilon))$.*

Proof. We will show that under the polymer mixing condition (Definition 2.6.1) the mixing time of the polymer dynamics is $O(n \log(n/\varepsilon))$ by applying the path coupling technique. We define a metric $D(\cdot, \cdot)$ on Ω_G by setting $D(\Gamma, \Gamma') = 1$ if $\Gamma' = \Gamma \cup \{\gamma\}$ or $\Gamma = \Gamma' \cup \{\gamma\}$, for some polymer γ, and extending this to a shortest path metric; i.e., $D(\Gamma, \Gamma') = |\Gamma \triangle \Gamma'|$ for any $\Gamma, \Gamma' \in \Omega_G$ where $\triangle$ denotes the symmetric difference of two sets.

Now suppose we couple two chains X_t and Y_t by attempting the same updates in both chains in each transition. Suppose that $X_t = Y_t \cup \{\gamma\}$ for some polymer γ. With probability $\frac{|V_\gamma|}{2n}$, we pick $v \in V_\gamma$ and remove γ_v which yields $X_{t+1} = Y_{t+1} = X_t$. On the other hand, we may attempt to add a polymer $\gamma' \nsim \gamma$ so that $Y_t \cup \{\gamma'\} \in \Omega_G$. That is, $X_{t+1} = X_t = Y_t \cup \{\gamma\}$ and $Y_{t+1} = Y_t \cup \{\gamma'\}$. This occurs with probability $\frac{|V_{\gamma'}|}{2n} \cdot w_G(\gamma')$ and in this case $D(X_{t+1}, Y_{t+1}) \leq 2$. Putting these together we can bound

$$\mathbb{E}[D(X_{t+1}, Y_{t+1})] \leq 1 + \frac{1}{2n} \left[-|V_\gamma| + \sum_{\gamma' \nsim \gamma} |V_{\gamma'}| \cdot w_G(\gamma') \right].$$

By the polymer mixing condition (Definition 2.6.1), we have that $\sum_{\gamma' \nsim \gamma} |V_{\gamma'}| \cdot w_G(\gamma') \leq \theta |V_\gamma|$ and so

$$\mathbb{E}[D(X_{t+1}, Y_{t+1})] \leq 1 - |V_\gamma| \frac{1-\theta}{2n} \leq 1 - \frac{1-\theta}{2n}.$$

By the path coupling (Lemma 2.4.5), and with W denoting the diameter of Ω_G under $D(\cdot, \cdot)$, we have that the mixing time is at most $\log(W/\varepsilon)2n/(1-\theta) = O(n \log(n/\varepsilon))$, using that $W \leq 2n$. This finishes the proof. □

To prove Theorem 3.2.2 we will show that a single update of the polymer dynamics can be computed in constant expected time. Recall that the polymer model is computationally feasible and that the polymer sampling condition (Definition 2.6.3) holds with constant $\tau \geq 5 + 3\log((q-1)\Delta)$. We will use the following algorithm. Let $r = \tau - 2 - \log((q-1)\Delta) \geq 3 + 2\log((q-1)\Delta)$ and let $\mathcal{A}_k(v) = \{\gamma \in \mathcal{A}(v) : |V_\gamma| \leq k\}$.

Algorithm 2: Single polymer sampler

1. Choose $\mathbf{k}$ according to the following geometric distribution: for k a non-negative integer,
$$\Pr[\mathbf{k} = k] = (1 - e^{-r})e^{-rk}.$$
This gives $\Pr[\mathbf{k} \geq k] = e^{-rk}$.
2. Enumerate all polymers in $\mathcal{A}_{\mathbf{k}}(v)$ and compute their weight functions.
3. Mutually exclusively output $\gamma \in \mathcal{A}_{\mathbf{k}}(v)$ with probability $w_G(\gamma) \cdot e^{r|V_\gamma|}$, and with all remaining probability output $\emptyset$. In particular if $\mathbf{k} = 0$, then output $\emptyset$ with probability 1.

In order to show that this algorithm has constant expected running time, we will require the following result on enumerating connected subgraphs of bounded degree graphs.

Lemma 3.3.2. *[78, Lemma 3.7] Let $\Delta \geq 3$ be an integer, let $G = (V_G, E_G)$ be a graph of maximum degree at most Δ, and let $v \in V_G$. There is an algorithm running in time $O(k^5(e\Delta)^{2k})$ that outputs a list of all connected subgraphs of G of size at most k containing v.*

We now proceed to prove the following lemma.

Lemma 3.3.3. *Let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}_\Delta\}$ be a family of computationally feasible (Definition 2.6.2) polymer models with degree bound Δ, that satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$. For all $G \in \mathcal{G}$ and all $v \in V_G$, the single polymer sampler samples a polymer from ν_v in expected constant time.*

Proof. We first show that the probabilities $w_G(\gamma) \cdot e^{r|V_\gamma|}$ sum to less than 1, which shows the last step of the sampling algorithm is well defined. Since $\tau - r = 2 + \log((q-1)\Delta)$, we have that

$$\sum_{\gamma \in \mathcal{A}(v)} w_G(\gamma) \cdot e^{r|V_\gamma|} \leq \frac{1}{2} \sum_{k \geq 1} (e\Delta)^{k-1} (q-1)^k e^{-\tau k + rk} = \frac{1}{2e\Delta} \sum_{k \geq 1} e^{-k} < 1.$$

We next show that the output of the algorithm has distribution ν_v. Given $\gamma \in \mathcal{A}(v)$, to output γ we must choose $\mathbf{k} \geq |V_\gamma|$. This happens with probability $e^{-r|V_\gamma|}$ by the distribution of $\mathbf{k}$. Conditioned on choosing such a $\mathbf{k}$, the probability we output γ is $w_G(\gamma) \cdot e^{r|V_\gamma|}$, and multiplying these probabilities together gives $w_G(\gamma)$ as desired. Since this is true for all $\gamma \in \mathcal{A}(v)$, the output distribution is exactly ν_v.

Finally we analyze the expected running time assuming that the model is computationally feasible. To do this, we observe that by Lemma 3.3.2, conditioned on the event that $\mathbf{k} = k$ the enumeration step of the algorithm takes time $O(k^5(e\Delta)^{2k})$, and the time to determine which polymers are allowed and computing their weights is $O(k^c(q-1)^k(e\Delta)^{k-1}/2)$ for some $c > 0$, since the polymer model is computationally feasible; here, the factor k^c accounts for the time to determine whether a single polymer of size k is 'allowed' and to compute its weight. Therefore, the expected running time is

$$\begin{aligned}
&O\left(1 + \sum_{k \geq 1} \Pr[\mathbf{k} = k] \left(k^5 (e\Delta)^{2k} + k^c (e(q-1)\Delta)^k\right)\right) \\
&= O\left(1 + \sum_{k \geq 1} e^{-rk} k^c (e(q-1)\Delta)^{2k}\right) \\
&= O\left(1 + \sum_{k \geq 1} k^c e^{-(\tau'+1)k}\right) \\
&= O(1),
\end{aligned}$$

where $\tau' = \tau - 5 - 3\log((q-1)\Delta) \geq 0$. □

Finally we prove the following theorem.

Theorem 3.2.2. *Let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\mathcal{F}_{\mathcal{G}}$ be a family of computationally feasible (Definition 2.6.2) polymer models with degree bound Δ, that satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$.*

There is an algorithm which takes as input a graph $G \in \mathcal{G}$ and an accuracy parameter $\varepsilon > 0$, and outputs an ε-sample from μ_G in $O(n\log(n/\varepsilon)\log(1/\varepsilon))$ time.

Proof. By Theorem 3.2.1, there is there is an integer $C_1 > 1$ (independent of n) so that if we start with the empty configuration $\Gamma_0 = \emptyset$ and run the polymer dynamics, then $\Gamma_{C_1\lceil n\log(n/\varepsilon)\rceil}$ has distribution within $\varepsilon/2$ total variation distance of μ_G. By Lemma 3.3.3, there is an integer $C_2 > 1$ (independent of n) such that the expected number of steps required to perform one update of the polymer dynamics is at most C_2. To compute an ε-sample from μ_G, we repeat the following $\lceil\log(2/\varepsilon)\rceil$ times, independently, and if no configuration is returned we return the empty configuration. Run the polymer dynamics for $3C_1C_2\lceil n\log(n/\varepsilon)\rceil$ steps starting from $\Gamma_0 = \emptyset$, and if at least $C_1\lceil n\log(n/\varepsilon)\rceil$ updates of the polymer dynamics were executed, return $\Gamma_{C_1\lceil n\log(n/\varepsilon)\rceil}$.

We next show that the probability that the algorithm does not timeout and return the empty configuration is at least $1 - \varepsilon/2$, which therefore yields that the output distribution has total variation distance at most ε from μ_G. Let X denote the total number of steps required to execute $C_1\lceil n\log(n/\varepsilon)\rceil$ updates of the polymer dynamics, and note that $\mathbb{E}[X] \leq C_1C_2\lceil n\log(n/\varepsilon)\rceil$. By Markov's inequality, it follows that $\Pr(X \geq 3\,\mathbb{E}[X]) < 1/e$. Thus, the probability that $X \geq 3\,\mathbb{E}[X]$ for each of $\lceil\log(2/\varepsilon)\rceil$ independent copies of X, is less than $(1/e)^{\log(2/\varepsilon)} = \varepsilon/2$. □

3.4 An approximate counting algorithm

In this section we show how to use a sampling oracle to approximately compute the partition function of the polymer model. One standard way is by self-reducibility. In [59] an efficient sampling algorithm for polymer models is derived

from an efficient approximate counting algorithm by applying self-reducibility on the level of polymers. While we could apply polymer self-reducibility in the other direction to obtain counting algorithms from our sampling algorithm, here we use the simulated annealing method instead (see [5, 61, 85]) to obtain a faster implementation of counting from sampling.

Let $q \geq 2$ and $\Delta \geq 3$ be integers, and let $\mathcal{G}$ be a class of graphs. Fix $G \in \mathcal{G}$ and fix a q-spin polymer model $(\mathcal{C}_G, w_G, J_G)$ from some computationally feasible family with degree bound Δ. Let ρ be a parameter and define a weight function

$$w_{G,\rho}(\gamma) = w_G(\gamma) \cdot e^{-\rho|V_\gamma|},$$

for all $\gamma \in \mathcal{C}_G$. For all $\rho \geq 0$, this defines a computationally feasible (Definition 2.6.2) polymer model $(\mathcal{C}_G, w_{G,\rho})$ on G, where setting $\rho = 0$ recovers the original model $(\mathcal{C}_G, w_G)$. If the original model $(\mathcal{C}_G, w_G)$ satisfies the polymer sampling condition (Definition 2.6.3), then so does $(\mathcal{C}_G, w_{G,\rho})$ for every $\rho \geq 0$ as the weight function $w_{G,\rho}(\gamma)$ is monotone decreasing in ρ.

Given the graph G, we write the partition function of the polymer model $(\mathcal{C}_G, w_{G,\rho})$ as follows

$$Z_{G,\rho} = \sum_{\Gamma \in \Omega_G} \prod_{\gamma \in \Gamma} w_{G,\rho}(\gamma) = \sum_{\Gamma \in \Omega_G} \prod_{\gamma \in \Gamma} w_G(\gamma) \cdot e^{-\rho|V_\gamma|}.$$

The associated Gibbs distribution is denoted by $\mu_{G,\rho}$. Since $\lim_{\rho\to\infty} w_{G,\rho}(\cdot) = 0$, we have that $\lim_{\rho\to\infty} Z_{G,\rho} = 1$ (only the empty configuration Γ contributes to this limit), and so we will use simulated annealing to interpolate between $Z_{G,\infty} = 1$ and our goal $Z_{G,0}$, assuming access to a sampling oracle for $(\mathcal{C}_G, w_{G,\rho})$ for all $\rho \geq 0$. To apply the simulated annealing method, roughly speaking, we find a sequence of parameters $0 = \rho_0 < \rho_1 < \cdots < \rho_\ell < \infty$ called a cooling schedule where $\ell \in \mathbb{Z}_{\geq 0}$, and then estimate $Z_{G,0}$ using the telescoping product

$$\frac{1}{Z_{G,0}} = \frac{1}{Z_{G,\rho_0}} = \frac{Z_{G,\rho_1}}{Z_{G,\rho_0}} \cdot \frac{Z_{G,\rho_2}}{Z_{G,\rho_1}} \cdots \frac{Z_{G,\rho_\ell}}{Z_{G,\rho_{\ell-1}}} \cdot \frac{1}{Z_{G,\rho_\ell}}.$$

To estimate each term $Z_{G,\rho_{i+1}}/Z_{G,\rho_i}$, we define independent random variables

$$W_i = \prod_{\gamma \in \Gamma_i} \frac{w_{G,\rho_{i+1}}(\gamma)}{w_{G,\rho_i}(\gamma)}, \qquad \text{where } \Gamma_i \sim \mu_{G,\rho_i}.$$

It is straightforward (see Lemma 3.4.1) to see that $\mathbb{E}[W_i] = Z_{G,\rho_{i+1}}/Z_{G,\rho_i}$. Using the sampling oracle for μ_{G,ρ_i}, we can sample W_i for all i, and by taking the product we get an estimate for $1/Z_{G,0}$.

The key ingredient of simulated annealing is finding a good cooling schedule. There are non-adaptive schedules [5] that depend only on n, and adaptive schedules [61, 85] that also depend on the structure of the Z_{G,ρ_i}. Usually, the latter leads to faster algorithms than the former. In this paper we will use a simple non-adaptive schedule: $\rho_i = i/n$ for $i = 0, \ldots, \ell$ where $\ell = O(n \log(n/\varepsilon))$. We will show that this cooling schedule already gives us a fast algorithm for the polymer model. The reason behind it is that the weight function $w_{G,\rho}(\gamma)$ decays exponentially fast, and so the partition function Z_{G,ρ_ℓ} is bounded by a constant when $\rho_\ell = O(\log n)$, leading to a short cooling schedule. See Lemma 3.4.2 for further details. Our algorithm is as follows.

Algorithm 3: Polymer approximate counting algorithm

---

1. Let $\rho_i = i/n$ for $i = 0, 1, \ldots, \ell$ where $\ell = \lceil n \log(4e(q-1)\Delta n/\varepsilon) \rceil$.
2. For $j = 1, \ldots, m$ where $m = \lceil 64\varepsilon^{-2} \rceil$:
   (a) For $0 \le i \le \ell - 1$:
      i. Sample $\Gamma_i^{(j)}$ from μ_{G,ρ_i}.
      ii. Let $W_i^{(j)} = \prod_{\gamma \in \Gamma_i^{(j)}} e^{-|V_\gamma|/n}$.
   (b) Let $W^{(j)} = \prod_{i=0}^{\ell-1} W_i^{(j)}$.
3. Let $\widehat{W} = \frac{1}{m} \sum_{j=1}^{m} W^{(j)}$ and output $\widehat{Z} = 1/\widehat{W}$.

Before proving Theorem 3.2.3, we first present a few useful lemmas. We shall use $\rho_i = i/n$ for $0 \le i \le \ell$ as our cooling schedule and we further define $\rho_{\ell+1} = (\ell+1)/n$ though it does not appear in the algorithm. For $0 \le i \le \ell - 1$ independently we define Γ_i to be a random sample from μ_{G,ρ_i} and $W_i = \prod_{\gamma \in \Gamma_i} e^{-|V_\gamma|/n}$. Finally, we let $W = \prod_{i=0}^{\ell-1} W_i$.

Lemma 3.4.1. *For* $0 \le i \le \ell - 1$,

$$\mathbb{E}[W_i] = \frac{Z_{G,\rho_{i+1}}}{Z_{G,\rho_i}} \qquad \text{and} \qquad \mathbb{E}[W_i^2] = \frac{Z_{G,\rho_{i+2}}}{Z_{G,\rho_i}}.$$

Therefore,

$$\mathbb{E}[W] = \frac{Z_{G,\rho_\ell}}{Z_{G,0}} \qquad \textit{and} \qquad \mathbb{E}[W^2] = \frac{Z_{G,\rho_\ell} Z_{G,\rho_{\ell+1}}}{Z_{G,0} Z_{G,\rho_1}}.$$

Proof. In the proof, we use $W_i(\Gamma_i)$ to denote $\prod_{\gamma\in\Gamma_i} \frac{w_{G,\rho_{i+1}}(\gamma)}{w_{G,\rho_i}(\gamma)}$. We deduce from the definition of W_i that

$$\begin{aligned}
\mathbb{E}[W_i] &= \sum_{\Gamma_i\in\Omega_G} \mu_{G,\rho_i}(\Gamma_i) W_i(\Gamma_i) \\
&= \frac{1}{Z_{G,\rho_i}} \sum_{\Gamma_i\in\Omega_G} \prod_{\gamma\in\Gamma_i} w_G(\gamma)\cdot e^{-i|V_\gamma|/n} \prod_{\gamma\in\Gamma_i} e^{-|V_\gamma|/n} \\
&= \frac{1}{Z_{G,\rho_i}} \sum_{\Gamma_i\in\Omega_G} \prod_{\gamma\in\Gamma_i} w_G(\gamma) e^{-(i+1)|V_\gamma|/n} = \frac{Z_{G,\rho_{i+1}}}{Z_{G,\rho_i}}
\end{aligned}$$

and that

$$\begin{aligned}
\mathbb{E}[W_i^2] &= \sum_{\Gamma_i\in\Omega_G} \mu_{G,\rho_i}(\Gamma_i) W_i(\Gamma_i)^2 \\
&= \frac{1}{Z_{G,\rho_i}} \sum_{\Gamma_i\in\Omega_G} \prod_{\gamma\in\Gamma_i} w_G(\gamma)\cdot e^{-i|V_\gamma|/n} \prod_{\gamma\in\Gamma_i} e^{-2|V_\gamma|/n} \\
&= \frac{1}{Z_{G,\rho_i}} \sum_{\Gamma_i\in\Omega_G} \prod_{\gamma\in\Gamma_i} w_G(\gamma)\cdot e^{-(i+2)|V_\gamma|/n} = \frac{Z_{G,\rho_{i+2}}}{Z_{G,\rho_i}}.
\end{aligned}$$

Since $W_0,\dots,W_{\ell-1}$ are mutually independent, we obtain that

$$\mathbb{E}[W] = \prod_{i=0}^{\ell-1} \mathbb{E}[W_i] = \prod_{i=0}^{\ell-1} \frac{Z_{G,\rho_{i+1}}}{Z_{G,\rho_i}} = \frac{Z_{G,\rho_\ell}}{Z_{G,\rho_0}}$$

and that

$$\mathbb{E}[W^2] = \prod_{i=0}^{\ell-1} \mathbb{E}[W_i^2] = \prod_{i=0}^{\ell-1} \frac{Z_{G,\rho_{i+2}}}{Z_{G,\rho_i}} = \frac{Z_{G,\rho_\ell} Z_{G,\rho_{\ell+1}}}{Z_{G,\rho_0} Z_{G,\rho_1}}.$$

□

Lemma 3.4.2. *Suppose that $w_G(\gamma) \le 1$ for all $\gamma \in \mathcal{C}_G$. Then we have that*

$$1 \le Z_{G,\rho_\ell} \le e^{\varepsilon/2}.$$

Proof. It is trivial that $Z_{G,\rho_\ell} \ge 1$ since $\emptyset \in \Omega_G$ has weight 1. Meanwhile, we have the crude bound

$$Z_{G,\rho_\ell} \le \prod_{\gamma\in\mathcal{C}_G} \left(1 + w_G(\gamma)\cdot e^{-\ell|V_\gamma|/n}\right).$$

We then deduce that

$$\log(Z_{G,\rho_\ell}) \le \sum_{\gamma\in\mathcal{C}_G} w_G(\gamma)\cdot e^{-\ell|V_\gamma|/n} \le \sum_{v\in V_G}\sum_{k\ge 1}\sum_{\substack{\gamma\in\mathcal{C}_G,\\ |V_\gamma|=k,\\ v\in V_\gamma}} \left(e^{-\ell/n}\right)^k$$
$$\overset{(a)}{\le} n\sum_{k\ge 1}\left(\frac{e(q-1)\Delta}{e^{\ell/n}}\right)^k \overset{(b)}{\le} n\sum_{k\ge 1}\left(\frac{\varepsilon}{4n}\right)^k \le n\cdot\frac{2\varepsilon}{4n} = \frac{\varepsilon}{2}$$

where (a) follows from Lemma 3.3.1 and (b) from $\ell \ge n\log(4e(q-1)\Delta n/\varepsilon)$. □

Lemma 3.4.3. *We have*

$$\frac{Z_{G,\rho_1}}{Z_{G,0}} \ge \frac{1}{e} \qquad \text{and} \qquad \frac{Z_{G,\rho_{\ell+1}}}{Z_{G,\rho_\ell}} \le 1.$$

Proof. Since the weight function $w_{G,\rho}(\gamma)$ is decreasing in ρ, the partition function $Z_{G,\rho}$ is also decreasing, which implies $Z_{G,\rho_{\ell+1}} \le Z_{G,\rho_\ell}$. On the other hand, recalling Lemma 3.4.1, we have

$$\frac{Z_{G,\rho_1}}{Z_{G,0}} = \mathbb{E}[W_0] = \mathbb{E}\left[\prod_{\gamma\in\Gamma_0} e^{-|V_\gamma|/n}\right]$$

where Γ_0 is sampled from μ_{G,ρ_0}. Notice that for any $\Gamma_0 \in \Omega_G$ we have

$$\prod_{\gamma\in\Gamma_0} e^{-|V_\gamma|/n} = \exp\left(-\frac{1}{n}\sum_{\gamma\in\Gamma_0}|V_\gamma|\right) \ge \frac{1}{e}.$$

Thus, the lemma follows. □

We are now ready to prove the following result.

Theorem 3.2.3. *Let $q \ge 2$ and $\Delta \ge 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\mathcal{F}_\mathcal{G}$ be a family of computationally feasible (Definition 2.6.2) polymer models with degree bound Δ, that satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \ge 5 + 3\log((q-1)\Delta)$.*

There is an algorithm which takes as input a graph $G \in \mathcal{G}$ and an accuracy parameter $\varepsilon > 0$, and outputs an ε-approximation to Z_G, with running time $O((n/\varepsilon)^2\log^3(n/\varepsilon))$ and success probability at least $3/4$.

Proof. We first assume that we have access to an exact sampler $\mathcal{S}_{\mathsf{exact}}$ that samples from $\mu_{G,\rho}$ for all $\rho \geq 0$. Using this sampler in the Polymer approximate counting algorithm, we find that, for each j and each i, $\Gamma_i^{(j)}$ is an exact sample from the distribution μ_{G,ρ_i} and hence $W_i^{(j)}$ is an exact sample of W_i, independently for every j and i. Thus, $W^{(j)}$ is a sample of W independently for every j, and $\widehat{W}$ is the sample mean of $W^{(j)}$'s. We deduce from Lemmas 3.4.1 and 3.4.2 that

$$(1+\varepsilon/2)\,\mathbb{E}[W] \leq \frac{e^{\varepsilon/2} Z_{G,\rho_\ell}}{Z_{G,0}} \leq \frac{e^{\varepsilon}}{Z_{G,0}}$$

and

$$(1-\varepsilon/2)\,\mathbb{E}[W] \geq \frac{e^{-\varepsilon} Z_{G,\rho_\ell}}{Z_{G,0}} \geq \frac{e^{-\varepsilon}}{Z_{G,0}}$$

where we use $1+\varepsilon/2 \leq e^{\varepsilon/2}$ and $e^{-\varepsilon} \leq 1-\varepsilon/2$ for all $0<\varepsilon<1$. Then

$$\Pr\left(\frac{e^{-\varepsilon}}{Z_{G,0}} \leq \widehat{W} \leq \frac{e^{\varepsilon}}{Z_{G,0}}\right) \geq \Pr\left(\left|\widehat{W} - \mathbb{E}[W]\right| \leq (\varepsilon/2)\,\mathbb{E}[W]\right).$$

By Chebyshev's inequality we have

$$\Pr\left(\left|\widehat{W} - \mathbb{E}[W]\right| \geq (\varepsilon/2)\,\mathbb{E}[W]\right) \leq \frac{4\,\mathrm{Var}(W)}{\varepsilon^2 m\,(\mathbb{E}[W])^2} \leq \frac{4(e-1)}{\varepsilon^2 m} \leq \frac{1}{8}$$

where the second to last inequality follows from Lemmas 3.4.1 and 3.4.3:

$$\frac{\mathrm{Var}(W)}{(\mathbb{E}[W])^2} = \frac{\mathbb{E}[W^2]}{(\mathbb{E}[W])^2} - 1 = \frac{Z_{G,0}}{Z_{G,\rho_1}} \frac{Z_{G,\rho_{\ell+1}}}{Z_{G,\rho_\ell}} - 1 \leq e - 1.$$

Thus, we deduce that

$$\Pr\left(e^{-\varepsilon} Z_{G,0} \leq \widehat{Z} \leq e^{\varepsilon} Z_{G,0}\right) = \Pr\left(\frac{e^{-\varepsilon}}{Z_{G,0}} \leq \hat{W} \leq \frac{e^{\varepsilon}}{Z_{G,0}}\right) \geq \frac{7}{8},$$

so the error probability is at most $1/8$. Note that the number of samples that we used is ℓm.

Now we replace the exact sampling oracle $\mathcal{S}_{\mathsf{exact}}$ by an approximate one. For every $\rho \geq 0$, the polymer model $(\mathcal{C}_G, w_{G,\rho})$ is computationally feasible and satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$. Thus, for any $\rho \geq 0$, Theorem 3.2.2 gives a randomised algorithm $\mathcal{S}$ that outputs a $1/(8\ell m)$-approximate sample from $\mu_{G,\rho}$. We then couple $\mathcal{S}$ and $\mathcal{S}_{\mathsf{exact}}$ optimally and run the algorithm with both $\mathcal{S}$ and $\mathcal{S}_{\mathsf{exact}}$ simultaneously, so that for any $\rho \geq 0$ samples from $\mathcal{S}$ and $\mathcal{S}_{\mathsf{exact}}$ for $\mu_{G,\rho}$ coincide

with probability at least $1 - 1/(8\ell m)$. Let $\mathcal{B}$ be the event that at least one of the ℓm samples from $\mathcal{S}$ in the algorithm does not couple with that from $\mathcal{S}_{\mathsf{exact}}$. Then a union bound yields $\Pr(\mathcal{B}) \leq 1/8$. Let $\mathcal{D}$ be the event that the algorithm using $\mathcal{S}_{\mathsf{exact}}$ fails. From our argument before we see that $\Pr(\mathcal{D}) \leq 1/8$. Note that if neither of $\mathcal{B}$ and $\mathcal{D}$ happens, then the algorithm with $\mathcal{S}$ will output a desired estimate. Hence, we conclude from the union bound that the algorithm with $\mathcal{S}$ fails with probability at most

$$\Pr(\mathcal{B}) + \Pr(\mathcal{D}) \leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Finally, we consider the running time of our algorithm. By Theorem 3.2.2 the running time of 2(a)i is $O(n \log(8\ell mn) \log(8\ell m)) = O(n \log^2(n/\varepsilon))$, and for step 2(a)ii the running time is $O(n)$. Thus, the running time of the algorithm is upper bounded by $\ell m \cdot O(n \log^2(n/\varepsilon)) = O((n/\varepsilon)^2 \log^3(n/\varepsilon))$. □

3.5 Polymer models for spin systems

Our eventual aim will be to use the algorithmic results of the previous two sections, in order to sample from the Gibbs distribution, and approximate the partition function, of an appropriately defined polymer model for some spin system. To this end, we first show how to capture the ferromagnetic Potts and hard-core models (defined on arbitrary graphs and bipartite graphs, respectively) as polymer models.

3.5.1 Ferromagnetic Potts model

Let $q \geq 2$ be an integer, let $\beta > 0$ be a real number, and let $G = (V_G, E_G)$ be an n-vertex graph. Fix a colour from $[q]$ – we'll call this red. We define a polymer model for the q-colour Potts model with ground state red. The underlying graph and the host graph for the polymer model are both G. We define the allowed polymers to be the polymers containing at most $n/2$ vertices; that is, the connected subgraphs of G of size at most $n/2$, with vertices labeled by colours from $[q] \setminus \{\text{red}\}$. A polymer γ has weight function $w_G^{\text{red}}(\gamma) = e^{-\beta B(\gamma)}$ where $B(\gamma)$ is the number of bichromatic edges internal to V_γ plus the number of edges of E_G connecting V_γ to $V_G \setminus V_\gamma$.

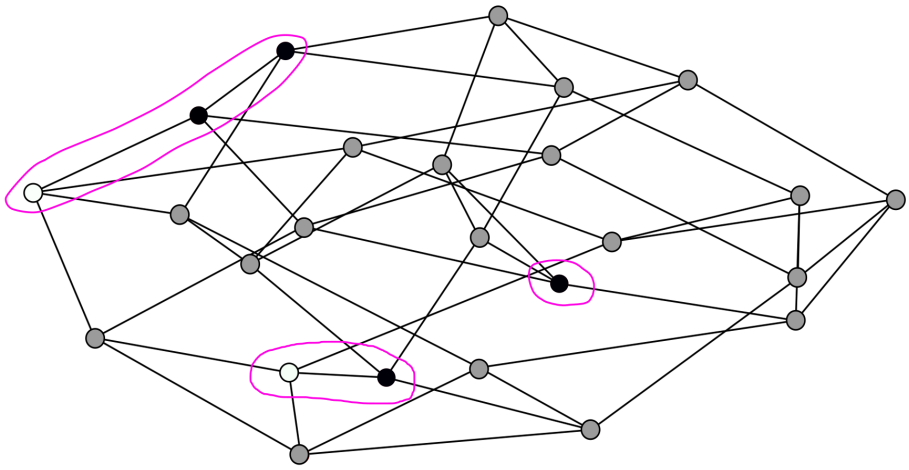


Figure 3.1: A polymer configuration for the grey phase of the Potts model.

We let $\mathcal{P}_G^{\text{red}}$ denote the set of all polymers and we let $\mathcal{C}_G^{\text{red}}$ denote the set of all allowed polymers. We let $\hat{\Omega}_G^{\text{red}}$ denote the set of all mutually compatible sets of allowed polymers, we let $\hat{Z}_G^{\text{red}}$ denote the partition function of the polymer model, and we let $\hat{\mu}_G^{\text{red}}$ denote its Gibbs distribution.

Note that a configuration of compatible polymers maps to a unique Potts configuration σ in which all connected components of non-red-coloured vertices have size at most $n/2$, and the weight of σ in the Potts model is exactly the product of the weight functions of the polymers. The polymer model partition function represents the contribution to the Potts model partition function of colourings where the colour red 'dominates', see also Section 3.6.1 for more details.

3.5.2 Hard-core model

Let $\lambda > 1$ be a real number and let $G = (V_G^0, V_G^1, E_G)$ be a bipartite graph. Here, we will consider two ground states, corresponding to V_G^0 and V_G^1. We will capture deviations from the two ground states using the 'even' and 'odd' polymer models of Jenssen, Keevash and Perkins [68]. We remark that similar models were considered independently by Liao, Lin, Lu, and Mao [74].

Fix $i \in \{0, 1\}$. We define a polymer model for the hard-core model with ground state corresponding to V_G^i. The underlying graph for the polymer model is G and the host graph J_G for the polymer model is G^2– the graph defined on

vertex set $V_G^0 \cup V_G^1$, in which two vertices are connected if the graph distance between them in G is at most 2. Note that since the host graph is G^2, polymers are G^2-connected vertex subsets, and they are compatible if the graph distance between them in G^2 is strictly greater than 1. The set of spins is $\{0, 1\}$, and the ground state spin for a vertex $v \in V_G$ is $g_v = 0$ if $v \in V_G^i$, and $g_v = 1$ if $v \in V_G^{1-i}$. We say a vertex set $S \subseteq V_G^i$ is *small* if $|S| \leq |V_G^i|/2$. We define the allowed polymers to be the polymers $\gamma = (V_\gamma, \sigma_\gamma)$ for which $V_\gamma \subseteq V_G^i$ and V_γ is small. Note that the spin assignment σ_γ for an allowed polymer γ assigns the spin 1 to each $v \in V_\gamma$. We let $\mathcal{P}_G^i$ denote the set of all polymers, we let $\mathcal{C}_G^i$ denote the set of all allowed polymers, and we let Ω_G^i denote the set of all mutually compatible sets of polymers. The weight of a polymer γ is defined to be

$$w_G^i(\gamma) = \frac{\lambda^{|V_\gamma|}}{(1+\lambda)^{|\partial_G V_\gamma|}}. \tag{3.1}$$

The key observation behind the definition of the weights is that for a set $\Gamma \in \Omega_G^i$ of compatible polymers from $\mathcal{C}_G^i$, the contribution to the hard-core partition function $Z_{G,\lambda}$ of all independent sets $I \in \mathcal{I}_G$ with $I \cap V_G^i = \bigcup_{\gamma \in \Gamma} V_\gamma$ is exactly

$$(1+\lambda)^{|V_G^{1-i}|} \prod_{\gamma \in \Gamma} w_G^i(\gamma),$$

see [66, Proof of Lemma 19] for further details.

We let $\hat{Z}_G^i$ denote the partition function of the polymer model and we let $\hat{\mu}_G^i$ denote its Gibbs distribution.

3.6 Efficient low temperature algorithms

In this section, we combine the approximate counting and sampling algorithms that we gave in Sections 3.3 and 3.4, with the polymer models for the ferromagnetic Potts and hard-core models that we described in Section 3.5.

3.6.1 Ferromagnetic Potts model on bounded-degree expanders

Let $q \geq 3$ be an integer, let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. Recall that $\mathcal{G}_{\Delta,\alpha}$ denotes the class of α-expander graphs of maximum degree at most Δ. We give an FPRAS and

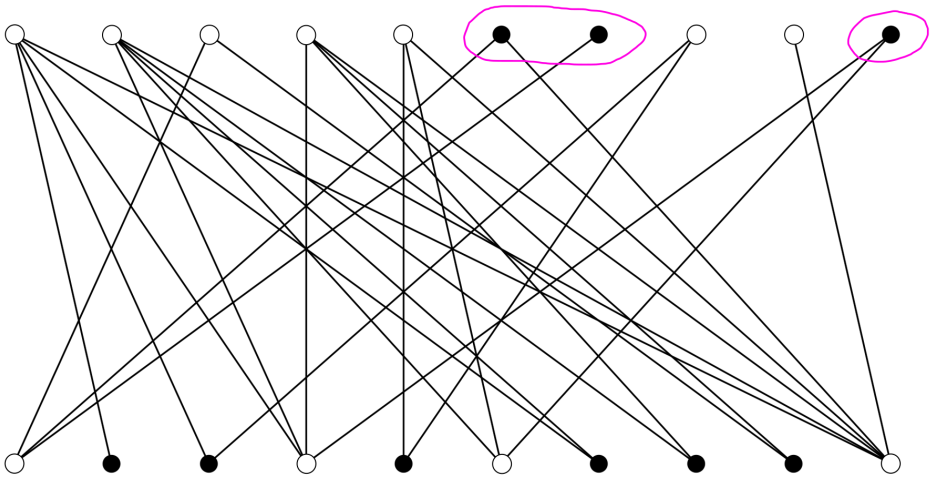


Figure 3.2: A polymer configuration for the all-even occupied phase of the hard-core model.

an efficient sampling algorithm for the q-colour ferromagnetic Potts model at inverse temperature β, on $\mathcal{G}_{\Delta,\alpha}$.

Consider the polymer model for the ferromagnetic Potts model which we introduced in 3.5.1, defined over the class $\mathcal{G}_{\Delta,\alpha}$. The following result is due to Jenssen, Keevash, and Perkins, and it shows that on a bounded-degree expander graph, the partition function of the low temperature Potts model is well-approximated by that of the polymer model.

Lemma 3.6.1. *[66, Lemma 12] Let $q \geq 3$ be an integer, let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\beta \geq 2\log(eq)/\alpha$ be a real number. For all $g \in [q]$ and all $G \in \mathcal{G}_{\Delta,\alpha}$, it follows that $q\hat{Z}_G^{red}$ is an e^{-n}-approximation to $Z_{G,q,\beta}$.*

This result tells us that in order to approximate the partition function of the Potts model, it suffices to approximate the partition function of the polymer model. To achieve this, we appeal to the algorithms presented in Sections 3.3 and 3.4.

Theorem 3.2.4. *Let $q \geq 3$ be an integer, let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. There is an algorithm which takes as input a graph $G \in \mathcal{G}_{\Delta,\alpha}$ and an accuracy parameter $qe^{-n} \leq \varepsilon < 1$, and outputs an ε-sample from $\mu_{G,q,\beta}$ in $O(n\log(n/\varepsilon)\log(1/\varepsilon))$ time. There is*

an algorithm which takes as input a graph $G \in \mathcal{G}_{\Delta,\alpha}$ and an accuracy parameter $qe^{-n} \leq \varepsilon < 1$, and outputs an ε-approximation to $Z_{G,q,\beta}$, with running time $O((n/\varepsilon)^2 \log^3(n/\varepsilon))$ and success probability at least $3/4$.

Proof. Consider the family of polymer models $\mathcal{F}^{\text{red}}_{\mathcal{G}_{\Delta,\alpha}} = \{(\mathcal{C}^{\text{red}}_G, w^{\text{red}}_G, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$ defined in Section 3.5. Note that it has degree bound Δ. Let $G \in \mathcal{G}_{\Delta,\alpha}$ and let $\gamma \in \mathcal{C}^{\text{red}}_G$ be a polymer. By expansion, we have that $B(\gamma) \geq e_G(V_\gamma, V_\gamma^c) \geq \alpha|V_\gamma|$, and therefore that $w^{\text{red}}_G(\gamma) \leq e^{-\alpha\beta|V_\gamma|}$. It therefore follows that $\mathcal{F}^{\text{red}}_{\mathcal{G}_{\Delta,\alpha}}$ satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau = \alpha\beta \geq 5 + 3\log((q-1)\Delta)$. Clearly, this polymer model is also computationally feasible, therefore Theorems 3.2.2 and 3.2.3 apply. Furthermore, since $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha} > \frac{2\log(eq)}{\alpha}$, Lemma 3.6.1 also applies.

For the sampling algorithm, we pick a color $g \in [q]$ (we can assume that this is red, by symmetry) uniformly at random and generate an (ε/q)-approximate sample from $\hat{\mu}^{\text{red}}_G$ using the algorithm of Theorem 3.2.2, in $O(n\log(n/\varepsilon)\log(1/\varepsilon))$ time. By Lemma 3.6.1, we conclude that the resulting output is an ε-approximate sample from $\mu_{G,q,\beta}$.

For the counting algorithm, we produce, using the algorithm of Theorem 3.2.3, a number $\hat{Z}$ in $O((n/\varepsilon)^2 \log^3(n/\varepsilon))$ time, which is an $\varepsilon/(2q)$-approximation to $\hat{Z}^{\text{red}}_G$ with probability at least $3/4$. By Lemma 3.6.1, we conclude that $q\hat{Z}$ is an ε-approximation to $Z_{G,q,\beta}$, with the same probability. □

The lower bound on ε in the statement of the above theorem is present so that we can apply, for example, Lemma 3.6.1. For smaller values of ε, we can instead compute the partition function by brute force in $\text{poly}(n/\varepsilon)$ time. This observation combined with the above result gives an FPRAS, but we can no longer guarantee a running time of $O((n/\varepsilon)^2 \log^3(n/\varepsilon))$ for exponentially small values of ε. A similar point also applies to the algorithm that we will give for the hard-core model.

3.6.2 Hard-core model on bounded-degree bipartite expanders

Let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (3\Delta)^{6/\alpha}$ be a real number. We give an FPRAS and an efficient sampling algorithm for the hard-core model at fugacity λ, on $\mathcal{G}^{\text{bip}}_{\Delta,\alpha}$.

Consider the polymer model for the hard-core model which we introduced in Section 3.5.2, defined over the class $\mathcal{G}^{\text{bip}}_{\Delta,\alpha}$. The following result is due to Jenssen, Keevash, and Perkins [68], and it shows that on a bounded-degree bipartite expander graph, the partition function of the high-fugacity hard-core model is well-approximated by that of the polymer model.

Lemma 3.6.2. *[66, Lemma 19] Let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (3\Delta)^{6/\alpha}$ be a real number. For all $G = (V^0_G, V^1_G, E_G) \in \mathcal{G}^{bip}_{\Delta,\alpha}$, it follows that*

$$(1+\lambda)^{|V^1_G|}\hat{Z}^0_G + (1+\lambda)^{|V^0_G|}\hat{Z}^1_G,$$

is an e^{-n}-approximation to $Z_{G,\lambda}$.

In particular, [67, Lemma 17] shows that $(1+\lambda)^{|V^1_G|}\hat{Z}^0_G + (1+\lambda)^{|V^0_G|}\hat{Z}^1_G$ counts the contribution to $Z_{G,\lambda}$ of every independent set $I \in \mathcal{I}_G$, but some independent sets are double counted: those independent sets I for which the 2-connected components of $V^0_G \cap I$ and $V^1_G \cap I$ are all small. We call these independent sets *sparse*. The proof of [66, Lemma 19] shows that the relative contribution to $Z_{G,\lambda}$ of sparse independent sets is at most e^{-n}.

This result tells us that in order to approximate the partition function of the hard-core model, it suffices to approximate the partition function of the polymer model. To achieve this, we appeal to the algorithms presented in Sections 3.3 and 3.4.

Theorem 3.2.5. *Let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (3\Delta)^{6/\alpha}$ be a real number. There is an algorithm which takes as input a graph $G = (V^0_G, V^1_G, E_G) \in \mathcal{G}^{bip}_{\Delta,\alpha}$ and an accuracy parameter $4e^{-n} \leq \varepsilon < 1$, and outputs an ε-sample from $\mu_{G,\lambda}$. There is an algorithm which takes as input a graph $G = (V^0_G, V^1_G, E_G) \in \mathcal{G}^{bip}_{\Delta,\alpha}$ and an accuracy parameter $4e^{-n} \leq \varepsilon < 1$, and outputs an ε-approximation to $Z_{G,\lambda}$, with probability at least $1-\varepsilon$. Both algorithms run in $O((n/\varepsilon)^2 \log^3(n/\varepsilon)\log(1/\varepsilon))$ time.*

Proof. First note that $\lambda \geq (3\Delta)^{6/\alpha} \geq 9^{6/\alpha} > e^{11/\alpha}$, so Lemma 3.6.2 applies. Fix arbitrary $i \in \{0,1\}$. Noting that the family of polymer models $\{(\mathcal{C}^i_G, w^i_G, G^2) \mid G \in \mathcal{G}^{\text{bip}}_{\Delta,\alpha}\}$ is computationally feasible and has degree bound Δ^2, we verify the

polymer sampling condition (Definition 2.6.3). As in [66, Section 4.2], we have the bound

$$w_G^i(\gamma) = \frac{\lambda^{|V_\gamma|}}{(1+\lambda)^{|\partial_G V_\gamma|}} \leq \frac{\lambda^{|V_\gamma|}}{(1+\lambda)^{(1+\alpha)|V_\gamma|}} \leq \lambda^{-\alpha|V_\gamma|},$$

so, using that $\lambda \geq (3\Delta)^{6/\alpha}$, we have that the models satisfy the polymer sampling condition (Definition 2.6.3) with $\tau = \alpha \log \lambda \geq 6\log(3\Delta) \geq 5 + 3\log(\Delta^2)$. Therefore, we may also apply Theorems 3.2.2 and 3.2.3.

For the counting algorithm, we apply Theorem 3.2.3. Namely, by taking the median of $O(\log(1/\varepsilon))$ trials, we can obtain $\tilde{Z}^0$ and $\tilde{Z}^1$ which are $(\varepsilon/32)$-approximations to $\hat{Z}_G^0$ and $\hat{Z}_G^1$, respectively, with probability at least $1 - \varepsilon/32$. Let $\mathcal{E}$ be the event that $\tilde{Z}^0$ and $\hat{Z}^1$ are indeed $(\varepsilon/32)$-approximations to $\hat{Z}_G^0$ and $\hat{Z}_G^1$, respectively. Conditioned on $\mathcal{E}$, the number

$$\tilde{Z} = (1+\lambda)^{|V_G^1|}\tilde{Z}^0 + (1+\lambda)^{|V_G^0|}\tilde{Z}^1$$

is an $(\varepsilon/32)$-approximation to the number $A = (1+\lambda)^{|V_G^1|}\hat{Z}_G^0 + (1+\lambda)^{|V_G^0|}\hat{Z}_G^1$. By Lemma 3.6.2 and since $\varepsilon \geq 4e^{-n}$, we have that A is an $(\varepsilon/4)$-approximation to $\hat{Z}_{G,\lambda}$ and hence $\tilde{Z}$ is an ε-approximation to $\hat{Z}_{G,\lambda}$. Since $\mathcal{E}$ occurs with probability at least $1-\varepsilon/16$, we obtain that $\tilde{Z}$ is the desired approximation for the counting algorithm.

For the sampling algorithm, let $\mathbf{i}$ be the random variable which takes the value 0 with probability $\frac{(1+\lambda)^{|V_G^1|}\tilde{Z}^0}{\tilde{Z}}$ and the value 1 otherwise, where $\tilde{Z}^0, \tilde{Z}^1, \tilde{Z}$ are the quantities computed earlier. Then, use Theorem 3.2.2 to obtain an $(\varepsilon/8)$-approximate sample from the Gibbs distribution corresponding to the polymer model $(\mathcal{C}_G^{\mathbf{i}}, w_G^{\mathbf{i}}, G^2)$, say $\tilde{\Gamma}^{\mathbf{i}}$. Obtain then an independent set $\tilde{I}$ by including each $v \in V_G^{1-\mathbf{i}} \backslash \partial_G(\bigcup_{\gamma\in\tilde{\Gamma}^{\mathbf{i}}} V_\gamma)$ in $\tilde{I}$ with probability $\frac{\lambda}{1+\lambda}$ and each vertex in $\bigcup_{\gamma\in\tilde{\Gamma}^{\mathbf{i}}} V_\gamma$ (with probability 1). We claim that the output distribution of $\tilde{I}$ is an ε-sample from the hard-core distribution $\mu_{G,\lambda}$.

To prove this, consider the random independent set I obtained by repeating the same steps above but using instead perfectly accurate computations, i.e., pick $i = 0$ with probability $\frac{(1+\lambda)^{|V_G^1|}\hat{Z}_G^0}{A}$ and the value 1 otherwise, then, sample (perfectly) $\hat{\Gamma}^i$ from the Gibbs distribution corresponding to the polymer model $(\mathcal{C}_G^i, w_G^i, G^2)$, and then obtain the independent set I by including each $v \in V_G^{1-i} \backslash \partial_G(\bigcup_{\gamma\in\hat{\Gamma}^i} V_\gamma)$ in I with probability $\frac{\lambda}{1+\lambda}$ and each vertex in $\bigcup_{\gamma\in\hat{\Gamma}^i} V_\gamma$

(with probability 1). Then, if I is not sparse, I is generated with probability $\lambda^{|I|}/A$ (cf. the observation below (3.1)). On the other hand, if I is sparse, then I is generated with probability $2\lambda^{|I|}/A$. But by Lemma 3.6.2 and the remark following, the total variation distance between the distribution of I and the hard-core distribution $\mu_{G,\lambda}$ is bounded by the relative weight of the sparse independent sets, which, by Lemma 3.6.2, is at most $e^{-n} \leq \varepsilon/4$.

We next observe that, conditioned on the event $\mathcal{E}$ (i.e., that $\tilde{Z}^0$ and $\tilde{Z}^1$ are $(\varepsilon/32)$-approximations to $\hat{Z}^0_G$ and $\hat{Z}^1_G$, respectively), there is a coupling between $\tilde{I}$ and I such that $\tilde{I} = I$ with probability at least $1 - \varepsilon/4$. Indeed, the total variation distance between $\mathbf{i}$ and i is at most $e^{\varepsilon/16} - 1 \leq \varepsilon/8$ and hence there is a coupling of $\mathbf{i}$ with i so that $\mathbf{i} = i$ with probability at least $1 - \varepsilon/8$. Analogously, there is a coupling of $\tilde{\Gamma}^{\mathbf{i}}$ with $\hat{\Gamma}^i$ so that $\tilde{\Gamma}^{\mathbf{i}} = \hat{\Gamma}^i$ with probability at least $1 - \varepsilon/8$. Since $\mathcal{E}$ occurs with probability at least $1 - \varepsilon/16$, it follows that the overall total variation distance between $\tilde{I}$ and I is at most $\varepsilon/2$. Hence, the output distribution of $\tilde{I}$ is ε-close to the hard-core distribution $\mu_{G,\lambda}$, finishing the proof. □

The extra factor in the running time of the sampling algorithm for the hard-core model as compared to the Potts model is due to the fact that the hard-core model on a bipartite graph does not in general exhibit exact symmetry between the ground states, and so we must approximate the partition functions of the even and odd dominant independent sets to sample.

3.6.3 Proper colourings

We can extend these algorithms to obtain fast sampling algorithms in most situations in which a counting problem can be put in the framework of subset polymer models. For instance, we can use Theorems 3.2.2 and 3.2.2 to improve the running times of the algorithms given by [68, 74] for sampling and counting proper q-colorings in Δ-regular bipartite graphs (for large Δ). The two papers give slightly different polymer models for proper q-colorings on Δ-regular bipartite graphs – see Section 5 of [66] and Section 5.2 of [74]. Section 5.2 of [74] shows that their polymer model is computationally feasible. Section 5.1 of [67] shows that their polymer model satisfies the Kotecký-Preiss condition; in fact, their proof establishes the polymer sampling condition (Definition 2.6.3).

It is easy to see (by comparing the polymer weights) that the polymer model of [74] therefore also satisfies the polymer sampling condition. Thus, we get the following corollary of Theorem 3.2.3 and 3.2.2.

Corollary 3.6.3. *There is an absolute constant $C > 0$ so that for all even $q \geq 3$, all $\Delta \geq Cq^2 \log^2 q$, and all $\varepsilon > e^{-n/(8q)}$, there is an ε-approximate sampling algorithm to sample a uniformly random proper q-colouring from a uniformly random Δ-regular bipartite graph running in time $O(n \log(n/\varepsilon) \log(1/\varepsilon))$. Furthermore, there is a randomised ε-approximation algorithm for the number of proper q-colourings with running time $O((n/\varepsilon)^2 \log^3(n/\varepsilon))$ and success probability at least $3/4$. For odd q, there are ε-approximate counting and sampling algorithms that both run in time $O((n/\varepsilon)^2 \log^3(n/\varepsilon) \log(1/\varepsilon))$.*

As with independent sets, the extra factor in the running time for odd q comes from the fact that the ground states (colourings in which one side of the bipartition is assigned $\lceil q/2 \rceil$ colors and the other side $\lfloor q/2 \rfloor$ colors) are exactly symmetric only if q is even.

Chapter 4

General spin systems

The work in this chapter is based on the following paper:

[40] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. *ACM Transactions on Computation Theory (TOCT)*, 13(4):1–18, 2021

An extended abstract also appeared at:

[39] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. In Javier Esparza and Daniel Kráĺ, editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik

A spin system (as introduced in Section 2.3) is a general framework that encompasses sampling and counting problems in computer science, graph homomorphism problems in combinatorics, and phase transition phenomena in statistical physics. In the spin system framework, the vertices of a graph are assigned spins from a finite set, and the interactions between neighbouring spins give rise to weights. The problem of approximating the partition function of a spin system (the aggregate weight of spin configurations) or of sampling from the resulting probability distribution over spin configurations, is typically intractable for general graphs.

In this chapter, we generalise the results of Chapter 3. We consider arbitrary spin systems on regular bipartite expander graphs, including the canonical class of random regular bipartite graphs, and develop efficient approximate sampling and counting algorithms for general spin systems. Our results apply when the degree and the spectral gap of the graph are sufficiently large, which roughly guarantees that the spin system is in the so-called low temperature regime. Our approach generalises the techniques of Jenssen et al. [68], Chen et al. [22], and the previous chapter, by showing that typical spin configurations on bipartite expanders, correspond to so-called bicliques of the spin system. Using suitable polymer models, we then show how to sample such configurations and approximate the partition function in $\tilde{O}(n^2)$ time[1], where n is the size of the graph.

4.1 Introduction

Recall from Section 2.3.3 the definitions relating to the spin system framework. Fix an integer $q \geq 2$. A q-spin system is specified by the set of spins $[q]$ and a symmetric interaction matrix $H \in \mathbb{R}_{\geq 0}^{q \times q}$. Henceforth, we will normalise H so that its largest entry is equal to 1. More formally, we will consider δ-matrices, defined as follows.

Definition 4.1.1. Let $q \geq 2$ and let $\delta \in (0,1)$. A symmetric matrix $H \in \mathbb{R}_{\geq 0}^{q \times q}$ is called a δ-matrix if $\max_{i,j \in [q]} H_{i,j} = 1$ and $\max_{i,j \in [q]: H_{i,j} \neq 1} H_{i,j} \leq \delta$. □

Note that every symmetric 0-1 matrix is a δ-matrix for every $\delta \in (0,1)$. Also, apart from the following trivial cases, we can always normalise the interaction matrix of any spin system to satisfy Definition 4.1.1 for some $\delta \in (0,1)$. If H has all of its entries equal to some $c > 0$, then $Z_{G,H} = q^{|V_G|} c^{|E_G|}$ for any graph $G = (V_G, E_G)$. Similarly, if H is the all-zeros matrix, then $Z_{G,H} = 0$ for any non-empty graph G. So, suppose that H has at least two entries with distinct values and let $h_{\max} = \max_{i,j \in [q]} H_{i,j}$. Then, the matrix $H' = \frac{1}{h_{\max}} H$ is a δ-matrix, for any $\delta \in (0,1)$ which is bigger than the second largest entry in H'. Moreover, for any graph $G = (V_G, E_G)$ we have that $Z_{G,H} = h_{\max}^{|E_G|} \cdot Z_{G,H'}$.

Approximately sampling from the Gibbs distribution of a spin system and approximating its partition function are well-studied problems in computer

[1] Ignoring logarithmic factors; that is, there exists some constant c such that the running time of the algorithm is $O(n^2 \log^c n)$.

science, since they appear in various applications. However, even for the most canonical models, such as the Potts model or graph homomorphisms, these computational problems are hard in general, even on bounded-degree graphs [31, 16, 54, 53, 38, 51, 46].

Despite these hardness results for specific models, there is currently no known characterisation classifying the complexity of approximating $Z_{G,H}$ (determining for which H approximation is tractable). However, there are some hardness results which apply to broad classes of H. For example, it has been shown [38] that if H is the adjacency matrix of any (non-trivial) bipartite graph (i.e., a bipartite graph whose connected components are not all complete) then approximating $Z_{G,H}$ is #BIS-hard[2], even when the input G is restricted to be bipartite.

In light of these hardness results, it is natural to consider whether efficient algorithms can be developed for more restricted classes of graphs. As we have seen through the work presented in Chapter 3 (and also in [66], [59], and [74]) positive algorithmic results have been obtained for specific instances of spin systems such as the hard-core and Potts models, on bounded-degree expander graphs. These results apply in a range of parameters where the problems are otherwise hard for general bounded-degree graphs.

In this chapter, we generalise all of these results by obtaining approximation algorithms of *any* spin system on regular bipartite expander graphs, whenever the degree is sufficiently large (but still independent of the size of the input graph). More precisely, we will consider regular bipartite graphs whose second eigenvalue is bounded by a small constant [60] (see Section 2.5.2 for precise details). One of the primary examples of bipartite expander graphs, and one of the main motivations behind this work, are random bipartite regular graphs. It is known [14] that, for any fixed $\varepsilon > 0$, with high probability over the choice of a random bipartite Δ-regular graph G, it holds that $\lambda(G) \leq 2\sqrt{\Delta - 1} + \varepsilon$. From a counting/sampling perspective these graphs are particularly interesting

[2]#BIS is the complexity class containing all approximate counting problems that are equivalent to approximately counting the independent sets of a bipartite graph. It is an important class in the field of approximate counting, but we will not need more details in this chapter. This hardness result does not represent a complexity classification, even in the 0-1 case. In fact, there are known examples [52, Theorem 5.1] of 0-1 matrices H for which approximating $Z_{G,H}$ is NP-hard, even when G is restricted to be bipartite.

since they have been key ingredients in obtaining inapproximability results [82, 17, 45]. Somewhat surprisingly, while we know constant factor[3] estimates of the partition function via (non-algorithmic) probabilistic methods that hold with high probability over the choice of the graph [45], it is not known how to approximately sample from the Gibbs distribution efficiently. In fact, even obtaining more refined estimates on the partition function is an open problem. As a corollary of our main result, we address both of these questions, provided that the degree Δ is sufficiently large relative to H.

Recall from Definition 2.5.4 that $\mathcal{G}^{\text{bip}}_{\Delta,\lambda}$ denotes the class of connected Δ-regular bipartite graphs of second largest eigenvalue at most λ. The results presented in this chapter will concern the following computational problem.

Name: $\textsc{Spin}_{H,\Delta,\lambda}$.
Parameters: A symmetric matrix $H \in \mathbb{R}^{q\times q}_{\geq 0}$, an integer $\Delta \geq 3$, and a real number $\lambda \in (0, \Delta)$.
Input: A graph $G \in \mathcal{G}^{\text{bip}}_{\Delta,\lambda}$.
Output: The value of $Z_{G,H}$.

In particular, we consider the problem of approximating $Z_{G,H}$ and sampling from $\mu_{G,H}$.

4.2 Main results

We give an FPRAS and an efficient approximate sampling algorithm for the above problem. Formally, we prove the following result.

Theorem 4.2.1. *Let $q \geq 2$ be an integer, δ be a real in $(0,1)$, $H \in \mathbb{R}^{q\times q}_{\geq 0}$ be a symmetric δ-matrix. Suppose that Δ, λ satisfy $\frac{\Delta}{\lambda} \geq \frac{100}{1-\delta} q^2 \log(q\Delta)$ and $\Delta \geq \left(\frac{10}{1-\delta} q \log(q\Delta)\right)^4$. Then, there is an FPRAS for $\textsc{Spin}_{H,\Delta,\lambda}$.*

In fact, there is a randomised algorithm that, given a graph $G \in \mathcal{G}^{bip}_{\Delta,\lambda}$ with $n = |V^0_G| = |V^1_G|$ vertices and an accuracy parameter $\varepsilon^* \geq e^{-n/(5q)}$, outputs an ε^*-approximation to $Z_{G,H}$ and an ε^*-sample from the Gibbs distribution $\mu_{G,H}$ in time $O((n/\varepsilon^*)^2 \log^4(n/\varepsilon^*))$.

[3]The constant factor estimates follow from applying the small subgraph conditioning method, which gives more accurate information about the partition function using the counts of small cycles, see [45, Lemma 6.12].

To prove Theorem 4.2.1, we first reduce the problem of approximating $Z_{G,H}$ to the problem of approximately sampling from the Gibbs distribution of a suitable polymer model. We then prove that this polymer model satisfies the conditions prescribed in Chapter 3; that is, we prove that it is computationally feasible and satisfies the polymer sampling condition (Definition 2.6.3) with the appropriate constant. This allows us to use (see Section 4.6 for details) the algorithm of Theorem 3.2.2 as the basis of our approximate counting and sampling algorithms for the spin system.

We remark that the restriction to bipartite expanders in Theorem 4.2.1 is necessary to have a result that holds for general spin systems (which is our goal in this paper), see Section 4.3 for a discussion of this point. As a corollary of Theorem 4.2.1, we have the following for random bipartite Δ-regular graphs.

Corollary 4.2.2. *Let $q \geq 2$ be an integer, δ be a real in $(0,1)$, and $H \in \mathbb{R}_{\geq 0}^{q\times q}$ be a symmetric δ-matrix. Then, for all integers $\Delta \geq \left(\frac{10}{1-\delta} q \log(q\Delta)\right)^4$, there is a randomised algorithm such that the following holds with high probability over the choice of a random Δ-regular bipartite graph G with $n = |V_G^0| = |V_G^1|$.*

The algorithm, on input G and an accuracy parameter $\varepsilon^* \geq e^{-n/(5q)}$, outputs in time $O((n/\varepsilon^*)^2 \log^4(n/\varepsilon^*))$ an ε^*-approximation to the partition function $Z_{G,H}$ and an ε^*-sample from the Gibbs distribution $\mu_{G,H}$.

Proof. Using the result in [14, Theorem 4], we have that, with high probability over the choice of G, it holds that $\lambda(G) \leq 2\sqrt{\Delta}$. It follows that $\frac{\Delta}{\lambda(G)} \geq \frac{1}{2}\sqrt{\Delta}$ and hence the result follows by applying Theorem 4.2.1. □

Our algorithms apply to a larger class of graphs when δ is small so that the interactions between spins are strong. By contrast, approaches such as MCMC and correlation decay apply when the interactions between spins are weak – this corresponds to the so-called 'high-temperature' regime, which is within the uniqueness phase of the infinite Δ-regular tree. Since our results concern regular graphs, they easily extend to models with external fields – the fields can be incorporated in the entries of the interaction matrix H.

4.3 Proof Outline

In order to prove our main Theorem 4.2.1, we appeal to the polymer model framework presented in Chapter 3. Our approach is inspired by, and generalises, the approaches of Chapter 3 and also [66], where counting and sampling algorithms are given for the hard-core and ferromagnetic Potts models on expander graphs at low temperatures. The main idea behind the use of polymer models is that, for graphs with good expansion properties, the partition function and the Gibbs distribution are dominated by configurations which are highly ordered, i.e., whose weight is large. As we shall see in detail in Section 4.4, each of these maximum weight configurations is close to a configuration that maps all vertices of G to a 'biclique' of H, as defined below.[4]

Definition 4.3.1. Let $q \geq 2$ be an integer, let $\delta \in (0,1)$ be a real number, and let $H \in \mathbb{R}_{\geq 0}^{q \times q}$ be a symmetric δ-matrix. A biclique of H is a pair (B_0, B_1) with $B_0, B_1 \subseteq [q]$, such that $H_{i,j} = 1$ for all $i \in B_0$ and all $j \in B_1$. We use $\mathcal{K}_H$ to denote the set of all bicliques of H and we use $\mathcal{K}_H^{\max}$ to denote the set of all inclusion maximal bicliques of H.

Given a bipartite graph G, configurations σ which assign vertices in V_G^0 a spin from B_0 and vertices in V_G^1 a spin from B_1 have weight 1. This is the largest possible weight that any configuration can have, since H is a δ-matrix. Polymer models allow us to capture deviations from such configurations and to approximate their contribution to the partition function. Using the results of Sections 4.4 and 4.5, we give the proof of Theorem 4.2.1 in Section 4.6.

In some situations where Theorem 4.2.1 provides an FPRAS for $\text{Spin}_{H,\Delta,\lambda}$ it is easy to see that traditional approaches such as the Glauber dynamics do not give good approximation algorithms. In particular, when multiple ground states make non-negligible contributions to the partition function, these ground states form a constriction in the state space which could also be used to prove that the Glauber dynamics mixes slowly. Note, however, that there are many interaction matrices with a unique inclusion-maximal biclique – in such settings it is unclear as to whether the Glauber dynamics would mix slowly.

[4] The 'biclique' terminology comes from the homomorphism problem (where H corresponds to a graph), but our interpretation here is analogous which justifies its use.

The inputs to $\textsc{Spin}_{H,\Delta,\lambda}$ need to be bipartite graphs so that we can obtain an algorithm for all possible H. When the input G is allowed to be non-bipartite, it is highly unlikely that there is an algorithm that can work for general spin systems. To see this, consider the case where H is the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Recall from the remark following Definition 4.1.1 that this is a δ-matrix for any $\delta \in (0,1)$. Spin configurations with this matrix H can be viewed as independent sets of G (where spin 0 means 'in the independent set'). However, it is unclear how to compute a maximum independent set on random Δ-regular graphs for large Δ, and the approximate counting problem that we study here is likely to be even harder.[5]

Despite the lack of a general result for non-bipartite graphs, for certain spin systems, our methods could be extended to allow non-bipartite inputs. For example, this occurs when all maximal bicliques (B_0, B_1) of H have $B_0 = B_1$ (so they can be viewed as 'cliques' of H). This is the case, for example, in the ferromagnetic Potts model, and it is the case more generally in 'ferromagnetic' spin systems (where the interaction matrix has only positive eigenvalues, see, e.g., [46]).

4.4 Ground states for spin configurations

In this section, we show that the partition function of a spin system is dominated by configurations which are 'close to maximal bicliques', cf. Definition 4.3.1. Let $q \geq 2$, $\Delta \geq 3$ be integers and λ, δ be reals with $\lambda \in (0, \Delta)$ and $\delta \in (0,1)$. Let $G \in \mathcal{G}^{\text{bip}}_{\Delta,\lambda}$ and let $H \in \mathbb{R}^{q\times q}_{\geq 0}$ be a symmetric δ-matrix.

We next describe more precisely the configurations which are 'close' (in some rigorous sense) to a maximal biclique of H. Given $\sigma\colon V_G \to [q]$ and a spin $i \in [q]$, we write $\sigma^{-1}(i)$ for the set of vertices of G whose image under σ is i. More generally, for a subset of spins $Q \subseteq [q]$, we let $\sigma^{-1}(Q) = \{v \in V_G \mid \sigma(v) \in Q\}$.

Definition 4.4.1. Let $\varepsilon \in (0,1)$ be a real number. For $(B_0, B_1) \in \mathcal{K}^{\max}_H$, define $\Sigma^{B_0,B_1}_{G,H,\varepsilon}$ to be the set of spin configurations $\sigma \in \Sigma_{G,H}$ for which

$$\left|\sigma^{-1}(B_0) \cap V^0_G\right| + \left|\sigma^{-1}(B_1) \cap V^1_G\right| \geq (1-\varepsilon)|V_G|.$$

[5]This is in contrast to the bipartite case, where finding a maximum weight configuration is trivial since this is just a configuration assigning spins from a biclique.

We define $\Sigma_{G,H,\varepsilon}$ to be the union of the sets $\Sigma_{G,H,\varepsilon}^{B_0,B_1}$ over all bicliques $(B_0, B_1) \in \mathcal{K}_H^{\max}$ and define $Z_{G,H,\varepsilon} := \sum_{\sigma \in \Sigma_{G,H,\varepsilon}} w_{G,H}(\sigma)$.

The following result shows that $Z_{G,H,\varepsilon}$ gives a close approximation to $Z_{G,H}$ whenever ε is sufficiently large relative to λ, Δ, q.[6]

Lemma 4.4.2. *Let $\varepsilon \in (0,1)$ be a real number for which $\varepsilon \geq 2q\lambda/\Delta$ and $\varepsilon^2 \geq \frac{8q^2 \log q}{\Delta \log(1/\delta)}$. Then, for $G \in \mathcal{G}_{\Delta,\lambda}^{bip}$ with $n = |V_G^0| = |V_G^1|$, we have that $Z_{G,H,\varepsilon}$ is an e^{-n}-approximation to $Z_{G,H}$.*

Proof. We associate each spin configuration $\sigma \in \Sigma_{G,H}$ with a pair of spin subsets $\big(B_0(\sigma), B_1(\sigma)\big)$ by setting for $i \in \{0,1\}$

$$B_i(\sigma) = \Big\{ j \in [q] : \big|\sigma^{-1}(j) \cap V_G^i\big| \geq \frac{\varepsilon n}{q} \Big\}.$$

Note that for $\sigma \in \Sigma_{G,H}$, there are fewer than εn vertices of V_G^i which are *not* assigned spins from $B_i(\sigma)$. Also note that, since $\varepsilon \in (0,1)$ and $|V_G^i| = n$, we have that $B_i(\sigma) \neq \emptyset$ for $i \in \{0,1\}$.

Fix arbitrary $\sigma \in \Sigma_{G,H}$. We first show that

$$\text{either } \big(B_0(\sigma), B_1(\sigma)\big) \in \mathcal{K}_H \text{ or } w_{G,H}(\sigma) \leq \delta^{\Delta\varepsilon^2 n/(2q^2)}, \tag{4.1}$$

i.e., either $(B_0(\sigma), B_1(\sigma))$ is a biclique of H or σ has small weight. For $i \in \{0,1\}$, consider arbitrary $j_i \in B_i(\sigma)$ and let $S_i = \sigma^{-1}(j_i) \cap V_G^i$. Since $|S_i| \geq \varepsilon n/q$, we have that $\frac{\Delta}{2n}\sqrt{|S_0||S_1|} \geq \Delta\varepsilon/(2q) \geq \lambda$, thus it follows from Corollary 2.5.6 that

$$e(S_0, S_1) \geq \frac{\Delta |S_0||S_1|}{2n} \geq \frac{\Delta \varepsilon^2 n}{2q^2}.$$

It follows that, if $H_{j_0,j_1} \leq \delta$, then $w_{G,H}(\sigma) \leq \delta^{\Delta\varepsilon^2 n/(2q^2)}$; otherwise, $H_{j_0,j_1} = 1$. Since j_0, j_1 were arbitrary spins in $B_0(\sigma), B_1(\sigma)$, respectively, we conclude (4.1).

Let σ be such that $\big(B_0(\sigma), B_1(\sigma)\big) \in \mathcal{K}_H$. Then there exists $(B_0, B_1) \in \mathcal{K}_H^{\max}$ such that $B_i(\sigma) \subseteq B_i$ for $i \in \{0,1\}$. Moreover, for $i \in \{0,1\}$ and $j \in [q] \setminus B_i$, we have that $\big|\sigma^{-1}(j) \cap V_G^i\big| < \varepsilon n/q$ and therefore that $|\sigma^{-1}([q] \setminus B_i) \cap V_G^i| < \varepsilon n$. Hence, we conclude that

$$\big|\sigma^{-1}(B_0) \cap V_G^0\big| + \big|\sigma^{-1}(B_1) \cap V_G^1\big| \geq (1-\varepsilon)|V_G|.$$

[6]Note, in Lemma 4.4.2, as in other lemmas as well, our assumed inequalities for ε impose some restrictions on Δ, λ, q to ensure that such an ε exists. These restrictions will be carefully accounted for when we apply these lemmas; namely, in the proof of Theorem 4.2.1.

Combining this with (4.1), we obtain that for all $\sigma \in \Sigma_{G,H} \setminus \Sigma_{G,H,\varepsilon}$ it holds that $w_{G,H}(\sigma) \leq \delta^{\Delta\varepsilon^2 n/(2q^2)}$, and hence

$$Z_{G,H} - Z_{G,H,\varepsilon} \leq \sum_{\sigma \in \Sigma_{G,H} \setminus \Sigma_{G,H,\varepsilon}} \delta^{\frac{\Delta\varepsilon^2 n}{2q^2}} \leq q^{2n} \delta^{\frac{\Delta\varepsilon^2 n}{2q^2}} \leq q^{-2n},$$

where in the last inequality we used that $\Delta \geq \frac{8q^2 \log(q)}{\varepsilon^2 \log(1/\delta)}$. The result follows since $Z_{G,H} \geq 1$; this bound can be seen by considering the configuration that maps V_G^0 to j_0 and V_G^1 to j_1, where $j_0, j_1 \in [q]$ are such that $H_{j_0,j_1} = 1$. □

For our approximation algorithms, it will be useful to consider the following quantities.

Definition 4.4.3. For a real number $\varepsilon \in (0,1)$, let

$$\widehat{Z}_{G,H,\varepsilon} = \sum_{(B_0,B_1) \in \mathcal{K}_H^{\max}} \sum_{\sigma \in \Sigma_{G,H,\varepsilon}^{B_0,B_1}} w_{G,H}(\sigma) \quad \text{and} \quad Z_{G,H,\varepsilon}^{\mathsf{overlap}} := \sum_{\sigma \in \Sigma_{G,H,\varepsilon}^{\mathsf{overlap}}} w_{G,H}(\sigma),$$

where $\Sigma_{G,H,\varepsilon}^{\mathsf{overlap}} := \bigcup_{(B_0,B_1) \in \mathcal{K}_H^{\max}} \bigcup_{(C_0,C_1) \in \mathcal{K}_H^{\max} \setminus \{(B_0,B_1)\}} \left(\Sigma_{G,H,\varepsilon}^{B_0,B_1} \cap \Sigma_{G,H,\varepsilon}^{C_0,C_1}\right)$.

The following lemma shows, given a value of ε that is sufficiently small, that $\widehat{Z}_{G,H,3\varepsilon}$ is a close approximation to $Z_{G,H,\varepsilon}$ by showing that $Z_{G,H,3\varepsilon}^{\mathsf{overlap}}$ is small relative to $Z_{G,H,\varepsilon}$.

Lemma 4.4.4. *Let $\varepsilon \in (0, \frac{1}{240q\log q}]$ be a real number for which $\varepsilon^2 \geq \frac{8q^2\log(q)}{\Delta\log(1/\delta)}$ and $\varepsilon \geq 2q\frac{\lambda}{\Delta}$. Then, for all $G \in \mathcal{G}_{\Delta,\lambda}^{bip}$ with $n = |V_G^0| = |V_G^1|$ sufficiently large, we have that $Z_{G,H,3\varepsilon}^{\mathsf{overlap}} \leq e^{-n/(3q)} Z_{G,H,\varepsilon}$ and that $\widehat{Z}_{G,H,3\varepsilon}$ is an $e^{-n/(4q)}$-approximation to $Z_{G,H,\varepsilon}$.*

Proof. We will show the result for all integers n satisfying $n^2 2^{6q} e^{3n/(5q)} \leq e^{2n/(3q)}\left(1 - \frac{2q}{n}\right)^2$.

Since $\Sigma_{G,H,3\varepsilon} = \bigcup_{(B_0,B_1) \in \mathcal{K}_H^{\max}} \Sigma_{G,H,3\varepsilon}^{B_0,B_1}$ and $\Sigma_{G,H,\varepsilon} \subseteq \Sigma_{G,H,3\varepsilon}$, we have

$$\widehat{Z}_{G,H,3\varepsilon} \geq Z_{G,H,3\varepsilon} \geq Z_{G,H,\varepsilon}.$$

Moreover, each $\sigma \in \Sigma_{G,H,3\varepsilon}^{\mathsf{overlap}}$ is accounted for at most $|\mathcal{K}_H^{\max}| \leq 2^{2q}$ times in $\widehat{Z}_{G,H,3\varepsilon}$, therefore

$$\widehat{Z}_{G,H,3\varepsilon} \leq 2^{2q} \widehat{Z}_{G,H,3\varepsilon}^{\mathsf{overlap}} + \sum_{\sigma \in \Sigma_{G,H,3\varepsilon}} w_{G,H}(\sigma).$$

Since $\Sigma_{G,H,\varepsilon} \subseteq \Sigma_{G,H,3\varepsilon}$, we therefore have that

$$\begin{aligned}\widehat{Z}_{G,H,3\varepsilon} - Z_{G,H,\varepsilon} &\leq 2^{2q}\widehat{Z}^{\mathsf{overlap}}_{G,H,3\varepsilon} + \sum_{\sigma\in\Sigma_{G,H,3\varepsilon}\setminus\Sigma_{G,H,\varepsilon}} w_{G,H}(\sigma) \\ &\leq 2^{2q}\widehat{Z}^{\mathsf{overlap}}_{G,H,3\varepsilon} + Z_{G,H} - Z_{G,H,\varepsilon} \\ &\leq 2^{2q}\widehat{Z}^{\mathsf{overlap}}_{G,H,3\varepsilon} + e^{-n}Z_{G,H,\varepsilon}, \end{aligned} \tag{4.2}$$

where the last inequality follows from Lemma 4.4.2. The lemma will thus follow by showing

$$2^{2q}\widehat{Z}^{\mathsf{overlap}}_{G,H,3\varepsilon} \leq e^{-n/(3q)}Z_{G,H,\varepsilon}, \tag{4.3}$$

since then, from (4.2) and (4.3), we obtain that

$$\widehat{Z}_{G,H,3\varepsilon} - Z_{G,H,\varepsilon} \leq \left(e^{-n} + e^{-n/(3q)}\right)\cdot Z_{G,H,\varepsilon} \leq e^{-n/(4q)}\cdot Z_{G,H,\varepsilon}.$$

It remains to prove (4.3). Consider $(B_0,B_1),(C_0,C_1) \in \mathcal{K}_H^{\max}$ such that $(B_0,B_1)\neq(C_0,C_1)$ and let us upper bound the aggregate weight of configurations in the set $F := \Sigma^{B_0,B_1}_{G,H,3\varepsilon}\cap\Sigma^{C_0,C_1}_{G,H,3\varepsilon}$; in fact, we will just upper bound $|F|$ and use the trivial upper bound of 1 on the weight of a configuration. For $\sigma\in F$, let $S = \cup_{i\in\{0,1\}}(V_G^i\backslash\sigma^{-1}(B_i))$ and $T = \cup_{i\in\{0,1\}}(V_G^i\backslash\sigma^{-1}(C_i))$, so that $|S|,|T| \leq 3\varepsilon|V_G| = 6\varepsilon n$. For $\varepsilon\in(0,1/10)$, there are at most $\sum_{k=0}^{\lfloor 6\varepsilon n\rfloor}\binom{2n}{k} \leq n\binom{2n}{\lfloor 6\varepsilon n\rfloor}$ ways to choose each of S and T, and then, crudely, $q^{12\varepsilon n}$ ways to assign them spins; further, for $i\in\{0,1\}$, the vertices in $V_G^i\backslash(S\cup T)$ can be coloured in at most $|B_i\cap C_i|^n$ ways, since they must have a colour in $B_i\cap C_i$. Observe now that at least one of the inequalities $|B_0\cap C_0|\leq|B_0|-1, |B_1\cap C_1|\leq|B_1|-1$ must hold since otherwise $B_0\subseteq C_0, B_1\subseteq C_1$, contradicting the maximality $(B_0,B_1)\in\mathcal{K}_H^{\max}$. We also have the bounds

$$\left(\frac{|B_i|-1}{|B_i|}\right)^n \leq (1-1/q)^n \leq e^{-n/q} \text{ for } i\in\{0,1\}.$$

Combining the above, we obtain

$$\sum_{\sigma\in\Sigma^{B_0,B_1}_{G,H,3\varepsilon}\cap\Sigma^{C_0,C_1}_{G,H,3\varepsilon}} w_{G,H}(\sigma) \leq n^2\binom{2n}{\lfloor 6\varepsilon n\rfloor}^2 q^{12\varepsilon n}e^{-n/q}|B_0|^n|B_1|^n.$$

We have $|\mathcal{K}_H^{\max}|\leq 2^{2q}$ and $\binom{2n}{\lfloor 6\varepsilon n\rfloor} \leq (e/3\varepsilon)^{6\varepsilon n}$, hence

$$\widehat{Z}^{\mathsf{overlap}}_{G,H,3\varepsilon} \leq n^2 2^{4q}(qe/3\varepsilon)^{12\varepsilon n}e^{-n/q}\sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}}|B_0|^n|B_1|^n. \tag{4.4}$$

For $c > 0$, the function $f(x) = (c/x)^x$ is increasing in the interval $(0, c/e]$, so using that $\varepsilon \leq \frac{1}{240q \log q}$, we further have that

$$(qe/3\varepsilon)^{\varepsilon} \leq (80qe^2 \log q)^{1/(240q \log q)} = e^{\frac{1+\log(80)+2\log q+\log\log q}{240q \log q}} \leq e^{1/(20q)}, \quad (4.5)$$

where in the last inequality we used that $\frac{1+\log(80)}{240 \log 2} + \frac{1}{120} + \frac{1}{240} \leq \frac{1}{20}$. By considering the set of surjective maps from V_G^0 to B_0 and V_G^1 to B_1, for each maximal biclique $(B_0, B_1) \in \mathcal{K}_H^{\max}$, we can also lower bound $Z_{G,H,\varepsilon}$. Lemma 17 of [38] (which is a simple corollary of an analogous bound in [28]) states that the number of surjective maps from a set of size m to a set of size k is at least $(1 - 2k/m)k^m$. Thus, we have that

$$Z_{G,H,\varepsilon} \geq \left(1 - \frac{2q}{n}\right)^2 \sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}} |B_0|^n |B_1|^n.$$

This, combined with (4.4), (4.5) and the choice of n, yields (4.3), finishing the proof. □

4.5 Applying the polymer model framework

In this section, we apply the polymer model framework and the algorithms of Chapter 3, to obtain an efficient algorithm for sampling from the Gibbs distribution, and approximating the partition function of a spin system. We begin by defining a polymer model for spin systems that captures the deviations that spin configurations take from maximal bicliques. The polymer model that we propose is a generalisation to arbitrary spin systems of a polymer model that was used in [66, Section 5] in the case of proper colourings.

4.5.1 A polymer model for spin systems

Let $H \in \mathbb{R}_{\geq 0}^{q\times q}$ be a symmetric matrix and $(B_0, B_1) \in \mathcal{K}_H^{\max}$ be a maximal biclique of H. Let $G \in \mathcal{G}_{\Delta,\lambda}^{\text{bip}}$ be a graph, and let $\varepsilon \in (0, 1)$. The host graph for the polymer model is $J_G = G^3$, where G^3 is the graph defined on V_G with two vertices connected by an edge if the distance between them in G is at most 3. Note that in the remainder of the chapter, we will always use ∂ to denote the boundary with respect to G of a vertex set, therefore we omit G from this

notation. For $v \in V_G^i$ with $i \in \{0,1\}$, the set of ground state spins g_v is B_i. Let $\mathcal{P}_{G,H}^{B_0,B_1}$ denote the set of all polymers, i.e., all pairs $\gamma = (V_\gamma, \sigma_\gamma)$ consisting of a G^3-connected set of vertices V_γ and an assignment $\sigma_\gamma \colon V_\gamma \to [q]$ such that $\sigma_\gamma(v) \in [q] \setminus g_v$ for all $v \in V_\gamma$. We define the set of allowed polymers as

$$\mathcal{C}_{G,H,\varepsilon}^{B_0,B_1} = \left\{ \gamma \in \mathcal{P}_{G,H}^{B_0,B_1} : |V_\gamma| \leq \varepsilon |V_G| \right\} \tag{4.6}$$

and let $\Omega_{G,H,\varepsilon}^{B_0,B_1}$ denote the set of all sets of mutually compatible polymers. We define the weight of a polymer $\gamma = (V_\gamma, \sigma_\gamma) \in \mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}$ as

$$w_{G,H}^{B_0,B_1}(\gamma) = \frac{\prod_{\{u,v\} \in E_G(V_\gamma)} H_{\sigma_\gamma(u),\sigma_\gamma(v)} \prod_{u \in \partial V_\gamma} F_u}{\prod_{i \in \{0,1\}} |B_i|^{|V_G^i \cap V_\gamma^+|}}, \tag{4.7}$$

where

$$\text{for } u \in V_G^i \text{ with } i \in \{0,1\},\ F_u := \sum_{j \in B_i} \prod_{v \in V_\gamma \cap \partial u} H_{j,\sigma_\gamma(v)}.$$

We let $Z_{G,H,\varepsilon}^{B_0,B_1}$ and $\mu_{G,H,\varepsilon}^{B_0,B_1}$ denote the partition function and the Gibbs distribution of the polymer model $(\mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}, w_{G,H}^{B_0,B_1}, J_G)$, as defined in Section 2.6.

The next lemma shows the motivation behind the definition of the weight of a polymer. For a polymer configuration $\Gamma \in \Omega_{G,H,\varepsilon}^{B_0,B_1}$, let $\cup\Gamma = \bigcup_{\gamma \in \Gamma} V_\gamma$, and let σ_Γ denote the assignment to vertices in $\cup\Gamma$ obtained by combining all of the assignments σ_γ, for $\gamma \in \Gamma$. Finally, recall that $(\cup\Gamma)^+ = (\cup\Gamma) \cup \partial(\cup\Gamma)$.

Definition 4.5.1. For $\Gamma \in \Omega_{G,H,\varepsilon}^{B_0,B_1}$, define $\Sigma_{G,H}^{B_0,B_1}(\Gamma)$ to be the set of configurations τ such that $\tau|_{\cup\Gamma} = \sigma_\Gamma$ and which map, for $i \in \{0,1\}$, $V_G^i \setminus (\cup\Gamma)$ to B_i.

Lemma 4.5.2. *Let $n = |V_G^0| = |V_G^1|$. For all $\varepsilon \in (0,1)$, and all polymer configurations $\Gamma \in \Omega_{G,H,\varepsilon}^{B_0,B_1}$, we have that*

$$|B_0|^n |B_1|^n \prod_{\gamma \in \Gamma} w_{G,H}^{B_0,B_1}(\gamma) = \sum_{\tau \in \Sigma_{G,H}^{B_0,B_1}(\Gamma)} w_{G,H}(\tau).$$

Proof. Let γ, γ' be two distinct polymers in Γ. Since the polymers are compatible, they correspond to distinct G^3-connected components; in particular,

$V_\gamma^+ \cap V_{\gamma'}^+ = \emptyset$. It follows that $|(\cup\Gamma)^+| = \sum_{\gamma\in\Gamma} |V_\gamma^+|$. Hence, by the definition in (4.7), we have that

$$\begin{aligned}
&|B_0|^n|B_1|^n \prod_{\gamma\in\Gamma} w_{G,H}^{B_0,B_1}(\gamma) \\
&\qquad = |B_0|^n|B_1|^n \prod_{\gamma\in\Gamma} \frac{\prod_{\{u,v\}\in E_G(V_\gamma)} H_{\sigma_\gamma(u),\sigma_\gamma(v)} \prod_{u\in\partial V_\gamma} F_u}{\prod_{i\in\{0,1\}} |B_i|^{|V_G^i\cap V_\gamma^+|}} \\
&\qquad = \prod_{i\in\{0,1\}} |B_i|^{|V_G^i\setminus(\cup\Gamma)^+|} \prod_{\{u,v\}\in E_G(\cup\Gamma)} H_{\sigma_\Gamma(u),\sigma_\Gamma(v)} \prod_{u\in\partial(\cup\Gamma)} F_u. \qquad (4.8)
\end{aligned}$$

On the other hand, for each $\tau \in \Sigma_{G,H}^{B_0,B_1}(\Gamma)$, we have that

$$w_{G,H}(\tau) = \prod_{\{u,v\}\in E_G(\cup\Gamma)} H_{\sigma_\Gamma(u),\sigma_\Gamma(v)} \prod_{u\in\partial(\cup\Gamma)} \prod_{v\in(\cup\Gamma)\cap\partial u} H_{\sigma_\Gamma(v),\tau(u)},$$

i.e., given Γ and that $\tau \in \Sigma_{G,H}^{B_0,B_1}(\Gamma)$, the weight of τ depends only on the assignment of $\partial(\cup\Gamma)$ — this is because $H_{\tau(u),\tau(v)} = 1$ for all edges $\{u,v\}$ with $u,v \notin \cup\Gamma$. Let $W(\Gamma)$ be the set of configurations $\eta\colon \partial(\cup\Gamma) \to B_0 \cup B_1$ such that vertices in $\partial(\cup\Gamma)\cap V_G^i$ take a spin in B_i, for $i \in \{0,1\}$. For $\eta \in W(\Gamma)$, the number of configurations τ in $\Sigma_{G,H}^{B_0,B_1}(\Gamma)$ with $\tau|_{\partial(\cup\Gamma)} = \eta$ is $\prod_{i\in\{0,1\}} |B_i|^{|V_G^i\setminus(\cup\Gamma)^+|}$, and

$$\begin{aligned}
\sum_{\eta\in W(\Gamma)} \prod_{u\in\partial(\cup\Gamma)} \prod_{v\in(\cup\Gamma)\cap\partial u} H_{\sigma_\Gamma(v),\eta(u)} &= \prod_{i\in\{0,1\}} \prod_{u\in V_G^i\cap\partial(\cup\Gamma)} \sum_{\eta\colon \{u\}\to B_i} \prod_{v\in(\cup\Gamma)\cap\partial u} H_{\sigma_\Gamma(v),\eta(u)} \\
&= \prod_{u\in\partial(\cup\Gamma)} F_u. \qquad (4.9)
\end{aligned}$$

It follows that

$$\sum_{\tau\in\Sigma_{G,H}^{B_0,B_1}(\Gamma)} w_{G,H}(\tau) = \prod_{i\in\{0,1\}} |B_i|^{|V_G^i\setminus(\cup\Gamma)^+|} \prod_{\{u,v\}\in E_G(V_\Gamma)} H_{\sigma_\Gamma(u),\sigma_\Gamma(v)} \prod_{u\in\partial(\cup\Gamma)} F_u,$$

which combined with (4.8) gives the desired equality. □

The following quantity combines the partition functions of the polymer models corresponding to maximal bicliques of H; we will use this as our approximation to $Z_{G,H}$.

Definition 4.5.3. For $\varepsilon \in (0,1)$, let

$$Z_{G,H,\varepsilon}^{\text{polymer}} = \sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}} |B_0|^n|B_1|^n \cdot Z_{G,H,\varepsilon}^{B_0,B_1}.$$

The following lemma is a minor adaptation of [66, Claim 29] in our setting, and will be used to bound the aggregate size of polymer configurations.

Lemma 4.5.4 ([66, Claim 29])**.** *Let $\varepsilon \in (0,1)$ be such that $\varepsilon \geq \lambda/\Delta$. Then, for $G \in \mathcal{G}^{bip}_{\Delta,\lambda}$ with $n = |V_G^0| = |V_G^1|$ sufficiently large, there is no set $S \subseteq V_G$ with $|S| > 6\varepsilon n$ whose G^3-connected components, say $S_1, S_2, \ldots, S_k$, satisfy $|S_i| \leq 2\varepsilon n$ for $i \in [k]$.*

Proof. We will show the result when $n \geq 1/(2\varepsilon^2\Delta)$. For the sake of contradiction, suppose that such a set S exists with $|S| > 6\varepsilon n$ and whose G^3-connected components $S_1, S_2, \ldots, S_k$ satisfy $|S_i| \leq 2\varepsilon n$ for $i \in [k]$. Then, we can partition $\{1, 2, \ldots, k\}$ into sets T_1 and T_2 such that, for $j \in \{1,2\}$, the set $U_j = \cup_{i \in T_j} S_i$ satisfies $|U_j| \geq 2\varepsilon n$. It follows by Corollary 2.5.6 that $e_G(U_1, U_2) \geq 2\varepsilon^2\Delta n \geq 1$, therefore there must exist $i \in T_1$ and $j \in T_2$ such that $e_G(S_i, S_j) \geq 1$. This yields a contradiction since two distinct G^3-connected components cannot be connected by an edge. □

Finally, we have the following result.

Lemma 4.5.5. *Let $\varepsilon \in (0, \frac{1}{240q\log q}]$ be such that $\Delta \geq \frac{8q^2\log(q)}{\varepsilon^2\log(1/\delta)}$ and $\varepsilon \geq 2q\frac{\lambda}{\Delta}$. Then, for all $G \in \mathcal{G}^{bip}_{\Delta,\lambda}$ with $n = |V_G^0| = |V_G^1|$ sufficiently large, we have that $Z^{polymer}_{G,H,\varepsilon}$ is an $e^{-n/(4q)}$-approximation to $Z_{G,H,\varepsilon}$.*

Proof. Using the definition of the polymer partition function $Z^{B_0,B_1}_{G,H,\varepsilon}$ (see Section 4.5.1) and Lemma 4.5.2, we can rewrite $Z^{\text{polymer}}_{G,H,\varepsilon}$ from Definition 4.5.3 as

$$\begin{aligned} Z^{\text{polymer}}_{G,H,\varepsilon} &= \sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}} \sum_{\Gamma\in\Omega^{B_0,B_1}_{G,H,\varepsilon}} |B_0|^n|B_1|^n \prod_{\gamma\in\Gamma} w^{B_0,B_1}_{G,H}(\gamma) \\ &= \sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}} \sum_{\Gamma\in\Omega^{B_0,B_1}_{G,H,\varepsilon}} \sum_{\sigma\in\Sigma^{B_0,B_1}_{G,H}(\Gamma)} w_{G,H}(\sigma). \end{aligned} \tag{4.10}$$

Recall from Definition 4.5.1 that $\Sigma^{B_0,B_1}_{G,H}(\Gamma)$ is the set of configurations τ such that $\tau|_{\cup\Gamma} = \sigma_\Gamma$ and which map, for $i \in \{0,1\}$, $V_G^i \backslash (\cup\Gamma)$ to B_i. Recall also from Definition 4.4.1 that $\Sigma^{B_0,B_1}_{G,H,\varepsilon}$ is the set of $\sigma \in \Sigma_{G,H}$ for which $\left|\sigma^{-1}(B_0) \cap V_G^0\right| + \left|\sigma^{-1}(B_1) \cap V_G^1\right| \geq (1-\varepsilon)|V_G|$ and that $Z_{G,H,\varepsilon} := \sum_{\sigma\in\Sigma_{G,H,\varepsilon}} w_{G,H}(\sigma)$.

We first prove that $Z^{\text{polymer}}_{G,H,\varepsilon} \geq Z_{G,H,\varepsilon}$. Index the bicliques in $\mathcal{K}_H^{\max}$ arbitrarily. For $\tau \in \Sigma_{G,H,\varepsilon}$, let $\left(B_{0,\tau}, B_{1,\tau}\right)$ be the biclique $(B_0, B_1) \in \mathcal{K}_H^{\max}$ with the smallest

index such that $\tau \in \Sigma_{G,H,\varepsilon}^{B_0,B_1}$. Let $T_i = V_G^i \cap \tau^{-1}([q] \backslash B_{i,\tau})$ for $i \in \{0,1\}$, so that $|T_0 \cup T_1| \leq \varepsilon |V_G|$. Let $S_1, \ldots, S_k$ denote the G^3-connected components of $T_0 \cup T_1$. Consider the polymer configuration Γ_τ which is the union of the polymers $(S_1, \tau|_{S_1}), \ldots, (S_k, \tau|_{S_k})$. Note that the map $\tau \mapsto (B_{0,\tau}, B_{1,\tau}, \Gamma_\tau)$ is injective. Moreover, we have that $\Gamma_\tau \in \Omega_{G,H,\varepsilon}^{B_{0,\tau},B_{1,\tau}}$ and $\tau \in \Sigma_{G,H}^{B_{0,\tau},B_{1,\tau}}(\Gamma_\tau)$, so from (4.10) we obtain that $Z_{G,H,\varepsilon}^{\text{polymer}} \geq Z_{G,H,\varepsilon}$.

We next prove that $Z_{G,H,\varepsilon}^{\text{polymer}} \leq (1 + e^{-n/(4q)}) Z_{G,H,\varepsilon}$. Recall from Definition 4.4.3 that

$$\widehat{Z}_{G,H,3\varepsilon} = \sum_{(B_0,B_1) \in \mathcal{K}_H^{\max}} \sum_{\sigma \in \Sigma_{G,H,3\varepsilon}^{B_0,B_1}} w_{G,H}(\sigma).$$

Fix any $(B_0, B_1) \in \mathcal{K}_H^{\max}$. For $\Gamma \in \Omega_{G,H,\varepsilon}^{B_0,B_1}$, we have from Lemma 4.5.4 that $|\cup \Gamma| \leq 6\varepsilon n = 3\varepsilon |V_G|$. Therefore, for $\sigma \in \Sigma_{G,H}^{B_0,B_1}(\Gamma)$ as in Definition 4.5.1, we have that $\left|\sigma^{-1}(B_0) \cap V_G^0\right| + \left|\sigma^{-1}(B_1) \cap V_G^1\right| \geq (1 - 3\varepsilon)|V_G|$ and hence $\sigma \in \Sigma_{G,H,3\varepsilon}^{B_0,B_1}$ (cf. Definition 4.4.1). Note that the map $(\Gamma, \sigma) \mapsto \sigma$ is injective (since Γ can be recovered from σ using the biclique (B_0, B_1)), yielding that

$$\sum_{\Gamma \in \Omega_{G,H,\varepsilon}^{B_0,B_1}} \sum_{\sigma \in \Sigma_{G,H}^{B_0,B_1}(\Gamma)} w_{G,H}(\sigma) \leq \sum_{\sigma \in \Sigma_{G,H,3\varepsilon}^{B_0,B_1}} w_{G,H}(\sigma),$$

and hence $Z_{G,H,\varepsilon}^{\text{polymer}} \leq \widehat{Z}_{G,H,3\varepsilon}$. By Lemma 4.4.4, we obtain that $Z_{G,H,\varepsilon}^{\text{polymer}} \leq (1 + e^{-n/(4q)}) Z_{G,H,\varepsilon}$.

This finishes the proof of Lemma 4.5.5. □

4.5.2 Sampling from the polymer model

Let $q \geq 2$ and $\Delta \geq 3$ be integers and let $\delta \in (0,1)$ and $\lambda \in (0, \Delta)$ be real numbers. Let $H \in \mathbb{R}_{\geq 0}^{q \times q}$ be a symmetric δ-matrix and let $(B_0, B_1) \in \mathcal{K}_H^{\max}$. For $\varepsilon \in (0,1)$, we now show that the family of polymer models $\{(\mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}, w_{G,H}^{B_0,B_1}, J_G) \mid G \in \mathcal{G}_{\Delta,\lambda}^{\text{bip}}\}$, which was defined in the previous subsection, is computationally feasible and satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$.

Lemma 4.5.6. *Let $\varepsilon \in (0,1)$ be such that $\varepsilon \geq \lambda^2/\Delta^2$ and $\varepsilon \leq \frac{1-\delta}{40q\log(q\Delta)}$. The family of polymer models $\{(\mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}, w_{G,H}^{B_0,B_1}, J_G) \mid G \in \mathcal{G}_{\Delta,\lambda}^{bip}\}$ is computationally*

feasible and satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta^3)$.

Proof. Consider an arbitrary polymer $\gamma = (V_\gamma, \sigma_\gamma) \in \mathcal{P}_{G,H}^{B_0,B_1}$. We first verify computational feasibility by showing that we can determine whether $\gamma \in \mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}$ and, if so, compute its weight $w_{G,H}^{B_0,B_1}(\gamma)$ in time $O(|V_\gamma|)$. Indeed, we have that $\gamma \in \mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}$ iff $|V_\gamma| \leq 2\varepsilon n$ and σ_γ maps $V_\gamma \cap V_G^i$ to $[q] \setminus B_i$ for each $i \in \{0,1\}$, which can be clearly checked in time $O(|V_\gamma|)$. To compute $w_{G,H}^{B_0,B_1}(\gamma)$, recall from (4.7) that

$$w_{G,H}^{B_0,B_1}(\gamma) = \frac{\prod_{\{u,v\}\in E_G(V_\gamma)} H_{\sigma_\gamma(u),\sigma_\gamma(v)} \prod_{u\in\partial V_\gamma} F_u}{\prod_{i\in\{0,1\}} |B_i|^{|V_G^i \cap V_\gamma^+|}},$$

where for a vertex $u \in V_G^i \cap \partial V_\gamma$ with $i \in \{0,1\}$, $F_u := \sum_{j\in B_i} \prod_{v\in\partial u\cap V_\gamma} H_{j,\sigma_\gamma(v)}$. It follows that the right hand side of (4.7) can be computed in $O(|V_\gamma|)$ time, using that $|\partial V_\gamma| \leq \Delta|V_\gamma|$.

We next verify the polymer sampling condition. From (4.7), using that the entries of H are at most 1, we have the bound

$$w_{G,H}^{B_0,B_1}(\gamma) \leq \frac{\prod_{u\in\partial V_\gamma\cap V_G} F_u}{\prod_{i\in\{0,1\}} |B_i|^{|V_G^i \cap V_\gamma^+|}}. \tag{4.11}$$

Let us now consider the factor F_u for $u \in \partial V_\gamma \cap V_G^i$ with $i \in \{0,1\}$. Let v be a neighbour of u in $V_\gamma \cap V_G^{i\oplus 1}$; such v exists since $u \in \partial V_\gamma \cap V_G^i$. Then, there exists $j \in B_i$ such that $H_{j,\sigma_\gamma(v)} \leq \delta$; otherwise, $(B_i, B_{i\oplus 1} \cup \{\sigma_\gamma(v)\})$ would also be a biclique of H, contradicting the maximality of (B_0, B_1) (since $\sigma_\gamma(v) \notin B_{i\oplus 1}$). It follows that $F_u \leq |B_i| - 1 + \delta$. Using this in (4.11), we get that

$$\begin{aligned} w_{G,H}^{B_0,B_1}(\gamma) &\leq \frac{\prod_{i\in\{0,1\}} (|B_i| - 1 + \delta)^{|\partial V_\gamma \cap V_G^i|}}{\prod_{i\in\{0,1\}} |B_i|^{|V_G^i \cap V_\gamma^+|}} \\ &\leq \Big(1 - \frac{1-\delta}{q}\Big)^{|\partial V_\gamma|} \\ &\leq e^{-|\partial V_\gamma|\left(\frac{1-\delta}{q}\right)}, \end{aligned} \tag{4.12}$$

where in the second to last inequality we used that $|V_\gamma^+ \cap V_G^i| \geq |\partial V_\gamma \cap V_G^i|$ for $i \in \{0,1\}$.

We next lower bound $|\partial V_\gamma|$ in terms of $|V_\gamma|$. For $i \in \{0,1\}$, let $\rho_i = |V_\gamma \cap V_G^i|/n$ and $\rho = |V_\gamma|/n$. Applying Lemma 2.5.7 to the set $V_\gamma \cap V_G^i$, we have that

$$|\partial(V_\gamma \cap V_G^i)| \geq \frac{|V_\gamma \cap V_G^i|}{\rho_i + \frac{\lambda^2}{\Delta^2}(1-\rho_i)} \geq \frac{|V_\gamma \cap V_G^i|}{\rho_i + \frac{\lambda^2}{\Delta^2}}.$$

Now observe that

$$\sum_{i\in\{0,1\}} \rho_i |V_\gamma \cap V_G^i| = \tfrac{1}{n} \sum_{i\in\{0,1\}} |V_\gamma \cap V_G^i|^2 \leq \tfrac{1}{n}|V_\gamma|^2 = \rho|V_\gamma|$$

and hence, using the inequality $\frac{a}{x} + \frac{b}{y} \geq \frac{(a+b)^2}{ax+by}$ which holds for all $a, b, x, y \geq 0$, we obtain

$$\sum_{i\in\{0,1\}} |\partial(V_\gamma \cap V_G^i)| \geq \sum_{i\in\{0,1\}} \frac{|V_\gamma \cap V_G^i|}{\rho_i + \frac{\lambda^2}{\Delta^2}} \geq \frac{|V_\gamma|}{\rho + \frac{\lambda^2}{\Delta^2}} \geq \frac{|V_\gamma|}{3\varepsilon},$$

using that $\rho \leq 2\varepsilon$ and $\varepsilon \geq \frac{\lambda^2}{\Delta^2}$. Since $\varepsilon \in (0, 1/20)$, we have that $|\partial V_\gamma| \geq \frac{|V_\gamma|}{3\varepsilon} - |V_\gamma| \geq \frac{|V_\gamma|}{4\varepsilon}$. Plugging this into (4.12), we obtain

$$w_{G,H}^{B_0,B_1}(\gamma) \leq e^{-|V_\gamma|\left(\frac{1-\delta}{4\varepsilon q}\right)} \leq e^{-\tau|V_\gamma|},$$

where $\tau := \frac{1-\delta}{4\varepsilon q} \geq 10\log(q\Delta) \geq 5 + 3\log((q-1)\Delta^3)$ using that $\varepsilon \leq \frac{1-\delta}{40q\log(q\Delta)}$ and $q \geq 2, \Delta \geq 3$. This finishes the proof of Lemma 4.5.6, after observing that the degree bound for the family $\{(\mathcal{C}_{G,H,\varepsilon}^{B_0,B_1}, w_{G,H}^{B_0,B_1}, J_G) \mid G \in \mathcal{G}_{\Delta,\lambda}^{\text{bip}}\}$ is Δ^3. □

Finally, we can apply Theorems 3.2.2 and 3.2.3, which gives us an efficient algorithm for approximating the partition function of the polymer model, and sampling from its Gibbs distribution.

Corollary 4.5.7. *Let $q \geq 2$ and $\delta \in (0,1)$. Let $H \in \mathbb{R}_{\geq 0}^{q\times q}$ be a symmetric δ-matrix and let $(B_0, B_1) \in \mathcal{K}_H^{max}$. Let $\varepsilon \in (0,1)$ be such that $\varepsilon \geq \lambda^2/\Delta^2$ and $\varepsilon \leq \frac{1-\delta}{40q\log(q\Delta)}$. Then there is a randomised algorithm that takes as input an n-vertex graph $G \in \mathcal{G}_{\Delta,\lambda}^{bip}$ and an accuracy parameter $\varepsilon^* \in (0,1)$ and outputs an ε^*-approximation to $Z_{G,H,\varepsilon}^{B_0,B_1}$ in time $O((n/\varepsilon^*)^2 \log^3(n/\varepsilon^*))$ with probability at least $3/4$. Moreover, there is a randomised algorithm that takes the same input and outputs an ε^*-approximate sample from $\mu_{G,H,\varepsilon}^{B_0,B_1}$ in time $O(n\log(n/\varepsilon^*)\log(1/\varepsilon^*))$.*

4.6 Proof of Theorem 4.2.1

In this section, we combine the results of Sections 4.4 and 4.5 to prove Theorem 4.2.1.

Theorem 4.2.1. *Let $q \geq 2$ be an integer, δ be a real in $(0,1)$, $H \in \mathbb{R}_{\geq 0}^{q \times q}$ be a symmetric δ-matrix. Suppose that Δ, λ satisfy $\frac{\Delta}{\lambda} \geq \frac{100}{1-\delta} q^2 \log(q\Delta)$ and $\Delta \geq \left(\frac{10}{1-\delta} q \log(q\Delta)\right)^4$. Then, there is an* FPRAS *for $\#\mathrm{SPIN}_{H,\Delta,\lambda}$.*

In fact, there is a randomised algorithm that, given a graph $G \in \mathcal{G}^{bip}_{\Delta,\lambda}$ with $n = |V_G^0| = |V_G^1|$ vertices and an accuracy parameter $\varepsilon^* \geq e^{-n/(5q)}$, outputs an ε^*-approximation to $Z_{G,H}$ and an ε^*-sample from the Gibbs distribution $\mu_{G,H}$ in time $O((n/\varepsilon^*)^2 \log^4(n/\varepsilon^*))$.

Proof. As input to the FPRAS, we are given a graph $G \in \mathcal{G}^{\mathrm{bip}}_{\Delta,\lambda}$ and an accuracy parameter $\varepsilon^* \in (0,1)$. We may assume that $n = |V_G^0| = |V_G^1|$ is sufficiently large, otherwise we can compute $Z_{G,H}$ exactly in constant time, by brute force. Similarly, we may assume that $\varepsilon^* \geq 9e^{-n/(4q)}$, otherwise we can compute $Z_{G,H}$ exactly in $O(nq^{2n}) = \mathrm{poly}(1/\varepsilon^*)$ time, by brute force.

Let $\varepsilon = \frac{1-\delta}{50q\log(q\Delta)}$ and observe that, using the lower bounds on $\frac{\Delta}{\lambda}$ and Δ, we have that

$$\varepsilon \leq \frac{1}{240q\log q}, \quad \varepsilon \geq 2q\frac{\lambda}{\Delta} \geq \frac{\lambda^2}{\Delta^2}, \quad \varepsilon^2 \geq \frac{8q^2\log q}{\Delta\log(1/\delta)},$$

where the first inequality follows from $\Delta \geq q^4$, the second from rearranging $\frac{\Delta}{\lambda} \geq 2q/\varepsilon$ (using the lower bound on $\frac{\Delta}{\lambda}$), and the last inequality from using that $\log(1/\delta) \geq 1-\delta$ for all $\delta \in (0,1)$. In particular, the assumption of Lemmas 4.4.2, 4.5.5 and Corollary 4.5.7 are satisfied.

By Corollary 4.5.7, for an arbitrary biclique $(B_0, B_1) \in \mathcal{K}_H^{\max}$, we can obtain an $(\varepsilon^*/8)$-approximation to $Z_{G,H,\varepsilon}^{B_0,B_1}$ in $O((n/\varepsilon^*)^2\log^3(n/\varepsilon^*))$ time, with probability at least $3/4$. Taking the median of $O(\log(1/\varepsilon^*))$ runs of this algorithm, we therefore obtain $\widehat{Z}_{G,H,\varepsilon}^{B_0,B_1}$ which is an $(\varepsilon^*/8)$-approximation to $Z_{G,H,\varepsilon}^{B_0,B_1}$ with probability at least $1-\varepsilon^*/(16|\mathcal{K}_H^{\max}|)$. By a union bound over the bicliques in $\mathcal{K}_H^{\max}$, it follows that

$$\widehat{Z}_{G,H,\varepsilon}^{\mathrm{polymer}} := \sum_{(B_0,B_1)\in\mathcal{K}_H^{\max}} |B_0|^n |B_1|^n \cdot \widehat{Z}_{G,H,\varepsilon}^{B_0,B_1},$$

is an $(\varepsilon^*/8)$-approximation to $Z^{\text{polymer}}_{G,H,\varepsilon}$ (cf. Definition 4.5.3) with probability at least $1-\varepsilon^*/16$. By Lemma 4.5.5, $Z^{\text{polymer}}_{G,H,\varepsilon}$ is an $(\varepsilon^*/8)$-approximation to $Z_{G,H,\varepsilon}$ which, by Lemma 4.4.2, is an $(\varepsilon^*/8)$-approximation to $Z_{G,H}$. It therefore follows that $\widehat{Z}^{\text{polymer}}_{G,H,\varepsilon}$ is a $(3\varepsilon^*/8)$-approximation, and hence an ε^*-approximation, to $Z_{G,H}$ with probability at least $1-\varepsilon^*/16$. The total run-time of the algorithm is therefore $O((n/\varepsilon^*)^2\log^4(n/\varepsilon^*))$.

For the sampling algorithm, we assume again that $\varepsilon^* \geq 9e^{-n/(4q)}$. We first sample a biclique $\widehat{\mathbf{B}} = (B_0, B_1) \in \mathcal{K}^{\max}_H$ with probability $\frac{|B_0|^n|B_1|^n\widehat{Z}^{B_0,B_1}_{G,H,\varepsilon}}{\widehat{Z}^{\text{polymer}}_{G,H,\varepsilon}}$, where $\widehat{Z}^{B_0,B_1}_{G,H,\varepsilon}$ and $\widehat{Z}^{\text{polymer}}_{G,H,\varepsilon}$ are as before. Then, using Corollary 4.5.7, we sample a polymer configuration $\widehat{\mathbf{\Gamma}}$ whose distribution is at distance at most $\varepsilon^*/6$ from $\mu^{B_0,B_1}_{G,H,\varepsilon}$. We output $\hat{\boldsymbol{\sigma}} = \mathtt{Spin}_{\widehat{\mathbf{B}}}(\widehat{\mathbf{\Gamma}})$, where for a biclique $B = (B_0, B_1)$ and a polymer configuration Γ, $\mathtt{Spin}_B(\Gamma)$ is a random configuration τ obtained as follows:

- For every vertex $u \in \cup\Gamma$, we set $\tau(u) = \sigma_\Gamma(u)$.
- For $u \in V^i_G \backslash (\cup\Gamma)^+$ with $i \in \{0,1\}$, we assign a random spin from B_i uniformly at random.
- For $u \in \partial(\cup\Gamma) \cap V^i_G$ with $i \in \{0,1\}$, for $j \in B_i$ we set $\tau(u) = j$ with probability $\frac{1}{F_u}\prod_{v\in\partial u\cap(\cup\Gamma)} H_{j,\sigma_\Gamma(v)}$ where $F_u := \sum_{j\in B_i}\prod_{v\in\partial u\cap(\cup\Gamma)} H_{j,\sigma_\Gamma(v)}$.

We claim that $\hat{\boldsymbol{\sigma}}$ is an ε^*-approximate sample from the Gibbs distribution $\mu_{G,H}$. To prove this, let $\mathbf{B}$, $\mathbf{\Gamma}$, and $\boldsymbol{\sigma}$ be the analogues of $\widehat{\mathbf{B}}$, $\widehat{\mathbf{\Gamma}}$, and $\hat{\boldsymbol{\sigma}}$, respectively, when there is no error, more precisely:

1. $\mathbf{B} = (B_0, B_1)$ with probability $|B_0|^n|B_1|^n\frac{Z^{B_0,B_1}_{G,H,\varepsilon}}{Z^{\text{polymer}}_{G,H,\varepsilon}}$.
2. Conditioned on $\mathbf{B} = (B_0, B_1)$, $\mathbf{\Gamma} \sim \mu^{B_0,B_1}_{G,H,\varepsilon}$ and $\boldsymbol{\sigma} = \mathtt{Spin}_{\mathbf{B}}(\mathbf{\Gamma})$.

With probability $1-\varepsilon^*/16$, we have that, for all bicliques $(B_0, B_1) \in \mathcal{K}^{\max}_H$, $\widehat{Z}^{B_0,B_1}_{G,H,\varepsilon}$ and $\widehat{Z}^{\text{polymer}}_{G,H,\varepsilon}$ are $(\varepsilon^*/8)$-approximations to $\widehat{Z}^{B_0,B_1}_{G,H,\varepsilon}$ and $\widehat{Z}^{\text{polymer}}_{G,H,\varepsilon}$ respectively, and conditioned on this the total variation distance between the distributions of $\widehat{\mathbf{B}}$ and $\mathbf{B}$ is at most $\frac{1+\varepsilon^*/8}{1-\varepsilon^*/8} - 1 \leq 3\varepsilon^*/8$. It follows that the total variation distance between the distributions of $\widehat{\mathbf{B}}$ and $\mathbf{B}$ is at most $3\varepsilon^*/8 + \varepsilon^*/16 \leq \varepsilon^*/2$, and so there is a coupling so that $\Pr[\widehat{\mathbf{B}} \neq \mathbf{B}] \leq \varepsilon^*/2$. Conditioned on $\widehat{\mathbf{B}} = \mathbf{B}$,

we have that the total variation distance between $\widehat{\mathbf{\Gamma}}$ and $\mathbf{\Gamma}$ is at most $\varepsilon^*/6$, so there is a coupling which further satisfies $\Pr[\widehat{\mathbf{\Gamma}} \neq \mathbf{\Gamma} \mid \widehat{\mathbf{B}} = \mathbf{B}] \leq \varepsilon^*/6$. Finally, conditioned on $\widehat{\mathbf{B}} = \mathbf{B}$ and $\widehat{\mathbf{\Gamma}} = \mathbf{\Gamma}$, we can clearly couple $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ so that they agree. It follows that the total variation distance between $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ it at most $2\varepsilon^*/3$, so the result will follow by showing that the distribution of $\boldsymbol{\sigma}$ and $\mu_{G,H}$ are at distance at most $3e^{-n/(4q)} \leq \varepsilon^*/3$.

For this, we will consider the set of configurations $\widehat{\Sigma} := \Sigma_{G,H,\varepsilon} \backslash \Sigma^{\mathsf{overlap}}_{G,H,3\varepsilon}$, where recall from Definition 4.4.1 that $\Sigma_{G,H,\varepsilon}$ is the set of configurations τ with $\sum_{i\in\{0,1\}} \left|\tau^{-1}(B_i) \cap V_G^i\right| \geq (1-\varepsilon)|V_G|$, and from Definition 4.4.3 that $\Sigma^{\mathsf{overlap}}_{G,H,3\varepsilon}$ is the set of configurations τ such that $\tau \in \Sigma^{B_0,B_1}_{G,H,3\varepsilon} \cap \Sigma^{C_0,C_1}_{G,H,3\varepsilon}$ for some distinct maximal bicliques $(B_0,B_1),(C_0,C_1)$.

Consider arbitrary $\tau \in \widehat{\Sigma}$ and let (B_0,B_1) be such that $\tau \in \Sigma^{B_0,B_1}_{G,H,\varepsilon}$. For $i \in \{0,1\}$, let $T_i = V_G^i \cap \tau^{-1}([q]\backslash B_i)$, and $S_1,\ldots,S_k$ denote the G^3-connected components of $T := T_0 \cup T_1$; note, since $\tau \in \widehat{\Sigma}$, we have $|T| \leq \varepsilon|V_G|$. Consider the polymer configuration Γ_τ which is the union of the polymers $(S_1,\ tau|_{S_1}),\ldots,(S_k,\tau|_{S_k})$. We next find how the sample $\boldsymbol{\sigma}$ can equal τ; we claim that

$$\boldsymbol{\sigma} = \tau \text{ iff } \mathbf{B} = (B_0,B_1),\ \mathbf{\Gamma} = \Gamma_\tau,\ \boldsymbol{\sigma}|_{V_G\backslash(\cup\Gamma)} = \tau|_{V_G\backslash T}. \tag{4.13}$$

The reverse direction of this equivalence is immediate, noting that the equality $\mathbf{\Gamma} = \Gamma_\tau$ implies that $(\cup\Gamma) = T$ and that $\boldsymbol{\sigma}|_{\cup\mathbf{\Gamma}} = \tau|_T$. So, we focus on the forward direction and assume that $\boldsymbol{\sigma} = \tau$. Then

1. $\mathbf{B} = (B_0,B_1)$. Otherwise, if $\mathbf{B} = (C_0,C_1)$ for some $(C_0,C_1) \neq (B_0,B_1)$, then by Lemma 4.5.4 we would have that $\tau \in \Sigma^{C_0,C_1}_{G,H,3\varepsilon}$, contradicting that $\tau \notin \Sigma^{\mathsf{overlap}}_{G,H,3\varepsilon}$.

2. $\cup\mathbf{\Gamma} = T_0 \cup T_1$. Since $\mathbf{B} = (B_0,B_1)$ from Item 1, for $i \in \{0,1\}$ we have that $u \in V_G^i \cap (\cup\mathbf{\Gamma})$ iff $\boldsymbol{\sigma}(u) \in [q]\backslash B_i$ iff $\tau(u) \in [q]\backslash B_i$ iff $u \in T_i$.

3. $\boldsymbol{\sigma}|_{\cup\mathbf{\Gamma}} = \tau|_T$ and $\boldsymbol{\sigma}|_{V_G\backslash(\cup\Gamma)} = \tau|_{V_G\backslash T}$. This follows from Item 2 and $\boldsymbol{\sigma} = \tau$.

4. $\mathbf{\Gamma} = \Gamma_\tau$. This follows from Items 2 and 3.

From (4.13) and the definition of the sampling procedure for $\boldsymbol{\sigma}$, we therefore have that $\boldsymbol{\sigma} = \tau$ with probability

$$\frac{|B_0|^n|B_1|^n Z^{B_0,B_1}_{G,H,\varepsilon}}{Z^{\text{polymer}}_{G,H,\varepsilon}} \cdot \frac{w^{B_0,B_1}_{G,H}(\Gamma_\tau)}{Z^{B_0,B_1}_{G,H,\varepsilon}} \times \prod_{i\in\{0,1\}} \frac{1}{|B_i|^{|V^i_G\setminus T^+|}} \prod_{u\in\partial T} \Big(\frac{1}{F_u} \prod_{v\in T\cap\partial u} H_{\tau(u),\tau(v)}\Big). \tag{4.14}$$

From (4.8) applied to the polymer configuration Γ_τ, we have, using that $\cup\Gamma_\tau = T$ and $\boldsymbol{\sigma}_{\Gamma_\tau} = \tau$, that

$$|B_0|^n|B_1|^n w^{B_0,B_1}_{G,H}(\Gamma_\tau) = \prod_{i\in\{0,1\}} |B_i|^{|V^i_G\setminus T^+|} \prod_{\{u,v\}\in E_G(T)} H_{\tau(u),\tau(v)} \prod_{u\in\partial T} F_u,$$

and hence we obtain that the expression in (4.14) equals

$$\frac{\prod_{\{u,v\}\in E_G(T)} H_{\tau(u),\tau(v)} \prod_{u\in\partial T}\prod_{v\in T\cap\partial u} H_{\tau(u),\tau(v)}}{Z^{\text{polymer}}_{G,H,\varepsilon}} = \frac{w_{G,H}(\tau)}{Z^{\text{polymer}}_{G,H,\varepsilon}},$$

where the last equality follows by noting that edges that are not in $E_G(T) \cup E_G(T,\partial T)$ contribute a factor of 1 in the weight of T (since their endpoints are assigned spins of the biclique). So, we have shown that $\boldsymbol{\sigma} = \tau$ with probability $w_{G,H}(\tau)/Z^{\text{polymer}}_{G,H,\varepsilon}$.

Let $p_{\boldsymbol{\sigma}}$ be the probability that $\boldsymbol{\sigma} \in \widehat{\Sigma}$ and p be the aggregate weight of configurations in the Gibbs distribution $\mu_{G,H}$ in $\widehat{\Sigma}$ (that is, p is the probability of seeing a configuration in $\hat{\Sigma}$ when a sample is drawn from $\mu_{G,H}$). Then, using that $Z^{\mathsf{overlap}}_{G,H,3\varepsilon} \le e^{-n/(3q)} Z_{G,H,\varepsilon}$ from Lemma 4.4.4 and $Z^{\text{polymer}}_{G,H,\varepsilon} \le (1+e^{-n/(4q)}) Z_{G,H,\varepsilon}$ from Lemma 4.5.5, we have that

$$\begin{aligned} p_{\boldsymbol{\sigma}} &\ge \frac{1}{Z^{\text{polymer}}_{G,H,\varepsilon}} \sum_{\tau\in\widehat{\Sigma}} w_{G,H}(\tau) \\ &\ge \frac{Z_{G,H,\varepsilon} - Z^{\mathsf{overlap}}_{G,H,3\varepsilon}}{Z^{\text{polymer}}_{G,H,\varepsilon}} \\ &\ge \frac{\big(1-e^{-n/(3q)}\big) Z_{G,H,\varepsilon}}{\big(1+e^{-n/(4q)}\big) Z_{G,H,\varepsilon}} \\ &\ge 1 - 2e^{-n/(4q)}, \end{aligned} \tag{4.15}$$

while for p, using that $Z_{G,H} \le (1+e^{-n}) Z_{G,H,\varepsilon}$ from Lemma 4.4.2, we have the bound

$$\begin{aligned} p &\geq \frac{1}{Z_{G,H}} \sum_{\tau \in \widehat{\Sigma}} w_{G,H}(\tau) \\ &\geq \frac{Z_{G,H,\varepsilon} - Z^{\mathsf{overlap}}_{G,H,3\varepsilon}}{Z_{G,H}} \\ &\geq \frac{\big(1 - e^{-n/(3q)}\big) Z_{G,H,\varepsilon}}{(1 + e^{-n}) Z_{G,H,\varepsilon}} \\ &\geq 1 - 2e^{-n/(3q)}. \end{aligned} \tag{4.16}$$

It follows that the total variation distance between the distribution of $\boldsymbol{\sigma}$ and $\mu_{G,H}$ is bounded above by

$$D := \frac{1}{2}\big((1 - p_{\boldsymbol{\sigma}}) + (1 - p) + M\big), \qquad \text{where } M := \sum_{\tau \in \widehat{\Sigma}} w_{G,H}(\tau) \Big| \frac{1}{Z^{\mathrm{polymer}}_{G,H,\varepsilon}} - \frac{1}{Z_{G,H}} \Big|.$$

Using Lemma 4.4.2 and Lemma 4.5.5, we have the bound

$$M \leq \Big| \frac{Z_{G,H}}{Z^{\mathrm{polymer}}_{G,H,\varepsilon}} - 1 \Big| \leq 2e^{-n/(4q)}. \tag{4.17}$$

Combining (4.15), (4.16) and (4.17), we obtain that $D \leq 3e^{-n/(4q)}$, i.e., the distance between the distribution of $\boldsymbol{\sigma}$ and $\mu_{G,H}$ is at most $3e^{-n/(4q)}$, as claimed. □

Chapter 5

Random graphs of unbounded degree

The work in this chapter is based on the following paper:

[42] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. *Information and Computation*, page 104894, 2022

An extended abstract also appeared at:

[41] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. In Mary Wootters and Laura Sanità, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, volume 207 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:13, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik

One of the key challenges in the analysis of algorithms for polymer models, is controlling the growth rate of the number of polymers. Typically, this is achieved by restricting the host graph of the polymer model to have bounded maximum degree. This assumption is often restrictive and obstructs the applicability of the method to more general graphs. For example, sparse random graphs typically have bounded average degree and good expansion properties, but they include vertices with unbounded degree, and therefore are excluded

from the polymer model framework presented in Chapter 3.

In this chapter, we develop a less restrictive framework for polymer models, which relaxes the standard bounded-degree assumption by measuring the 'size' of a polymer in terms of its number of edges. This allows us to bound the growth rate of the number of polymers in terms of the total degree of these polymers, which in turn can be more easily related to the expansion properties of the underlying graph. To apply our methods, we consider random graphs having fixed degree sequence (with minimum degree at least 3, but possibly unbounded maximum degree) and obtain approximation algorithms for the ferromagnetic Potts model, which is a standard benchmark for the polymer model framework.

5.1 Introduction

Algorithms that use the polymer model framework typically apply to classes of graphs that have strong expansion properties relative to their local growth rates. The local growth rate is often controlled by constraining the maximum degree of the host graph to be at most some fixed constant. Examples of such algorithms include bounded-degree expanders [68, 74, 22, 18, 19, 3, 40, 58] and the d-dimensional grid [59, 11, 68, 62]. The purpose of the work in this chapter is to extend the polymer model framework, by relaxing the bounded-degree assumption and using alternative methods to capture the growth of the graph.

As we have already seen, recent work [59, 68] has resulted in algorithms that bypass the worst-case hardness of approximate counting and sampling from low temperature spin systems, on certain classes of bounded-degree graphs with good expansion properties. The rough intuition for the q-colour Potts model is that, for large β, due to the expansion properties, the colourings with non-negligible weight are close to the so-called ground-states of the model, i.e., the q configurations in which all vertices get the same colour. Polymer models capture the deviations that configurations take from these ground states. The Potts model can then be decomposed into q polymer models, one for each phase, each of which can be studied using relatively streamlined algorithmic methods (based on either interpolation [2] or the Markov chain based approach introduced in Chapter 3).

The polymer model framework has found multiple algorithmic applications in far more general settings [59, 11, 62, 68, 65, 74, 22, 37, 18, 19, 40, 58]. In spite of these advances, the application of polymer models relies on the fact that the maximum degree of the host graph is bounded. This fact is used to control the number of polymers of a given size (which is a crucial step in the analysis of these algorithms). As a result of this limitation, applications to several other interesting classes of graphs are ruled out, including, for example, sparse random graphs which have bounded average degree and good expansion properties, but include vertices with unbounded degree.

5.2 Main results

In this chapter, we propose a framework for polymer models that overcomes the bounded-degree limitations of previous algorithms, by revisiting the Markov chain approach of Chapter 3. We introduce a new condition (Definition 5.3.2) which requires that the weight of each polymer decays exponentially in its total degree (the sum of the degrees of the vertices in the polymer) instead of decaying exponentially in the polymer's size. This new condition allows us to prove rapid mixing for a Markov chain which is an adaptation (which considers edges instead of vertices) of the polymer dynamics Markov chain of Chapter 3. Crucially, the fact that the new condition is formulated in terms of the total degree of a polymer allows us to relax the assumption that the instance has bounded degree.

Recall the definitions relating to the polymer model framework introduced in Section 2.6. To simplify the presentation of the new framework, our definitions will apply to polymer models in which the host graph is the same as the underlying graph. We may therefore write $\{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ to identify a family of polymer models corresponding to the class of graphs $\mathcal{G}$, since $J_G = G$ for all $G \in \mathcal{G}$. We will also dispense with the notion of a degree bound for a family of polymer models, as it is no longer relevant. Subject to new conditions, we give an efficient algorithm for sampling from the Gibbs distribution of a polymer model, and approximating its partition function.

Lemma 5.2.1. *Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G}$ be a family of computationally feasible (Definition 5.3.5) q-spin polymer*

models satisfying the polymer sampling condition (Defintion 5.3.2) with constant $\tau \geq 3\log(8e^3(q-1))$.

There are randomised algorithms which, given as input a graph $G \in \mathcal{G}$ *with* m *edges and an accuracy parameter* $\varepsilon > 0$*, output an* ε*-sample from* μ_G *in* $O\big(m\log(m/\varepsilon)\log(1/\varepsilon)\big)$ *time, and an* ε*-approximation to* Z_G*, with probability at least* $3/4$*, in* $O\big((m/\varepsilon)^2\log^3(m/\varepsilon)\big)$ *time.*

As an application of this revised method, we consider the q-state ferromagnetic Potts model on sparse random graphs with a given degree sequence (of possibly unbounded maximum degree), as detailed below.

Definition 5.2.2. Let d be a positive real number and n be a positive integer. We define $\mathcal{D}_{n,d}$ to be the set of all degree sequences $\{x_1, x_2, ..., x_n\}$ that satisfy

1. For all $i \in [n]$, $3 \leq x_i \leq n^\rho$ where $\rho = \frac{1}{50}$, and
2. $\sum_{i\in[n]} x_i^2 \leq dn$.

Note that $\mathcal{D}_{n,d}$ is empty unless $d \geq 9$. The assumption that all degrees are greater than or equal to 3 (rather than 2) guarantees that the random graph G is connected and has good expansion properties. The degree lower bound also means that our results do not apply to Erdős-Rényi random graphs. The upper bound on the degrees is mild and can in fact be relaxed somewhat further (but in general cannot be made to be linear in n due to the sparsity assumption in Item 2). Note that d is fixed and independent of n.

We give an efficient algorithm for approximately sampling from and approximating the partition function of the ferromagnetic Potts model on random graphs with a given degree sequence, for all sufficiently large β.

Theorem 5.2.3. *Let* d *be a real number and let* $q \geq 3$ *be an integer. For the ferromagnetic Potts model, there is* β_0 *such that for all* $\beta \geq \beta_0$ *there is a poly-time approximate sampling algorithm for* $\mu_{G,q,\beta}$ *and an* FPRAS *for* $Z_{G,q,\beta}$ *that work with high probability on random graphs* $G \sim \mathcal{G}(n,\vec{x})$ *for any degree sequence* $\vec{x} \in \mathcal{D}_{n,d}$.

Remark 5.2.4. Note that β_0 depends on d and q, and our arguments later (see Remark 5.6.7) show that $\beta_0 = Cd\log d\log q$ for some $C > 0$ (independent

of d or q). If the desired accuracy ε is at least e^{-n} then the running time of the sampling algorithm is $O\big(n\log(n/\varepsilon)\log(1/\varepsilon)\big)$ and the running time of the FPRAS is $O\big((n/\varepsilon)^2\log^3(n/\varepsilon)\big)$.

Whilst this result applies to classes of graphs for which the results of Chapter 3 do not, it is not a strict generalisation. For example, Theorem 5.2.3 does not apply to classes of bounded-degree graphs which are expanders, but not degree-expanders (see Definition 5.5.1). Whilst we could apply Theorem 5.2.3 to obtain an algorithm for the low temperature Potts model on the random regular graph (as obtained in Chapter 3), this would result in a smaller range of parameters due to the more restrictive conditions that we require here.

Chapter 3, We further remark that the bounded-degree assumption has also been relaxed in [62] for the ferromagnetic Potts model on lattice graphs; the approach therein, however, is tailored to a certain flow representation of the Potts model, which is used as a basis for the corresponding polymer models and therefore does not extend to general spin systems. Our approach applies to general polymer models, as detailed in the next section, and our focus on the ferromagnetic Potts model is mainly to illustrate the method without further technical overhead; the approach, we expect, can be adapted to general spin systems on bipartite random graphs with a given degree sequence (analogously to the work of Chapter 4).

5.3 Revised polymer model framework

We will present a Markov chain for sampling from the Gibbs distribution of a polymer model, subject to a certain condition on the weights of its polymers. This condition is analogous to the polymer mixing condition of Definition 2.6.1, except that we consider edges instead of vertices. Consider a polymer model $(\mathcal{C}_G, w_G)$ defined for a host graph $G = (V_G, E_G)$. For a polymer $\gamma \in \mathcal{P}_G$, let E_γ denote the set of edges of G with at least one endpoint in V_γ.

Definition 5.3.1. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}_\mathcal{G}$ satisfies the polymer mixing condition with constant $\theta \in (0,1)$ if $\sum_{\gamma' \nsim \gamma} |E_{\gamma'}| \cdot w_G(\gamma') \leq \theta |E_\gamma|$ for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{C}_G$.

As was the case in the bounded-degree setting of Section 3.3, in order to use the modified polymer dynamics as an efficient algorithm for computing an approximate sample from μ_G, we need to introduce some further conditions on the weights of polymers. We refine the polymer model framework by introducing a new polymer sampling condition (originally stated in Definition 2.6.3), which requires that the weight of each polymer decays exponentially in its total degree. This condition is as follows.

Definition 5.3.2. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_{\mathcal{G}} = \{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}_{\mathcal{G}}$ satisfies the polymer sampling condition with constant $\tau > 0$ if $w_G(\gamma) \leq e^{-\tau \deg_G(V_\gamma)}$ for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{C}_G$.

The new polymer sampling condition is analogous to the original one in Definition 2.6.3, except that the original condition requires the weight of a polymer to decay exponentially in its size and, in particular, that the constant τ is sufficiently large relative to the maximum degree of G. The new condition relaxes this, allowing us to choose the constant τ in Definition 5.3.2 so that it does not depend on the maximum degree of the host graph, which is how we overcome the limitations of previous work. This new condition also fits well with the original abstract polymer model framework of [71], where the notion of the "size" of a polymer is an abstract function.

Technically, the improvement comes from the fact that previous work relies on bounding the number of connected vertex subsets of a given size (with bounds that depend on the maximum degree of the graph), but here we are able to instead rely on the following lemma, which bounds the number of connected vertex subsets with a given total degree and this enables us to avoid restricting the maximum degree of the graph. Recall that $\deg_G(S)$ denotes the total degree of vertices in S, for a vertex set $S \subseteq V_G$,

Lemma 5.3.3. *Let $G = (V_G, E_G)$ be a graph, let $v \in V_G$, and let $\ell \geq 1$ be an integer. The number of connected vertex subsets $S \subseteq V_G$ such that $v \in S$ and $\deg_G(S) = \ell$ is at most $(2e)^{2\ell-1}$.*

Proof. Let $N(G, v, \ell)$ be the set of subtrees $T = (V_T, E_T)$ of G such that $v \in V_T$, $\deg_G(V_T) = \ell$. We will show that $|N(G, v, \ell)| \leq (2e)^{2\ell-1}$, which gives us the

desired result for the following reason. Let $S \subseteq V_G$ be a connected vertex subset such that $v \in S$ and $\deg_G(S) = \ell$. Since S is connected, it has at least one spanning tree $T = (V_T = S, E_T)$ such that $v \in V_T$ and $\deg_G(V_T) = \ell$. Since S is the unique connected vertex subset that T spans, this gives us an injective map from the set of all connected vertex subsets containing v with total degree ℓ, to $N(G, v, \ell)$.

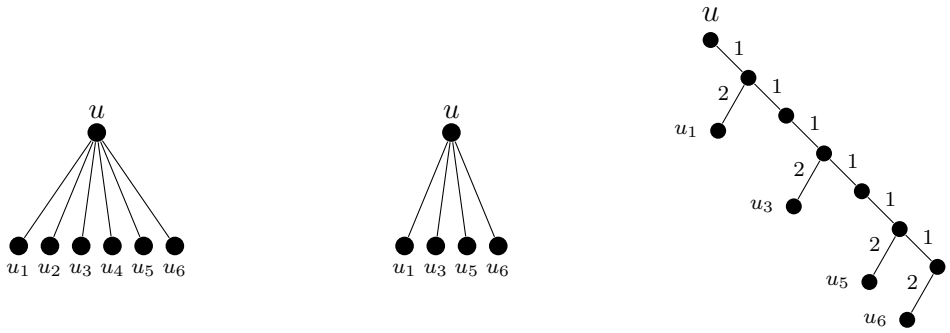
We now give an injective map from $N(G, v, \ell)$ to $T^*(2\ell, 3)$ – the set of subtrees of size 2ℓ that contain the root, of the infinite rooted 3-regular tree. By a result of Bollobás [7, p. 129], we know that $|T^*(2\ell, 3)|$ is at most $(2e)^{2\ell-1}$. Let $T = (V_T, E_T)$ be a subtree from $N(G, v, \ell)$. We will map T to a rooted subtree $T' = (V_{T'}, E_{T'})$ from $T^*(2\ell, 3)$. For each vertex of V_G, fix an ordering of its neighbours. In the infinite rooted 3-regular tree, label the edges incident to the root with $\{1, 2, 3\}$, and for each other vertex label the edges connecting it to its two children with $\{1, 2\}$. As we construct T', we will label its edges so that it is clear which subtree from $T^*(2\ell, 3)$ we are constructing, we will also label some of its vertices. We construct T' as follows (see Figure 5.1 for an example of the following construction).

1. Add the root to $V_{T'}$ and label it v.

2. While there is a labelled vertex of T' (call its label u) such that u has a child w in T but no vertex of T' is labelled w, then we do the following. First, we create a path P of length $\deg_G(u)$ where each edge is labelled 1. We then connect the vertex of T' labelled u to P via an edge labelled 1. Finally, for $1 \leq i \leq \deg_G(u)$, we connect a vertex labelled w to the i^{th} vertex of P via an edge labelled 2, if w is the i^{th} neighbour of u in G and w is a child of u in T.

Each $T \in N(G, v, \ell)$ maps to a different $T' \in T^*(2\ell, 3)$. When constructing T', we used edge labels from $\{1, 2, 3\}$, therefore the maximum degree of T' is 3. For each $v \in V_T$, we added at most $2 \deg_G(v)$ vertices to T', therefore the size of T' is at most $2 \deg_G(V_T) = 2\ell$. □

In addition to the above bound, we will use the fact that these connected vertex subsets can be enumerated in time exponential in the total degree ℓ (see Lemma 5.4.3). Although the bound in Lemma 5.3.3 is exponential in ℓ, this



(a) Neighbourhood of u in G (b) Neighbourhood of u in T (c) Neighbourhood of u in T'

Figure 5.1: Neighbourhoods of a vertex.

will be mitigated by the fact that the new polymer sampling condition ensures that the weight of each polymer is exponentially small in its total degree. This allows us to prove that if the new polymer sampling condition holds, then the new polymer mixing condition does, also.

The following lemma shows that if a family of polymer models satisfies the polymer sampling condition then it also satisfies the polymer mixing condition. This will be convenient when applying our algorithms as it is, in general, easier to show that the polymer sampling condition holds for a given family of polymer models, than to show directly that the polymer mixing condition does. However, the polymer mixing condition is used to bound the mixing time of the polymer dynamics.

Lemma 5.3.4. *Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_{\mathcal{G}} = \{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. If $\mathcal{F}_{\mathcal{G}}$ satisfies the polymer sampling condition (Definition 5.3.2) with constant $\tau \geq 3\log(8e^3(q-1))$, then $\mathcal{F}_{\mathcal{G}}$ satisfies the polymer mixing condition (Definition 5.3.1) with constant $\theta = \frac{1}{e}$.*

Proof. Let $G \in \mathcal{G}$ and let $\gamma \in \mathcal{C}_G$ be an arbitrary polymer. We have that

$$\sum_{\gamma' \nsim \gamma} |E_{\gamma'}| \cdot w_G(\gamma') \leq \sum_{\{u,v\} \in E_\gamma} \left(\sum_{\gamma' : u \in V_{\gamma'}} |E_{\gamma'}| \cdot w_G(\gamma') + \sum_{\gamma' : v \in V_{\gamma'}} |E_{\gamma'}| \cdot w_G(\gamma') \right),$$

therefore the result will follow if we are able to show for all $v \in V_\gamma \cup \partial_G V_\gamma$ that

$$\sum_{\gamma' : v \in V_{\gamma'}} |E_{\gamma'}| \cdot w_G(\gamma') \leq \frac{1}{2e}.$$

We can re-write the left-hand side of the above as follows

$$\sum_{\gamma' : v \in V_{\gamma'}} |E_{\gamma'}| \cdot w_G(\gamma') \leq \sum_{\ell \geq 1} \sum_{\substack{\gamma' :\ v \in V_{\gamma'}, \\ \deg_G(V_{\gamma'}) = \ell}} |E_{\gamma'}| \cdot w_G(\gamma').$$

By Lemma 5.3.3, we know that there are at most $(2e)^{2\ell-1}$ connected vertex subsets S such that $v \in S$, and $\deg_G(S) = \ell$. For each such S there are $(q-1)^{|S|} \leq (q-1)^\ell$ ways to assign spins to its vertices, therefore there are at most $(2e)^{2\ell-1}(q-1)^\ell$ polymers γ' such that $v \in V_{\gamma'}$ and $\deg_G(V_{\gamma'}) = \ell$. Furthermore, since $\mathcal{F}_\mathcal{G}$ satisfies the polymer sampling condition, we know that each of these polymers satisfies $w_G(\gamma') \leq e^{-\tau\ell}$ where $\tau \geq 3\log(8e^3(q-1))$. Combining these facts with the observation that $|E_{\gamma'}| \leq \deg_G(V_{\gamma'}) \leq e^{\deg_G(V_{\gamma'})}$ for all $\gamma' \in \mathcal{C}_G$, we obtain that

$$\sum_{\gamma' : v \in V_{\gamma'}} |E_{\gamma'}| \cdot w_G(\gamma') \leq \sum_{\ell \geq 1} (2e)^{2\ell-1}(q-1)^\ell e^\ell e^{-\tau\ell} \leq \frac{1}{2e} \sum_{\ell \geq 1} \left(\frac{4e^3(q-1)}{e^\tau} \right)^\ell.$$

Finally, since $\tau \geq \log(8e^3(q-1))$, the result follows. □

In contrast to the conditions in Section 2.6, the two new conditions that we have introduced consider edges – this is because we modify the polymer dynamics algorithm to sample edges instead of vertices. Subject to these new conditions, the techniques of Chapter 3 can be adapted to show that the modified polymer dynamics mixes rapidly, therefore giving the efficient algorithm for sampling from the Gibbs distribution of a polymer model. To obtain these efficient algorithms, we will also require the following mild condition on the polymer model.

Definition 5.3.5. Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_\mathcal{G} = \{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models. We say that $\mathcal{F}_\mathcal{G}$ is computationally feasible if for all $G \in \mathcal{G}$ and all $\gamma \in \mathcal{P}_G$, it is possible to decide whether $\gamma \in \mathcal{C}_G$ and to compute $w_G(\gamma)$, if it is, in $O(e^{\deg_G(V_\gamma)})$ time.

Computational feasibility serves exactly the same purpose as it did in Definition 2.6.2, which required that the same operations are able to be carried out in time depending on $|V_\gamma|$ (instead of $\deg_G(V_\gamma)$ as is the case here).

5.4 Revised polymer dynamics

We will define a Markov chain to sample from μ_G, that mixes rapidly if the polymer mixing condition is met. This Markov chain is essentially the same as the polymer dynamics of Algorithm 1, except that we sample edges instead of vertices. For $e \in E_G$, let $\mathcal{A}(e)$ be the set of polymers γ such that $e \in E_\gamma$. Let ν_e be the probability distribution on $\mathcal{A}(e) \cup \emptyset$ defined by

$$\nu_e(\gamma) = w_G(\gamma) \quad \text{and} \quad \nu_e(\emptyset) = 1 - \sum_{\gamma \in \mathcal{A}(e)} \nu_e(\gamma).$$

If the polymer mixing condition is satisfied, then ν_e gives a well-defined probability distribution. To see this, let $\gamma \in \mathcal{A}(e)$ be such that $|E_\gamma|$ is least. Applying the polymer mixing condition, we obtain that

$$\sum_{\substack{\gamma' \nsim \gamma \\ \gamma' \in \mathcal{A}(e)}} |E_{\gamma'}| \cdot w_G(\gamma') \leq \sum_{\gamma' \nsim \gamma} |E_{\gamma'}| \cdot w_G(\gamma') < |E_\gamma|.$$

Thus,

$$\sum_{\gamma' \in \mathcal{A}(e)} \nu_e(\gamma') = \sum_{\substack{\gamma' \nsim \gamma \\ \gamma' \in \mathcal{A}(e)}} w_G(\gamma') \leq \sum_{\substack{\gamma' \nsim \gamma \\ \gamma' \in \mathcal{A}(e)}} \frac{|E_{\gamma'}|}{|E_\gamma|} \cdot w_G(\gamma') \leq 1.$$

Let $t \geq 0$ be an integer. The polymer dynamics Markov chain transitions from $\Gamma_t \in \Omega_G$ to $\Gamma_{t+1} \in \Omega_G$ according to the following rules.

Algorithm 4: Polymer dynamics

1. Choose $e \in E_G$ uniformly at random. Let γ_e be the (unique) polymer in $\Gamma_t \cap \mathcal{A}(e)$, if it exists; otherwise, let $\gamma_e = \emptyset$.
2. Mutually exclusively
   (a) with probability 1/2, let $\Gamma_{t+1} = \Gamma_t \setminus \{\gamma_e\}$, and
   (b) with probability 1/2, sample a polymer γ from ν_e, and let $\Gamma_{t+1} = \Gamma_t \cup \{\gamma\}$ if this is compatible; otherwise, let $\Gamma_{t+1} = \Gamma_t$.

The following lemma shows that the unique stationary distribution of the polymer dynamics is the Gibbs distribution of the polymer model.

Lemma 5.4.1. *Let $G = (V_G, E_G)$ be a graph and let $(\mathcal{C}_G, w_G)$ be a polymer model. The unique stationary distribution of the polymer dynamics is μ_G.*

Proof. One update of the polymer dynamics changes the polymer configuration by at most one polymer. Let $\Gamma' = \Gamma \cup \{\gamma\}$, then

$$\frac{\mu_G(\Gamma')}{\mu_G(\Gamma)} = w_G(\gamma) = \frac{w_G(\gamma)|E_\gamma|/2|E_G|}{|E_\gamma|/2|E_G|} = \frac{\Pr(\Gamma \to \Gamma')}{\Pr(\Gamma' \to \Gamma)}.$$

It therefore follows by detailed balance that μ_G is a stationary distribution of the polymer dynamics.

Finally, the polymer dynamics are irreducible since we can move from any $\Gamma \in \Omega_G$ to any $\Gamma' \in \Omega_G$ by adding and removing polymers (for example, via the empty set). The polymer dynamics are also aperiodic since there are self-loops. □

The following lemma shows that if the polymer mixing condition is satisfied, then the polymer dynamics Markov chain mixes rapidly.

Lemma 5.4.2. *Let $\mathcal{G}$ be a class of graphs and let $\{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of polymer models that satisfies the polymer mixing condition (see Definition 5.3.1). For all $G \in \mathcal{G}$ and all $\varepsilon > 0$, the mixing time of the polymer dynamics is $T_{mix}(\varepsilon) = O(|E_G| \log(|E_G|/\varepsilon))$.*

Proof. Let $G \in \mathcal{G}$. We proceed by path coupling. Let $D(\cdot, \cdot)$ be a metric on Ω_G which we define by setting $D(\Gamma, \Gamma') = 1$ for polymer configurations $\Gamma, \Gamma' \in \Omega_G$ such that $\Gamma' = \Gamma \cup \{\gamma\}$ for some polymer $\gamma \in \mathcal{C}_G$. This can be extended to a shortest path metric on Ω_G (the symmetric difference).

Let (X_t, Y_t) be a coupling for the polymer dynamics where both chains make the same random choices. Suppose that $D(X_t, Y_t) = 1$ and that $X_t = Y_t \cup \{\gamma\}$ for some $\gamma \in \mathcal{C}_G$. We obtain (X_{t+1}, Y_{t+1}) from (X_t, Y_t) by performing an update according to the polymer dynamics, and consider the various possibilities for $D(X_{t+1}, Y_{t+1})$. With probability $|E_\gamma|/2|E_G|$ we choose γ and remove it from both X_t and Y_t – in this case $D(X_{t+1}, Y_{t+1}) = 0$. Alternatively, if we add a polymer γ' such that $\gamma' \nsim \gamma$, then the update will be rejected in one chain and performed in the other and $D(X_{t+1}, Y_{t+1}) \leq 2$ – this occurs with probability $|E_{\gamma'}|/2|E_G| \cdot w_G(\gamma')$. Combining these two cases, we obtain that

$$\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq 1 + \frac{1}{2|E_G|}\left(-|E_\gamma| + \sum_{\gamma' \nsim \gamma} |E_{\gamma'}| \cdot w_G(\gamma')\right),$$

and therefore by the polymer mixing condition that

$$\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq 1 - \frac{|E_\gamma|(1-\theta)}{2|E_G|} \leq 1 - \frac{(1-\theta)}{2|E_G|},$$

for some constant $\theta \in (0,1)$. Let W be the diameter of Ω_G with respect to the metric $D(\cdot,\cdot)$ and note that $W \leq 2n$. By the path coupling lemma [30, Section 6], it follows that the mixing time is at most $\log(W/\varepsilon)2|E_G|/(1-\theta) = O(|E_G|\log(|E_G|/\varepsilon))$. □

5.4.1 Single polymer sampler

We now show how to use the polymer dynamics as part of an efficient algorithm for constructing an approximate sample from μ_G. The input to this algorithm will be a graph and an accuracy parameter, and we will assume that we are given the graph both as an adjacency matrix and an array of linked adjacency lists, where each adjacency list contains all of the neighbours of a vertex. These data structures will allow us to decide whether two given vertices are adjacent in constant time (in the unit cost model), and will also allow us to access the next entry in any adjacency list in constant time. We note that other choices of data structure will increase the running time of our algorithms by at most a polynomial factor.

Fix an edge $e \in E_G$. We now show how to sample a polymer from ν_e in expected constant time. Suppose that our polymer model satisfies the polymer sampling condition (Definition 5.3.2) with constant $\tau \geq 3\log(8e^3(q-1))$, and let $r = \tau - \log(12e^2(q-1)) > 0$. For $\ell \geq 0$ let $\mathcal{A}_\ell(e) = \{\gamma \in \mathcal{A}(e) : \deg_G(V_\gamma) \leq \ell\}$. We use the following algorithm to to sample a polymer from ν_e.

Algorithm 5: Sampling a polymer from ν_e.

1. Sample $\boldsymbol{\ell}$ from the following distribution: $\Pr(\boldsymbol{\ell} = \ell) = (1-e^{-r})e^{-r\ell}$ for all integers $\ell \geq 1$. This is so that $\Pr(\boldsymbol{\ell} \geq \ell) = e^{-r\ell}$.
2. Enumerate $\mathcal{A}_{\boldsymbol{\ell}}(e)$ and compute $w_G(\gamma)$ for each $\gamma \in \mathcal{A}_{\boldsymbol{\ell}}(e)$.
3. Mutually exclusively, output each $\gamma \in \mathcal{A}_{\boldsymbol{\ell}}(e)$ with probability $w_G(\gamma) \cdot e^{r\deg_G(\gamma)}$. With all remaining probability, output $\emptyset$.

In order to perform the second step of Algorithm 5, we must be able to enumerate all connected vertex subsets $S \subseteq V_G$ with $\deg_G(S) \leq \boldsymbol{\ell}$, that contain an endpoint of e. In order to do this efficiently, we require the following result, which is an adaptation of Lemma 3.4 of [78].

Lemma 5.4.3. *Let $G = (V_G, E_G)$ be a graph, let $v \in V_G$, and let $\ell \geq 1$ be an integer. There is an algorithm which enumerates all connected vertex subsets $S \subseteq V_G$ such that $v \in S$ and $\deg_G(S) \leq \ell$. The running time of this algorithm is $O(\ell^7(2e)^{4\ell})$.*

Proof. Let $C(G, v, \ell)$ be the set of all connected vertex subsets $S \subseteq V_G$ such that $v \in S$ and $\deg_G(S) \leq \ell$. For $\ell < \deg_G(v)$, we simply output $\emptyset$. For $\ell \geq \deg_G(v)$, we construct $C(G, v, \ell)$ recursively. For the base case, observe that $C(G, v, \deg_G(v))$ is $\{v\}$. For $\ell' \geq \deg_G(v)$, in order to construct $C(G, v, \ell' + 1)$ given $C(G, v, \ell')$, we first construct the multiset

$$C''(G, v, \ell' + 1) = \{S \cup \{u\} : S \in C(G, v, \ell'),\ u \in \partial_G S,\ \deg_G(S \cup \{u\}) = \ell' + 1\},$$

then remove the repeat elements from it to obtain $C'(G, v, \ell' + 1)$, and finally set $C(G, v, \ell' + 1) = C'(G, v, \ell' + 1) \cup C(G, v, \ell')$.

To construct $C''(G, v, \ell' + 1)$, we consider each element $S \in C(G, v, \ell')$ and each vertex $u \in \partial_G S$. Iterating through $C(G, v, \ell')$ requires $O(\ell'(2e)^{2\ell'-1})$ time, by Lemma 5.3.3. For each $S \in C(G, v, \ell')$, we can iterate through $\partial_G S$ in $O(\ell')$ time, given that we have constant-time access to the elements of the adjacency list of each vertex in S. The total time required is therefore $O(\ell'^2(2e)^{2\ell'-1})$. We then remove the repeat elements from $C''(G, v, \ell' + 1)$ by pairwise comparison, which takes $O((\ell' + 1)^2 \cdot |C''(G, v, \ell' + 1)|^2) = O(\ell'^6(2e)^{4\ell'})$ time.

To construct $C(G, v, \ell)$, we begin with the base case and then perform the above $\ell - \deg_G(v)$ times. The total running time is therefore $O(\ell^7(2e)^{4\ell})$. We now show that the algorithm returns every element of $C(G, v, \ell)$. For the base case, it is clearly true. Furthermore, for all $\ell' > \deg_G(v)$ and all $S \in C(G, v, \ell')$, we know that S consists of $T \in C(G, v, \ell'')$ and $u \in \partial_G T$, for some $\ell'' < \ell'$. □

We can now show the correctness and expected constant running time of Algorithm 5.

Lemma 5.4.4. *Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\{(\mathcal{C}_G, w_G) \mid G \in \mathcal{G}\}$ be a family of computationally feasible (Definition 5.3.5) q-spin polymer models that satisfies the polymer sampling condition (see Definition 5.3.2) with constant $\tau \geq 3\log(8e^3(q-1))$. For all $G \in \mathcal{G}$ and all $e \in E_G$, Algorithm 5 samples a polymer from ν_e in expected constant time.*

Proof. We begin by showing that step 3 of the algorithm is well-defined, by showing that $\sum_{\gamma \in \mathcal{A}(e)} w_G(\gamma) \cdot e^{r \deg_G(V_\gamma)} \leq 1$ for all $e \in E_G$. For all $v \in V_G$, we have that

$$\sum_{\gamma : u \in V_\gamma} w_G(\gamma) \cdot e^{r \deg_G(V_\gamma)} \leq \sum_{\ell \geq 1} \sum_{\substack{\gamma : u \in V_\gamma, \\ \deg_G(V_\gamma) = \ell}} w_G(\gamma) \cdot e^{r\ell} \leq \sum_{\ell \geq 1} (4e^2)^\ell (q-1)^\ell e^{-\tau \ell} e^{r\ell} \leq \frac{1}{2},$$

where the second-to-last inequality follows from Lemma 5.3.3 and the final inequality follows from the fact that $4e^2(q-1)e^{-\tau+r} \leq 1/3$.

We now show that the expected running time of the algorithm is constant. For $\ell \geq 1$, the time taken to enumerate $\mathcal{A}_\ell(e)$ is $O((q-1)^\ell \ell^7 (2e)^{4\ell})$, by Lemma 5.4.3. The time taken to then iterate through $\mathcal{A}_\ell(e)$ and compute $w_G(\gamma)$ for each $\gamma \in \mathcal{A}_\ell(e)$ is $O((q-1)^\ell \ell (2e)^{2\ell-1} e^\ell)$, by Lemma 5.3.3 and the fact that the family of polymer models is computationally feasible. The expected running time of Algorithm 5 is therefore

$$\begin{aligned} &= O\left(\sum_{\ell \geq 1} \Pr(\boldsymbol{\ell} = \ell) \left((q-1)^\ell \ell^7 (2e)^{4\ell} + (q-1)^\ell \ell (2e)^{2\ell-1} e^\ell \right) \right) \\ &= O\left(\sum_{\ell \geq 1} \ell^7 \left(\frac{16e^4(q-1)}{e^r} \right)^\ell \right). \end{aligned}$$

Since $r = \tau - \log(12e^2(q-1)) \geq 3\log(8e^3(q-1)) - \log(12e^2(q-1)) \geq 2\log(8e^3(q-1))$, it follows that $16e^4(q-1)/e^r < 1/2$, and therefore that the expected running time of Algorithm 5 is $O(1)$.

Finally, let $\gamma \in \mathcal{A}(e)$ be a polymer. In order for the algorithm to sample γ, it must first sample $\boldsymbol{\ell} \geq \deg_G(V_\gamma)$ in step 1 of the algorithm, then conditioned on this choice, it must output γ in step 3. This occurs with probability $e^{-r \deg_G(V_\gamma)} w_G(\gamma) \cdot e^{r \deg_G(V_\gamma)} = w_G(\gamma)$, and therefore the output distribution is ν_e. □

We now combine the polymer dynamics with Algorithm 5 to give an efficient algorithm for computing an approximate sample from the Gibbs distribution of a polymer model.

Lemma 5.2.1. *Let $q \geq 2$ be an integer, let $\mathcal{G}$ be a class of graphs, and let $\mathcal{F}_{\mathcal{G}}$ be a family of computationally feasible (Definition 5.3.5) q-spin polymer models satisfying the polymer sampling condition (Defintion 5.3.2) with constant $\tau \geq 3\log(8e^3(q-1))$.*

There are randomised algorithms which, given as input a graph $G \in \mathcal{G}$ with m edges and an accuracy parameter $\varepsilon > 0$, output an ε-sample from μ_G in $O\big(m\log(m/\varepsilon)\log(1/\varepsilon)\big)$ time, and an ε-approximation to Z_G, with probability at least $3/4$, in $O\big((m/\varepsilon)^2\log^3(m/\varepsilon)\big)$ time.

Proof. We focus on the sampling algorithm, since it is shown in [22, Section 3] how to convert this into the desired counting algorithm for the polymer model (although the approach in [22] was stated for bounded-degree graphs, it carries over to graphs of unbounded degree).

By Lemma 5.4.1, the unique stationary distribution of the polymer dynamics is μ_G, thus our sampling algorithm is based on the polymer dynamics. By Lemma 5.4.4, there is an integer $C_1 > 1$ (independent of G) such that the expected number of steps taken to execute a single update of the polymer dynamics is at most C_1. Lemma 5.3.4 shows that $\mathcal{F}_{\mathcal{G}}$ also satisfies the polymer mixing condition. This allows us to apply Lemma 5.4.2, which shows that there is an integer $C_2 > 1$ (also independent of G) such that the mixing time of the polymer dynamics satisfies $T_{\text{mix}}(\varepsilon/2) \leq C_2 m \log \frac{m}{\varepsilon}$.

We use the following algorithm to compute an ε-sample from μ_G. Repeat the following $\lceil \log(2/\varepsilon) \rceil$ times, and if no polymer configuration is returned, return $\emptyset$. Run the polymer dynamics for $3C_1C_2m\lceil \log \frac{m}{\varepsilon} \rceil$ steps, starting from $\emptyset$, and if at least $C_2 m \lceil \log \frac{m}{\varepsilon} \rceil$ updates of the polymer dynamics were executed, then return the configuration.

We claim that the distribution of the output configuration from the above algorithm is within total variation distance ε of μ_G. This will follow once we have shown that the probability that no run of the polymer dynamics returns a configuration, is at most $\varepsilon/2$. This is because the configuration we would output

if one is returned by a run of the polymer dynamics, is within total variation distance $\varepsilon/2$ of μ_G.

Consider one of the $\lceil \log(2/\varepsilon) \rceil$ independent runs of the polymer dynamics that is made by the algorithm. Let the random variable X denote the total number of steps required to execute $C_2 m \lceil \log \frac{m}{\varepsilon} \rceil$ updates of the polymer dynamics. We have that $\mathbb{E}[X] \leq C_1 C_2 m \lceil \log \frac{m}{\varepsilon} \rceil$, and therefore by Markov's inequality that $\Pr(X \geq 3\,\mathbb{E}[X]) \leq 1/3 < 1/e$. Since the runs are independent, the probability that no run performs at least $C_2 m \lceil \log \frac{m}{\varepsilon} \rceil$ updates of the polymer dynamics, is at most $(1/e)^{\log(2/\varepsilon)} = \varepsilon/2$. □

5.5 Application to unbounded-degree graphs

Whilst the algorithms presented in Chapter 3 applied to α-expander graphs (as per Definition 2.5.2), here we will require a slightly modified notion of expansion.

Definition 5.5.1. Let $\alpha > 0$ be a real number. We say that a graph $G = (V_G, E_G)$ is an α-degree expander if, for all connected vertex subsets $S \subseteq V_G$ with $|S| \leq |V_G|/2$, we have that $e_G(S, S^c) \geq \alpha \deg_G(S)$. We let $\mathcal{G}_\alpha$ denote the set of all α-degree expanders.[1]

Let $q \geq 2$ be an integer, let $\beta > 0$ be a real number, and fix a ground state colour from $[q]$ – we'll call this red. We will consider the ferromagnetic Potts model defined over the class $\mathcal{G}_\alpha$. The following definition captures the red phase of the ferromagnetic Potts model.

Definition 5.5.2. Let $G = (V_G, E_G)$ be a graph. Let q be a positive integer, and let $\beta > 0$ be a real number. We define $\Omega^{\text{red}}_{G,q}$ to be set of q-colourings of G that colour more than half of the vertices of G with the colour red. We define $Z^{\text{red}}_{G,q,\beta} = \sum_{\sigma \in \Omega^{\text{red}}_{G,q}} e^{-\beta b_G(\sigma)}$ and for $\sigma \in \Omega^{\text{red}}_{G,q}$, we define $\mu^{\text{red}}_{G,q,\beta}(\sigma) = e^{-\beta b_G(\sigma)}/Z^{\text{red}}_{G,q,\beta}$.

Consider the polymer model for the ferromagnetic Potts model introduced in Section 3.5.1, defined over the class of α-degree expanders $\mathcal{G}_\alpha$. Note that our new framework applies to this polymer model, since its host graph is the same as the underlying graph G. Let $\hat{\Omega}^{\text{red}}_{G,q}$ denote the set of all sets of mutually compatible

[1]Note that every connected $G \in \mathcal{G}_\alpha$ is also an α-expander as per Definition 2.5.2 (i.e., $e_G(S, S^c) \geq \alpha|S|$).

allowed polymers, let $\hat{\mu}^{\mathrm{red}}_{G,q,\beta}$ denote the Gibbs distribution of the polymer model, and let $\hat{Z}^{\mathrm{red}}_{G,q,\beta}$ denote its partition function. Observe that there is a bijection between the polymer configurations of $\hat{\Omega}^{\mathrm{red}}_{G,q}$ and the Potts configurations of $\Omega^{\mathrm{red}}_{G,q}$, where a colouring $\sigma \in \Omega^{\mathrm{red}}_{G,q}$ maps to the polymer configuration $\Gamma \in \hat{\Omega}^{\mathrm{red}}_{G,q}$ consisting of the connected components of vertices that do not get coloured red by σ. Moreover, the weight of each $\Gamma \in \hat{\Omega}^{\mathrm{red}}_{G,q}$ is closely related to the weight of the Potts configuration $\sigma \in \Omega^{\mathrm{red}}_{G,q}$ by

$$\prod_{\gamma \in \Gamma} w_G(\gamma) = e^{-\beta b_G(\sigma)}.$$

Sampling from this polymer model is equivalent to sampling from the Potts model restricted to colourings which colour more than half of the vertices with red, and we will formalise this idea later. We denote this family of polymer models by $\mathcal{F}^{\mathrm{red}}_{\mathcal{G}_\alpha} = \{(\mathcal{C}^{\mathrm{red}}_G, w^{\mathrm{red}}_G) \mid G \in \mathcal{G}_\alpha\}$.

We will prove the following facts about the above polymer model.

1. The partition function of the polymer model $\hat{Z}^{\mathrm{red}}_{G,q,\beta}$ is a close approximation to the partition function of the red phase of the Potts model $Z^{\mathrm{red}}_{G,q,\beta}$, when β is sufficiently large with respect to q and α.

2. The polymer model is computationally feasible and satisfies the polymer sampling condition with constant $\tau \geq 3\log(8e^3(q-1))$, when β is sufficiently large with respect to q and α.

Item 1 tells us that in order to approximate the partition function of the low temperature Potts model on $\mathcal{G}_\alpha$, it suffices to approximate the partition function of the above polymer model. By Item 2, we also have an efficient algorithm to do this with (that of Lemma 5.2.1).

We first deal with Item 1, by appealing to the following result of Jenssen, Keevash, and Perkins [68]. We note that although this result is applied to bounded-degree graphs in [68], it remains true for arbitrary α-expanders (and therefore also arbitrary α-degree expanders).

Lemma 5.5.3. *[68, Lemma 12] Let $\alpha > 0$ and let $G = (V_G, E_G)$ be an α-expander. Let $q \geq 2$ be an integer, fix an arbitrary spin from $[q]$ (by symmetry, we can assume that this is red), and let $\beta > 2\log(eq)/\alpha$ be a real number. We have that $qZ^{red}_{G,q,\beta}$ is an $e^{-|V_G|}$-approximation to $Z_{G,q,\beta}$.*

We now show that the algorithms presented in Section 5.4 can be used to sample from the Gibbs distribution of the polymer model, and approximate its partition function.

Lemma 5.5.4. *Let $\alpha > 0$ be a real number. Let $q \geq 3$ be an integer and let $\beta \geq \frac{3}{\alpha}\log(8e^3(q-1))$ be a real number. Fix an arbitrary spin from $[q]$ (by symmetry, we can assume that this is red). The family of polymer models $\mathcal{F}^{red}_{\mathcal{G}_\alpha}$ is computationally feasible (Definition 5.3.5) and satisfies the polymer sampling condition (Definition 5.3.2) with constant $\tau = \alpha\beta \geq 3\log(8e^3(q-1))$.*

Proof. Fix $G \in \mathcal{G}_\alpha$. To see that the polymer sampling condition is satisfied, consider $\gamma \in \mathcal{C}^r_G$ and observe that since $B_\gamma \geq e_G(V_\gamma, V_\gamma^c)$ and $|V_\gamma| < |V_G|/2$, it follows that

$$w_{G,\beta}(\gamma) \leq \exp\{-\alpha\beta \deg_G(V_\gamma)\} = e^{-\tau \deg_G(V_\gamma)},$$

where $\tau \geq 3\log(8e^3(q-1))$ if $\beta \geq \frac{3}{\alpha}\log(8e^3(q-1))$.

The polymer model is computationally feasible, since determining whether $\gamma \in \mathcal{C}^r_{G,q}$ can be done in $O(|V_\gamma|) = O(\exp(\deg_G(V_\gamma)))$ time (one just needs to check whether $|V_\gamma| < n/2$). Computing $w_{G,\beta}(\gamma) = e^{-\beta B_\gamma}$ can be done by examining all $O(\deg_G(V_\gamma))$ edges with endpoints in V_γ, by iterating through V_γ and counting the neighbours of each vertex that are not in V_γ. The total running time required to do this is therefore $O(|V_\gamma|^2 \deg_H(V_\gamma)) = O(\exp(\deg_H(V_\gamma)))$. □

We are now in a position to prove the following theorem, which gives an efficient algorithm for sampling from the low temperature Potts model on expanders.

Theorem 5.5.5. *Let $\alpha > 0$ be a real number. Let $q \geq 3$ be an integer and let $\beta \geq \frac{3}{\alpha}\log(8e^3(q-1))$ be a real. For the Potts model on $G \in \mathcal{G}_\alpha$, there is a poly-time approximate sampling algorithm for $\mu_{G,q,\beta}$ and an* FPRAS *for $Z_{G,q,\beta}$.*

In fact, for $n = |V_G|$ and $m = |E_G|$, if the desired accuracy ε satisfies $\varepsilon \geq e^{-n}$ then the running time of the sampler is $O(m\log(m/\varepsilon)\log(1/\varepsilon))$ and the running time of the FPRAS *is $O\big((m/\varepsilon)^2\log^3(m/\varepsilon)\big)$.*

Proof. Assume without loss of generality that G is connected (otherwise, consider the connected components separately). Fix an arbitrary spin from $[q]$ (by symmetry, we can assume that this is red) and consider the family of polymer

models $\mathcal{F}^{\mathrm{red}}_{\mathcal{G}_\alpha}$. From Lemma 5.5.4, we know that this family of polymer models is computationally feasible and satisfies the polymer sampling condition (Definition 5.3.2) with constant $\tau = \alpha\beta \geq 3\log(8e^3(q-1))$. We may therefore apply the algorithms of Lemma 5.2.1.

The input to both algorithms is a graph $G \in \mathcal{G}_\alpha$ and an accuracy parameter $\varepsilon > 0$. If $\varepsilon < e^{-n}$ then we can construct an ε-sample from $\mu_{G,q,\beta}$ in $\mathrm{poly}(n, 1/\varepsilon)$ time, by brute force. If $\varepsilon \geq e^{-n}$ then we can construct an ε-sample from $\mu_{G,q,\beta}$ as follows. Choose a colour from $[q]$ (by symmetry we can assume that this is red) uniformly at random. By Lemma 5.2.1 there is an algorithm which, given ε/q and G as input, outputs an (ε/q)-sample from $\hat{\mu}^{\mathrm{red}}_{G,q,\beta}$ (and therefore also from $\mu^{\mathrm{red}}_{G,q,\beta}$) in $O\big(m\log(m/\varepsilon)\log(1/\varepsilon)\big)$ time. Since G is also an α-expander, then by Lemma 5.5.3, this is an ε-sample from $\mu_{G,q,\beta}$. For the counting algorithm, if $\varepsilon < e^{-n}$ then we can compute $Z_{G,q,\beta}$ exactly in $\mathrm{poly}(n, 1/\varepsilon)$ time, by brute force. If $\varepsilon \geq e^{-n}$, then by Lemma 5.2.1, we can compute an ε-approximation to $Z^{\mathrm{red}}_{G,q,\beta}$ in $O\big(m^2\log^3(m/\varepsilon)\big)$ time. By Lemma 5.5.3, it follows that $qe^{\beta m} \cdot Z^{\mathrm{red}}_{G,q,\beta}$ is an ε-approximation to $Z_{G,q,\beta}$. □

5.6 Random graphs with a given degree sequence

Let d be a real number. In this section, we will show that a graph $G \sim \mathcal{G}(n, \vec{x})$ for a degree sequence $\vec{x} \in \mathcal{D}_{n,d}$ is whp an α-degree-expander for some constant $\alpha > 0$, i.e., that $G \in \mathcal{G}_\alpha$. This is so that we can apply the results of the previous section (specifically Theorem 5.5.5) in order to provide efficient algorithms for sampling from the low temperature Potts model for random graphs with a given degree sequence.

As explained in Section 2.5, we typically prove properties of random graphs by first proving them for random multigraphs generated by the configuration model. These properties can then be transferred back to the random graph model using Lemma 2.5.1. See Section 2.5 for more details. For a (multi)graph $H = (V_H, E_H)$ we define the tree-excess to be $t_H = |E_H| - (|V_H| - 1)$; that is, the number of edges more than a tree that H has. First, we show that multigraphs drawn from the configuration model have locally bounded tree excess.

Lemma 5.6.1. *Let d be a positive real number. The following is true with high probability when $H = (V_H, E_H)$ is drawn from $\mathsf{CM}(n, \vec{x})$ uniformly over all degree sequences $\vec{x} \in \mathcal{D}_{n,d}$. For all connected vertex sets $S \subseteq V_H$ with $|S| \leq (\log n)^2$ and $\deg_H(S) \geq 36$, we have that $t_{H[S]} \leq \frac{1}{6}\deg_H(S)$.*

Proof. For positive integers k and ℓ, and a non-negative integer t, let the random variable $X_{k,\ell,t}$ denote the number of connected vertex subsets $S \subseteq V_H$ such that $|S| = k$, $\deg_H(S) = \ell$, and $t_{H[S]} = t$. To prove the lemma, we will show that whp

$$\sum_{k \leq \lfloor (\log n)^2 \rfloor} \sum_{\ell \geq 36} \sum_{t \geq \lfloor \ell/6 \rfloor + 1} X_{k,\ell,t} = 0.$$

In fact, we can further restrict the range of summation. From the lower bound in Item 1 of Definition 5.2.2, we have that $x_i \geq 3$ for all i, and therefore $\ell \geq 3k$. Item 2 shows that $\sum_i x_i \leq dn$, and therefore $\ell \leq dn$ and $t \leq \ell/2 \leq dn/2$. So, consider any integer ℓ in the range $36 \leq \ell \leq dn$, any integer k in the range $1 \leq k \leq \min\{(\log n)^2, \ell/3\}$, and any integer $t > \ell/6$. There are at most $\binom{n}{k}$ vertex subsets $S \subseteq V_H$ with $|S| = k$ and $\deg_G(S) = \ell$. Let $j = k - 1 + t$ be the number of edges with both endpoints in S. Given such a set S, there are at most $\binom{\ell}{2j}$ possibilities for the set of half-edges in these j edges. On a given set of $2j$ half-edges, there are $(2j-1)!! = \frac{(2j)!}{2^j j!}$ perfect matchings. Using the upper bound on the degrees from Item 1 of Definition 5.2.2, the probability that a set of j edges is present in H is at most

$$\frac{n^{2\rho}}{2m-1}\frac{n^{2\rho}}{2m-3}\cdots\frac{n^{2\rho}}{2m-2j+1} \leq \left(\frac{n^{2\rho}}{2m-2j}\right)^j \leq \left(\frac{n^{2\rho}}{n}\right)^j.$$

To see this, consider the probability that two half-edges emanating from vertices in the set (of which there are at most n^ρ per vertex, by the assumed degree bound) are paired j times. Note also that the final inequality follows from the fact that $k \leq (\log n)^2$ and therefore that $2m - 2j \geq \deg_G(S^c) \geq 3|S^c| = 3(n-k) > n$ (as long as n is sufficiently big). We also have that

$$\binom{\ell}{2j} \cdot \frac{(2j)!}{2^j j!} < \frac{\ell!}{(\ell-2j)!j!} < \frac{\ell^{2j}e^j}{j^j} \leq \left(\frac{e\ell^2}{t}\right)^j.$$

Putting everything together, it follows that

$$\mathbb{E}[X_{k,\ell,t}] \leq \binom{n}{k}\left(\frac{e\ell^2}{t}\right)^{k-1+t}\left(\frac{n^{2\rho}}{n}\right)^{k-1+t} < \left(\frac{e^2\ell^2}{t}\right)^{k-1+t}\frac{n^{2\rho(k-1+t)}}{n^{t-1}}.$$

Furthermore, since $t > \ell/6$, $k < 2t$, and (by the upper bound in Item 1 of Definition 5.2.2) $\ell \leq kn^{\rho} \leq n^{2\rho}$, we have that

$$\mathbb{E}[X_{k,\ell,t}] < \frac{(6e^2 n^{4\rho})^{3t-1}}{n^{t-1}}. \tag{5.1}$$

Let

$$X = \sum_{\ell=36}^{dn} \sum_{k=1}^{\lfloor \min\{(\log n)^2, \ell/3\} \rfloor} \sum_{t=\lfloor \ell/6 \rfloor + 1}^{\lfloor dn/2 \rfloor} X_{k,\ell,t}.$$

Since $t > \ell/6 \geq 6$, it follows that $t \geq 7$. For big enough n, (5.1) shows that $\mathbb{E}[X_{k,\ell,t}] \leq n^{13\rho t}/n^{t-1}$. Recall from Definition 5.2.2 that $\rho = 1/50$. Since $\rho \leq 2/91$ and $t \geq 7$, $1 - 13\rho \geq 5/7 \geq 5/t$ so $13\rho t \leq t - 5$ and $\mathbb{E}[X_{k,\ell,t}]$ is at most n^{-4}. Taking a union bound over all permissible values for ℓ, k, and t, we find that $\mathbb{E}[X] = o(1)$. Applying Markov's inequality, we have that $\Pr(X > 0) = \Pr(X \geq 1) \leq \mathbb{E}[X] = o(1)$, and the result follows. □

We will split the remaining cases in two depending on the size of the vertex set (see Lemmas 5.6.3 and 5.6.4 which prove expansion properties for 'large' and 'small' vertex subsets, respectively). To obtain these expansion bounds, we will use the following result from Fountoulakis and Reed [34]. Although this result is stated in [34] in terms of the random graph model, it is first proved for the configuration model, so this is how we state it. Also, the result in [34] requires that the vector $\vec{x}$ be in $\mathcal{D}_{n,d}$ but this is only important for lifting their result to the random graph model, so it is not relevant for us.

Lemma 5.6.2 ([34, Proposition 4.5])**.** *When $H = (V_H, E_H)$ is drawn from $\mathsf{CM}(n, \vec{x})$ for some length-n degree sequence $\vec{x}$ the following is true for any set $S \subseteq V_H$.*

$$\Pr(e_H(S, S^c) = 0) \leq \binom{m}{\deg_H(S)/2}^{-1},$$

where $m = \frac{1}{2}\sum_i x_i$.

Note that Lemma 5.6.2 was stated in [34] for S such that $\deg_H(S)$ is even, but if $\deg_H(S)$ is odd, it is not possible to have $e_H(S, S^c) = 0$. Next, we show that in a multigraph H drawn from the configuration model, small vertex subsets satisfy certain expansion properties.

Lemma 5.6.3. *Let d be a positive real number. The following is true with high probability when $H = (V_H, E_H)$ is drawn from $\mathsf{CM}(n, \vec{x})$ uniformly over all degree sequences $\vec{x} \in \mathcal{D}_{n,d}$. For all connected vertex sets $S \subseteq V_H$ with $|S| \leq (\log n)^2$, we have that $e_H(S, S^c) \geq |S|/4$.*

Proof. For positive integers k and ℓ, and a non-negative integer j, let the random variable $X_{k,j,\ell}$ denote the number of connected vertex subsets $S \subseteq V_H$ with $|S| = k$, $e_H(S, S^c) = j$, and $\deg_H(S) = \ell$. By Item 1 of Definition 5.2.2, we need only consider ℓ satisfying $3k \leq \ell \leq kn^\rho$. Let

$$X = \sum_{k=1}^{\lfloor (\log n)^2 \rfloor} \sum_{j=0}^{\lfloor k/4 \rfloor} \sum_{\ell=3k}^{\lfloor kn^\rho \rfloor} X_{k,j,\ell}.$$

To prove the lemma we will show that $X = 0$, whp. Consider any integer k in the range $1 \leq k \leq (\log n)^2$, any integer j in the range $0 \leq j < k/4$, and any integer ℓ in the range $3k \leq \ell \leq kn^\rho$. There are at most $\binom{n}{k}$ candidates for vertex sets S with $|S| = k$ and $\deg_H(S) = \ell$. There are then at most $\binom{\ell}{j}$ choices for the j half-edges emanating from vertices of S that will be matched with half-edges emanating from vertices of S^c, once H is drawn. Applying Lemma 5.6.2 to the degree sequence derived from $\vec{x}$ by removing the j half-edges (and their partners), the probability that the remaining $\ell - j$ half-edges are matched amongst themselves is at most

$$\binom{m'}{(\ell - j)/2}^{-1} \leq \left(\frac{(\ell - j)}{2m'}\right)^{\frac{(\ell-j)}{2}} \leq \left(\frac{kn^\rho}{n}\right)^{\frac{11k}{8}},$$

where $2m' = (\sum_{i=1}^{n} x_i) - 2j$ and the last inequality follows (for big enough n) since $11k/4 \leq \ell - j \leq kn^\rho$ and $2m' \geq 3n - 2j > n$. We therefore have that

$$\begin{aligned}
\mathbb{E}[X] &\leq \sum_{k=1}^{\lfloor (\log n)^2 \rfloor} \sum_{j=0}^{\lfloor k/4 \rfloor} \sum_{\ell=3k}^{\lfloor kn^\rho \rfloor} \binom{n}{k} \binom{\ell}{j} \left(\frac{kn^\rho}{n}\right)^{\frac{11k}{8}} \\
&\leq \sum_{k=1}^{\lfloor (\log n)^2 \rfloor} \sum_{j=0}^{\lfloor k/4 \rfloor} \sum_{\ell=3k}^{\lfloor kn^\rho \rfloor} \left(\frac{ne}{k}\right)^k \left(\frac{e\ell}{j}\right)^j \left(\frac{kn^\rho}{n}\right)^{\frac{11k}{8}} \\
&\leq \sum_{k=1}^{\lfloor (\log n)^2 \rfloor} \left(\frac{(\log n)^{O(1)} n^{\rho(2+11/8)}}{n^{3/8}}\right)^k.
\end{aligned}$$

This is $o(1)$ since $\rho < 1/9$. Applying Markov's inequality, we have that $\Pr(X > 0) = \Pr(X \geq 1) = o(1)$, and the result follows. □

We will also require the following lemma.

Lemma 5.6.4. *Let d be a positive real number. There is a positive real number α (depending on d) such that the following is true with high probability when $H = (V_H, E_H)$ is drawn from $\mathsf{CM}(n, \vec{x})$ uniformly over all degree sequences $\vec{x} \in \mathcal{D}_{n,d}$. For all connected vertex sets $S \subseteq V_H$ with $(\log n)^2 \leq |S| \leq n/2$, we have that $e_H(S, S^c) \geq \alpha \deg_H(S)$.*

Proof. We split the proof into three cases based on the degree of the vertex set $S \subseteq V_H$:

1. $\deg_H(S) \leq 100d|S|$,
2. $\deg_H(S) > 100d|S|$ and $\deg_H(S) \leq n/2$, and
3. $\deg_H(S) > 100d|S|$ and $\deg_H(S) > n/2$.

Throughout, we let $m = \frac{1}{2}\sum_{i=1}^{n} x_i$.

Case 1. The proof of this case is similar to that of Lemma 4.1 of [34], but we include it for completeness. We will show that there is a positive real number $\alpha' < 1/4$ such that whp, every vertex set $S \subseteq V_H$ with $(\log n)^2 \leq |S| \leq n/2$ and $\deg_H(S) \leq 100d|S|$ satisfies $e_H(S, S^c) \geq \alpha'|S| \geq \alpha' \deg_H(S)/(100d)$. We will eventually require α' to be sufficiently small depending on d.

For positive integers k and ℓ, and a non-negative integer j, let the random variable $X_{k,\ell,j}$ denote the number of connected vertex subsets $S \subseteq V_H$ with $|S| = k$, $e_H(S, S^c) = j$, and $\deg_H(S) = \ell$. Let

$$X = \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j=0}^{\lfloor \alpha' k \rfloor} \sum_{\ell=1}^{100dk} X_{k,\ell,j}.$$

Our aim will be to show that $\mathbb{E}[X] = o(1)$, and we begin by further restricting the range of summation in the above. Let $S \subseteq V_H$ be a connected vertex subset with $|S| = k$, $e_H(S, S^c) = j$, and $\deg_H(S) = \ell$. Since there are $(\ell - j)/2$ edges in $e_H(S, S)$, we need only consider ℓ such that $\ell - j$ is even. Furthermore, since S is connected, we can assume that $e_H(S, S) \geq k - 1$, so we need only consider values of ℓ satisfying $\ell \geq j + 2(k-1)$. Finally, since $x_i \geq 3$ for all i, and $\sum_i x_i = 2m$, we have that $2m \geq \ell + 3(n-k)$, so we need only consider values of ℓ satisfying $\ell \leq 2m - 3(n-k)$. Let $J_k = \{j \in \mathbb{Z} : 0 \leq j \leq \alpha' k\}$ and let

$L_{k,j} = \{\ell \in \mathbb{Z} : j + 2(k-1) \le \ell \le \min\{2m - 3(n-k), 100dk\},\ \ell - j \text{ is even}\}$, we can therefore re-write X as follows

$$X = \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j \in J_k} \sum_{\ell \in L_{k,j}} X_{k,\ell,j}.$$

Given a set of k vertices, there are at most $\binom{\ell}{j}$ ways to choose the half-edges that will make up $E_H(S, S^c)$ once H is drawn. The remaining total degree of S after choosing these edges is $\ell - j$. By Lemma 5.6.2, the probability that every other edge of of $E_H(S, V_H)$ is an edge of $E_H(S, S)$ is at most $\binom{m'}{(\ell-j)/2}^{-1}$, where $m' = m - j$. Since $\ell \in L_{k,j}$, $(\ell - j)/2 \ge k - 1$ and $(\ell - j)/2 \le m - 3(n-k)/2 - j/2$. Since $j < \alpha' k < k/4 \le n/8$ it follows that $n - k > j$ and therefore that $(\ell - j)/2 \le m - 3(n-k)/2 - j/2 \le m' - n + k$. Thus, we have that

$$\binom{m'}{(\ell - j)/2}^{-1} \le \max\left\{ \binom{m'}{k-1}^{-1}, \binom{m'}{n-k}^{-1} \right\}.$$

Using Stirling's approximation, there exists a constant C (see [34, Equation 7]) such that for all n, m', and k

$$\frac{\binom{n}{k}}{\binom{m'}{k-1}} \le C \cdot \frac{m'}{k} \left(\frac{n}{m'}\right)^k \quad \text{and} \quad \frac{\binom{n}{k}}{\binom{m'}{n-k}} \le C \cdot \frac{m'}{m' - n + k + 1} \left(\frac{n}{m'}\right)^{n-k}.$$

Since $2m' = 2m - 2j \ge 3n - n/4$, it follows that $n/m' \le 8/11$. Since $k \le n/2$, it follows that $n - k \ge k$. Thus,

$$\frac{\binom{n}{k}}{\binom{m'}{k-1}} \le C \cdot m' \left(\frac{8}{11}\right)^k \quad \text{and} \quad \frac{\binom{n}{k}}{\binom{m'}{n-k}} \le C \cdot m' \left(\frac{8}{11}\right)^k,$$

therefore

$$\binom{n}{k} \binom{m'}{(\ell - j)/2}^{-1} \le C \cdot m' \left(\frac{8}{11}\right)^k.$$

From the above, we have that

$$\begin{aligned} \mathbb{E}[X] &= O(1) \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j \in J_k} \sum_{\ell \in L_{k,j}} \binom{n}{k} \binom{\ell}{j} \binom{m'}{(\ell - j)/2}^{-1} \\ &= O(1) \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j \in J_k} \sum_{\ell \in L_{k,j}} m' \binom{\ell}{j} \left(\frac{8}{11}\right)^k. \end{aligned}$$

We wish to show that for each term $\binom{\ell}{j} \leq (9/8)^k$. Since $j \leq \alpha' k \leq \alpha' \ell < \ell/4$ we have $\binom{\ell}{j} \leq \binom{\ell}{\alpha' k} \leq (\frac{e\ell}{\alpha' k})^{\alpha' k}$ and since $\ell \leq 100dk$ this is at most $(100ed/\alpha')^{\alpha' k}$ which is at most $(9/8)^k$ if α' is sufficiently small, depending on d. (It suffices to take $\alpha' = 10^{-4}/\log d$, for example.)

We conclude that we can upper bound each term by $m'(9/11)^k \leq dn(9/11)^k$ and therefore

$$\mathbb{E}[X] = O(n) \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j \in J_k} \sum_{\ell \in L_{k,j}} \left(\frac{9}{11}\right)^k = O(n) \sum_{k=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} k^2 \left(\frac{9}{11}\right)^k.$$

It follows from the above that $\mathbb{E}[X] = o(1)$. Applying Markov's inequality, we have that $\Pr(X > 0) = \Pr(X \geq 1) = o(1)$, and the result follows.

Case 2. We will do more than is required, and show that every connected vertex set $S \subseteq V_H$ with $(\log n)^2 \leq \deg_H(S) \leq n/2$ satisfies $e_H(S, S^c) \geq \deg_H(S)/100$.

For every positive integer ℓ satisfying $(\log n)^2 \leq \ell \leq n/2$ and every non-negative integer j satisfying $j \leq \ell/100$, let $X_{\ell,j}$ denote the number of connected vertex subsets $S \subseteq V_H$ with $e_H(S, S^c) = j$ and $\deg_H(S) = \ell$. Let $J_\ell = \{j \in \mathbb{Z} : 0 \leq j \leq \ell/100, \ell - j \text{ is even}\}$ and let

$$X = \sum_{\ell=\lceil (\log n)^2 \rceil}^{\lfloor n/2 \rfloor} \sum_{j \in J_\ell} X_{\ell,j}.$$

Our aim will be to show that $\mathbb{E}[X] = o(1)$.

In order for a set $S \subseteq V_H$ to contribute to $X_{\ell,j}$ its size, s, must be at most $(\ell - j)/2 + 1$ (otherwise, the $(\ell - j)/2$ internal edges won't be able to connect S). The number of size-s sets S is at most $\binom{n}{s}$ which is at most $\binom{n}{(\ell-j)/2+1}$, as long as n is sufficiently large. The number of possible values of s is at most $(\ell - j)/2 \leq \ell$. Thus, the number of possibilities for S is at most $\ell \cdot \binom{n}{(\ell-j)/2+1}$

Given S, there are at most $\binom{\ell}{j}$ ways to choose the half-edges that will make up $E_H(S, S^c)$ once H is drawn. The remaining total degree of S after choosing these edges is $\ell - j$. By Lemma 5.6.2, the probability that every other edge of of $E_H(S, V_H)$ is an edge of $E_H(S, S)$ is at most $\binom{m'}{(\ell-j)/2}^{-1}$, where $m' = m - j$. As noted in Case 1, it is shown using Stirling's approximation in [34, Equation 7] that there exists a constant C such that for all n, ℓ, j and m',

$$\binom{n}{(\ell - j)/2 + 1} \binom{m'}{(\ell - j)/2}^{-1} \leq C m' \left(\frac{n}{m'}\right)^{\frac{\ell - j}{2} + 1}.$$

Since $2m \geq 3n$ by Item 1 of Definition 5.2.2 and $2j \leq 2\ell/100 \leq n/100$, $2m' = 2m - 2j \geq 3n - n/100$ so $n/m' < 7/10$. Thus,

$$\begin{aligned}\binom{n}{(\ell-j)/2+1}\binom{m'}{(\ell-j)/2}^{-1} &\leq Cm'\left(\frac{7}{10}\right)^{\frac{\ell-j}{2}+1} \\ &\leq Cm'\left(\frac{7}{10}\right)^{\frac{\ell-\ell/100}{2}} \\ &\leq Cm'\left(\frac{17}{20}\right)^{\ell}.\end{aligned}$$

Furthermore, we have that

$$\binom{\ell}{j} \leq \binom{\ell}{\ell/100} \leq (100e)^{\ell/100}.$$

Putting things together, it follows that

$$\begin{aligned}\mathbb{E}[X] &\leq \sum_{\ell=\lceil(\log n)^2\rceil}^{\lfloor n/2\rfloor} \sum_{j\in J_\ell} \ell \cdot \binom{\ell}{j}\binom{n}{(\ell-j)/2+1}\binom{m'}{(\ell-j)/2}^{-1} \\ &\leq C \sum_{\ell=\lceil(\log n)^2\rceil}^{\lfloor n/2\rfloor} \ell^2 m' \left(\frac{17}{20}\right)^{\ell} (100e)^{\ell/100}.\end{aligned}$$

Now $m' \leq m$ and this is $O(n)$ by Item 2 of Definition 5.2.2, so

$$\mathbb{E}[X] = O(n) \sum_{\ell=\lceil(\log n)^2\rceil}^{\lfloor n/2\rfloor} \ell^2 \cdot \left(\frac{9}{10}\right)^{\ell} = o(1).$$

Applying Markov's inequality, we have that $\Pr(X > 0) = \Pr(X \geq 1) = o(1)$, and the result follows.

Case 3. Finally, we deal with vertex sets $S \subseteq V_H$ with $(\log n)^2 \leq |S| \leq n/2$ and $\deg_H(S) > \max\{100d|S|, n/2\}$. Let $C = 10^4 d$. By the Cauchy-Schwarz inequality, we have that $|S| \sum_{i\in S} x_i^2 \geq (\deg_H(S))^2 \geq 10^4 d^2 |S|^2$, so using Item 2 of Definition 5.2.2 which ensures that $\sum_{i\in S} x_i^2 \leq dn$, we find that $|S| \leq n/C$.

Let $f = (eC)^{1/C}$. The number of sets S satisfying $|S| \leq n/C$ is at most $n\binom{n}{n/C} \leq nf^n$ since there are at most n possibilities for $|S|$ to consider, and for each of them $\binom{n}{|S|} \leq \binom{n}{n/C}$.

Fix any set $S \subseteq V_H$ with $|S| \leq n/C$ and consider the random construction of H, starting from half-edges in S^c (and choosing their mates in the pairing).

Let

$$j = \left\lfloor \frac{\deg_H(S^c)}{2} \right\rfloor \geq \left\lfloor \frac{3|S^c|}{2} \right\rfloor \geq \left\lfloor \frac{3n\left(1-\frac{1}{C}\right)}{2} \right\rfloor \geq \frac{3n\left(1-\frac{2}{C}\right)}{2},$$

where the first inequality uses the fact that each x_i is at least 3 (from Item 1 of Definition 5.2.2) and the final inequality uses the fact that n is sufficiently large.

Note that the process initiates a pairing from at least j half-edges in S^c. For each $i \in [j]$, let Y_i be the indicator random variable for the event that the i'th half-edge from which pairing is initiated connects to an endpoint in S (conditioned on the pairings of the first $i-1$ half-edges initiated from S^c).

Let $\varepsilon = 3(1-2/C)/(8\sqrt{d}) \leq 1/2$. From Item 2 in Definition 5.2.2 and the Cauchy-Schwarz inequality, $\left(\sum x_i\right)^2 \leq n \sum x_i^2 \leq n^2 d$ so $\sum_{i=1}^n x_i \leq \sqrt{d}n$. For any $t \in [j]$ satisfying $\sum_{i=1}^{t-1} Y_t < \varepsilon n/2$ we have

$$Pr(Y_t = 1) \geq \frac{\deg_H(S) - \varepsilon n/2}{\sqrt{d}n} > \frac{1-\varepsilon}{2\sqrt{d}} \geq \frac{1}{4\sqrt{d}}.$$

Now let $X_1, \ldots, X_j$ be i.i.d. Bernoulli random variables which are 1 with probability $1/(4\sqrt{d})$. We can couple the evolution of these variables so that, for any $t \in [j]$ satisfying $\sum_{i=1}^{t-1} Y_i < \varepsilon n/2$, we have $\sum_{i=1}^{t} Y_i \geq \sum_{i=1}^{t} X_i$. We conclude that $\Pr(\sum_{i=1}^{j} Y_i < \varepsilon n/2) \leq \Pr(\sum_{i=1}^{j} X_i < \varepsilon n/2)$.

To conclude we will show that $nf^n \Pr(\sum_{i=1}^{j} X_i < \varepsilon n/2) = o(1)$, implying that we can take $\alpha'' = \varepsilon/(2\sqrt{d})$ since $\varepsilon n/2 = \alpha''\sqrt{d}n \geq \alpha'' \deg_H(S)$.

Let $X = \sum_{i=1}^{j} X_i$ and $\delta = 1/2$. Note that $\mathbb{E}[X] = j/(4\sqrt{d})$ and that

$$\frac{(1-\delta)j}{4\sqrt{d}} \geq \frac{(1-\delta)3n\left(1-\frac{2}{C}\right)}{8\sqrt{d}} = \frac{\varepsilon n}{2}.$$

By a Chernoff bound, $\Pr(X \leq \varepsilon n/2) \leq \Pr(X \leq (1-\delta)j/(4\sqrt{d})) \leq \exp(-j\delta^2/(8\sqrt{d}))$. To conclude that $nf^n \exp(-j\delta^2/(8\sqrt{d})) = o(1)$ we observe that $f < \exp(3(1-2/C)\delta^2/(16\sqrt{d}))$. So with $\alpha'' = \varepsilon/(2\sqrt{d})$, we conclude Case 3.

To prove the lemma, combine the three cases by taking $\alpha = \min\{\alpha'/(100d), 1/100, \alpha''\}$. □

We can now prove the following result, which establishes the desired expansion properties of the multigraphs generated by the configuration model.

Lemma 5.6.5. *Let d be a positive real number. There is a positive real number α (depending on d) such that the following is true with high probability when $H = (V_H, E_H)$ is drawn from $\mathsf{CM}(n, \vec{x})$ uniformly over all degree sequences $\vec{x} \in \mathcal{D}_{n,d}$. For all connected vertex sets $S \subseteq V_H$ with $|S| \leq n/2$, we have that $e_H(S, S^c) \geq \alpha \deg_H(S)$.*

Proof. We consider three cases.

Case 1. Consider all connected subsets $S \subseteq V_H$ with $(\log n)^2 \leq |S| \leq n/2$. By Lemma 5.6.4 there is a positive real number α' such that, whp, every such subset S has $e_H(S, S^c) \geq \alpha' \deg_H(S)$.

Case 2. Consider all connected subsets $S \subseteq V_H$ with $|S| \leq (\log n)^2$ and $\deg_H(S) \geq 36$.

- Consider first those subsets S with $|S| \leq \frac{1}{6}\deg_H(S)$. We have that

$$\begin{aligned} e_H(S, S^c) &= \deg_H(S) - 2(t_{H[S]} + |S| - 1) \\ &\geq \frac{2}{3}\deg_H(S) - 2|S| \\ &\geq \frac{1}{3}\deg_H(S), \end{aligned}$$

  by Lemma 5.6.1 and our assumption on the size of S.

- Now consider those subsets S with $|S| > \frac{1}{6}\deg_H(S)$, then by Lemma 5.6.3, we have that $e_H(S, S^c) \geq |S|/4 \geq \deg_H(S)/24$.

Case 3. Finally, consider connected subsets $S \subseteq V_H$ with $|S| \leq (\log n)^2$ and $\deg_H(S) < 36$. By Lemma 5.6.3, we have that $e_H(S, S^c) \geq |S|/4 \geq 1/4 = 36/144 > \deg_H(S)/144$.

The result follows from the three cases by taking $\alpha = \min\{1/144, \alpha'\} = \alpha'$. □

Using the definition of $\mathcal{G}_\alpha$ and Lemma 2.5.1, we have the following corollary of Lemma 5.6.5.

Corollary 5.6.6. *Let d be a real number. There is a positive real number α (depending on d) such that the following holds. With high probability, when $G \sim \mathcal{G}(n, \vec{x})$ for some $\vec{x} \in \mathcal{D}_{n,d}$, it holds that $G \in \mathcal{G}_\alpha$.*

Combining Corollary 5.6.6 with Theorem 5.5.5 implies our main theorem.

Theorem 5.2.3. *Let d be a real number and let $q \geq 3$ be an integer. For the ferromagnetic Potts model, there is β_0 such that for all $\beta \geq \beta_0$ there is a poly-time approximate sampling algorithm for $\mu_{G,q,\beta}$ and an* FPRAS *for $Z_{G,q,\beta}$ that work with high probability on random graphs $G \sim \mathcal{G}(n, \vec{x})$ for any degree sequence $\vec{x} \in \mathcal{D}_{n,d}$.*

Proof. Let d be a real number and let $q \geq 2$ be an integer. Let α be the positive real number from Corollary 5.6.6. Let $\beta_0 = \frac{3}{\alpha}\log(8e^3(q-1))$. Consider $\vec{x} \in \mathcal{D}_{n,d}$ and let G be drawn from $\mathcal{G}(n, \vec{x})$. By Corollary 5.6.6, $G \in \mathcal{G}_\alpha$ whp. The result then follows by using the algorithms from Theorem 5.5.5. □

Remark 5.6.7. The bounds on β in Remark 5.2.4 follow from the choice of β_0 in the proof of Theorem 5.2.3 and from the fact that $\alpha = \Omega(\frac{1}{d\log d})$ which follows from the proofs of Lemmas 5.6.4 and 5.6.5. The running time bounds in Remark 5.2.4 come from those in Theorem 5.5.5 using the fact that $|E_G| = O(n)$ which follows from Item 2 of Definition 5.2.2.

Chapter 6

Glauber dynamics within a phase

The work in this chapter is based on Section 5 of the following paper:

[22] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. *Random Struct. Algorithms*, 58(2):294–321, 2021. Theorems 5 and 6 from arxiv.org/abs/1901.0665

An extended abstract also appeared at:

[21] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik

For spin systems such as the ferromagnetic Potts model or the hard-core model, it is known that the Glauber dynamics typically mixes exponentially slowly at low temperatures. In Chapter 1, we saw how this is the case for the low temperature Potts model on expander graphs. The proof of this fact exhibited a set of exponentially small bottlenecks in the state space that separate the

phases of the Potts model, which single vertex update Markov chains, such as the Glauber dynamics, struggle to cross. This is a well-understood phenomenon that has been exhibited more generally (see, for example, the work of Galvin and Tetali [47]), and is a fundamental obstacle to using the Glauber dynamics as the basis of an efficient low temperature sampling algorithm. In light of this fact, significant effort has been made to somehow circumvent these low temperature bottlenecks, and develop efficient algorithms for approximate sampling at low temperatures. Examples of such work include the truncated Taylor series approach [59, 68], the polymer dynamics Markov chain presented in Chapter 3, and the Swendsen-Wang dynamics [44].

Whilst the work of Chapter 3 introduces a Markov chain for sampling from the polymer representation of a spin system, a much more natural idea is to use a single-spin update Markov chain like the Glauber dynamics, but to start in one of the ground states of the model chosen at random. For example, in the case of the q-colour ferromagnetic Potts model, we might choose one of the q-colours with equal probability then start the Potts model Glauber dynamics in the monochromatic configuration corresponding to that colour. The intuition is that the Glauber dynamics will mix well within the portion of the state space close to the chosen ground state, and the randomness in the choice of ground state will ensure that an accurate sample from the full measure is obtained. Analysing this algorithm was suggested in [59] and [68]. While we are not yet able to show that this algorithm succeeds, we make partial progress. In this chapter, we show that Glauber dynamics, restricted to remain in a portion of the state space, mixes rapidly (in polynomial time).

6.1 Main results

For convenience, we present the results of this chapter in terms of the low temperature Potts model. Let $\alpha > 0$ be a real number and let $G = (V_G, E_G)$ be an α-expander. Let $q \geq 2$ be an integer and let $\beta > 0$ be a real number. For a ground state colour $g \in [q]$ and an integer M, we let $\Omega^g_{G,M}$ denote the set of q-colourings of the vertices of G so that every connected component of G coloured with the palette of colours $[q] \setminus \{g\}$ is of size at most M. In [68] it is shown that for an appropriate choice of M, the set $\{\Omega^g_{G,M}\}_{g\in[q]}$ forms an 'almost partition'

of the set of all colourings, in that the weight of both the overlap of the almost partition and the set of colourings uncovered by the almost partition is at most ε under the conditions of Theorem 3.2.4. In particular, an ε-approximate sample from the Potts model restricted to $\Omega^g_{G,M}$ for $M = O(\log(n/\varepsilon))$ is enough (by symmetry) to obtain a $(q\varepsilon)$-approximate sample from the Gibbs distribution of the Potts model $\mu_{G,q,\beta}$ (cf. Lemma 6.4.1). Using Markov chain comparison techniques, we show in this chapter that this can be done using the usual spin Glauber dynamics, restricted to remain in $\Omega^g_{G,M}$.

Theorem 6.1.1. *Let $\alpha > 0$ be a real number, let $q \geq 2$ and $\Delta \geq 3$ be integers, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. Fix a colour from $[q]$, say, red. For any n-vertex α-expander graph $G = (V_G, E_G)$ of maximum degree at most Δ and any $\varepsilon \in (0,1)$, for $M = O(\log(n/\varepsilon))$ the Glauber dynamics restricted to $\Omega^{red}_{G,M}$ has mixing time $T_{mix}(\varepsilon)$ polynomial in n and $1/\varepsilon$.*

We remark that the polynomial bound in Theorem 6.1.1 depends on q, Δ, and α exponentially, see the relevant Theorem 6.3.2 and Section 6.4.1 for details. Theorem 6.1.1 shows that despite exponentially slow mixing of the Glauber dynamics on the full state space [8], it can still be used to obtain a polynomial-time approximate sampling algorithm. We leave for future work two important extensions that would complete the picture:

1. showing that *unrestricted* Glauber dynamics starting from a well chosen configuration works
2. lowering the running time to $O(n \log n)$ from the large polynomial we obtain in the theorem.

In Section 6.3.1, we state a general theorem (Theorem 6.3.2) comparing the polymer model dynamics to spin model dynamics as well as a specific result for the hard-core model (Theorem 6.4.2).

6.2 Preliminaries

The results presented in this chapter are obtained, using fairly standard Markov chain comparison techniques, from those presented in Chapter 3. Before we begin, we first present some definitions relating to the restricted Glauber dynamics, and the comparison techniques we use to bound its mixing time.

6.2.1 Restricted Glauber dynamics for polymer models

Here, we define the restricted Glauber dynamics for subset polymer models, and show the upcoming Theorem 6.3.2 which bounds its mixing time under some appropriate conditions. The restricted Glauber dynamics is a Markov chain on spin configurations, where the state space is restricted to those which correspond to the polymers of a certain truncated polymer model.

Let $q \geq 2$ and let $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and consider a family of q-spin polymer models $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ with degree bound Δ. Fix a graph $G = (V_G, E_G) \in \mathcal{G}$. Recall from Section 2.6 that g_v denotes the set of ground state spins for each vertex $v \in V_{J_G}$, and that Ω_G denotes the set of all mutually compatible sets of polymers from $\mathcal{C}_G$. There is a natural map $f \colon \Omega_G \to \{0, 1, \ldots, q-1\}^{V_G}$ between allowed polymer configurations and spin configurations, given by $f(\Gamma)_v = \sigma_\gamma(v)$ if $\gamma \in \Gamma$ and $v \in V_\gamma$ and $f(\Gamma)_v = g_v$ if $v \notin \cup_{\gamma \in \Gamma} V_\gamma$. Let $\Omega_G^{\text{spin}} = f(\Omega_G)$ be the spin configurations obtainable as images of the map f. It will be helpful to consider the inverse map f^{-1} and extend its domain to all $\sigma \in \{0, 1, \ldots, q-1\}^{V(G)}$, so that $f^{-1}(\sigma)$ is the polymer configuration consisting of polymers that are connected components of vertices which do not receive their ground state spin; note that the range of the extended f^{-1} is not limited to Ω_G anymore.

The restricted Glauber dynamics is defined as follows, starting from $\Gamma_t \in \Omega_G$.

1. Choose $v \in V_G$ and $s \in \{0, \ldots, q-1\}$ uniformly.
2. Γ' is formed from Γ_t by assigning v to spin s (formally, by letting $\sigma = f(\Gamma_t)$, forming σ' from σ by assigning v to spin s, and finally letting $\Gamma' = f^{-1}(\sigma')$).
3. If $\Gamma' \in \Omega_G$ let $p = \min(1, w(\Gamma')/w(\Gamma))$.
   (a) With probability p, $\Gamma_{t+1} = \Gamma'$.
   (b) With probability $1 - p$, $\Gamma_{t+1} = \Gamma_t$.
4. If $\Gamma' \notin \Omega_G$ then $\Gamma_{t+1} = \Gamma_t$.

We will use the Markov chain comparison technique to show that the restricted Glauber dynamics is rapidly mixing. To do this, we need a mild condition on the set of allowed polymers $\mathcal{C}_G$. A polymer model is said to be *single-update-compatible* if, for every size-k polymer $\gamma \in \mathcal{C}_G$, there is an ordering $v_1, \ldots, v_k$ of the vertices of V_γ such that, for all $i \in [k]$, the set $S_i = \{v_1, \ldots, v_i\}$ induces a connected subgraph and we have that $(S_i, \sigma_\gamma|_{S_i})$ is a valid polymer itself, i.e., $(S_i, \sigma_\gamma|_{S_i}) \in \mathcal{C}_G$.

We will use the comparison method of Diaconis and Saloff-Coste [25, 26] as applied to mixing times by Randall and Tetali [80]. In order to avoid discussion of eigenvalues here, we use the version from Observation 13 of the survey paper [29]. We first show that the restricted Glauber dynamics is a reversible ergodic Markov chain with stationary distribution μ_G, which is easy to see from its definition.

Lemma 6.2.1. *Let $q \geq 2$ be an integer. Let G be a graph and let $(\mathcal{C}_G, w_G, J_G)$ be a single-update-compatible q-spin polymer model. The restricted Glauber dynamics is ergodic and reversible with stationary distribution μ_G.*

Proof. The restricted Glauber dynamics is aperiodic since we remain in the same state with positive probability after performing an update. It is irreducible since we can transition from any $\Gamma \in \Omega_G$ to any $\Gamma' \in \Omega_G$ by adding and removing vertices one-by-one. This shows that the restricted Glauber dynamics is ergodic. To show that it is reversible and has stationary distribution μ_G, we check detailed balance. Suppose $\Gamma, \Gamma' \in \Omega_G$ with $\Gamma \neq \Gamma'$ and $P(\Gamma, \Gamma') > 0$ where P is the transition matrix of the restricted Glauber dynamics. Then,

$$\frac{P(\Gamma, \Gamma')}{P(\Gamma', \Gamma)} = \frac{\frac{1}{n} \cdot \frac{1}{q} \cdot \min\{1, w_G(\Gamma')/w_G(\Gamma)\}}{\frac{1}{n} \cdot \frac{1}{q} \cdot \min\{1, w_G(\Gamma)/w_G(\Gamma')\}} = \frac{w_G(\Gamma')}{w_G(\Gamma)} = \frac{\mu_G(\Gamma')}{\mu_G(\Gamma)},$$

where n is the number of vertices of G. □

6.2.2 Comparison of Markov chains

We next give some standard definitions that will be used in our comparison proof. Let $\mathcal{M}$ denote the restricted Glauber dynamics and P be its transition matrix. Let $\mathcal{M}'$ be the polymer dynamics and denote its transition matrix by P'. Define $E^*(\mathcal{M})$ to be the set of pairs of configurations (Γ, Γ') that can be achieved

by one transition of the restricted Glauber dynamics; i.e., $E^*(\mathcal{M}) = \{(\Gamma, \Gamma') \in \Omega_G^2 : P(\Gamma, \Gamma') > 0\}$. Similarly, define $E^*(\mathcal{M}') = \{(\Gamma, \Gamma') \in \Omega_G^2 : P'(\Gamma, \Gamma') > 0\}$ for the polymer dynamics.

For every $(\Gamma, \Gamma') \in E^*(\mathcal{M}')$, we define a path $\mathcal{P}_{\Gamma,\Gamma'}$ from Γ to Γ' to be a sequence of configurations such that every adjacent pair is a transition of the restricted Glauber dynamics; i.e, every adjacent pair of configurations is in $E^*(\mathcal{M})$. For this, we assume that the polymer model is single-update-compatible (see Section 6.2.1). If $\Gamma = \Gamma'$, then the choice is easy — we let $\mathcal{P}_{\Gamma,\Gamma'} = (\Gamma, \Gamma')$. Suppose instead that $\Gamma' = \Gamma \cup \gamma$ for some polymer $\gamma \in \Omega_G$. Recall that there is a natural one-to-one mapping $f \colon \Omega_G \to \Omega_G^{\text{spin}}$ between the set of all (polymer) configurations Ω_G and the set of spin configurations Ω_G^{spin}. Let $\sigma = f(\Gamma)$ and $\sigma' = f(\Gamma')$ be the corresponding spin configurations. If $|V_\gamma| = k$, let $v_1, \ldots, v_k$ be the ordering of vertices of V_γ from the definition of single-update-compatible so that, for all $i \in [k]$, the polymer induced by vertices $v_1, \ldots, v_i$ is in $\mathcal{C}_G$. Let $(\sigma = \sigma_0, \sigma_1, \ldots, \sigma_k = \sigma')$ be the sequence of spin configurations such that each σ_j is obtained from σ_{j-1} by changing the spin of v_j from $\sigma(v) = g_v$ to $\sigma'(v)$. The path $\mathcal{P}_{\Gamma,\Gamma'}$ is then defined to be $(f^{-1}(\sigma_0), \ldots, f^{-1}(\sigma_k))$. If $\Gamma' = \Gamma \setminus \{\gamma\}$ for some $\gamma \in \Omega_G$, we can define the path $\mathcal{P}_{\Gamma,\Gamma'}$ in a similar manner. Note that in both cases the length of the path is $|\mathcal{P}_{\Gamma,\Gamma'}| = k = |V_\gamma|$.

For every $(\Gamma_0, \Gamma_0') \in E^*(\mathcal{M})$, the congestion of the edge (Γ_0, Γ_0') is defined to be

$$A(\Gamma_0, \Gamma_0') = \frac{1}{\mu_G(\Gamma_0) P(\Gamma_0, \Gamma_0')} \sum_{\substack{(\Gamma,\Gamma') \in E^*(\mathcal{M}'): \\ \mathcal{P}_{\Gamma,\Gamma'} \ni (\Gamma_0, \Gamma_0')}} \mu_G(\Gamma) P'(\Gamma, \Gamma') |\mathcal{P}_{\Gamma,\Gamma'}|.$$

The congestion of the choice of paths is the quantity

$$A = \max_{(\Gamma_0, \Gamma_0') \in E^*(\mathcal{M})} A(\Gamma_0, \Gamma_0').$$

6.3 Comparison to the polymer dynamics

The following comparison lemma gives an upper bound on the mixing time of the restricted Glauber dynamics by the mixing time of the polymer dynamics.

Lemma 6.3.1 ([29, Observation 13])**.** *Let* $c_1 = \min_{\Gamma\in\Omega_G} P(\Gamma,\Gamma)$ *and* $c_2 = \min_{\Gamma\in\Omega_G} \mu_G(\Gamma)$*. Then, for any* $\varepsilon \in (0,1)$ *we have*

$$T_{mix}(\mathcal{M},\varepsilon) \le \max\left\{A\left(T_{mix}\left(\mathcal{M}',\frac{1}{2e}\right)+1\right),\frac{1}{2c_1}\right\}\ln\frac{1}{\varepsilon c_2}.$$

6.3.1 A mixing time bound for spin Glauber

We now proceed to establish the mixing time of the restricted Glauber dynamics, which is the main result of this section. We will apply this to both the hard-core model (on bipartite α-expander graphs) and the ferromagnetic Potts model (on α-expander graphs), for which we will define appropriate single-update compatible polymer models. Furthermore, in both of these applications, M below will be logarithmic in n/ε, giving polynomial mixing time for the restricted Glauber dynamics.

Theorem 6.3.2. *Let* $q \ge 2$ *and* $\Delta \ge 3$ *be integers. Let* $\mathcal{G}$ *be a class of graphs and let* $\{(\mathcal{C}_G, w_G, J_G) \mid G \in \mathcal{G}\}$ *be a family of* q*-spin polymer models with degree bound* Δ*, that satisfy the polymer mixing condition (Definition 2.6.1). Consider a graph* $G \in \mathcal{G}$ *such that* $(\mathcal{C}_G, w_G, J_G)$ *is single-update-compatible. Let* $M = \max\{|V_\gamma|\colon \gamma \in \mathcal{C}_G\}$*. Suppose that, for every pair of configurations* $\Gamma,\Gamma' \in \Omega_G$ *whose corresponding spin configurations* $f(\Gamma), f(\Gamma') \in \Omega_G^{\text{spin}}$ *differ at exactly one vertex, we have*

$$\frac{1}{\eta} \le \frac{\mu_G(\Gamma)}{\mu_G(\Gamma')} \le \eta, \tag{6.1}$$

for some constant $\eta > 1$*. Then for any* $\varepsilon \in (0,1)$*, the restricted Glauber dynamics has mixing time*

$$T_{mix}(\mathcal{M},\varepsilon) \le O\left(M\eta^{M+1}n^2\log n\log(\eta/\varepsilon)\right).$$

Proof. By Lemma 6.3.1, it suffices to upper bound the congestion $A(\Gamma_0,\Gamma_0')$ for every $(\Gamma_0,\Gamma_0') \in E^*(\mathcal{M})$ where

$$A(\Gamma_0,\Gamma_0') = \sum_{\substack{(\Gamma,\Gamma')\in E^*(\mathcal{M}'):\\ \mathcal{P}_{\Gamma,\Gamma'}\ni(\Gamma_0,\Gamma_0')}} \frac{\mu_G(\Gamma)}{\mu_G(\Gamma_0)}\cdot\frac{P'(\Gamma,\Gamma')}{P(\Gamma_0,\Gamma_0')}\cdot|\mathcal{P}_{\Gamma,\Gamma'}|.$$

If $\Gamma_0 = \Gamma_0'$, then for our choices of paths to get $(\Gamma_0, \Gamma_0') \in \mathcal{P}_{\Gamma,\Gamma'}$ we must have $\Gamma = \Gamma' = \Gamma_0 = \Gamma_0'$. It follows that

$$A(\Gamma_0, \Gamma_0) = \frac{P'(\Gamma_0, \Gamma_0)}{P(\Gamma_0, \Gamma_0)} \leq \frac{1}{1/q} = q$$

since $P(\Gamma_0, \Gamma_0) \geq 1/q$ by the update rule of the restricted Glauber dynamics.

Now suppose $\Gamma_0 \neq \Gamma_0'$. Let $\sigma_0 = f(\Gamma_0)$ and $\sigma_0' = f(\Gamma_0')$ be the corresponding spin configurations. Notice that σ_0 and σ_0' differ at exactly one vertex, which we denote by v. If $\sigma_0(v) \neq g_v$ and $\sigma_0'(v) \neq g_v$ then no path $\mathcal{P}_{\Gamma,\Gamma'}$ would contain (Γ_0, Γ_0') by our choice of paths, and thus $A(\Gamma_0, \Gamma_0) = 0$. Assume next that $\sigma_0(v) = g_v$ and $\sigma_0'(v) \neq g_v$. Then, if $(\Gamma_0, \Gamma_0') \in \mathcal{P}_{\Gamma,\Gamma'}$ for some $(\Gamma, \Gamma') \in E^*(\mathcal{M}')$, we must have $\Gamma' = \Gamma \cup \{\gamma\}$ for some polymer $\gamma \in \mathcal{C}_G$ and also $v \in V_\gamma$. Moreover, the spin configurations $f(\Gamma), f(\Gamma'), \sigma_0, \sigma_0'$ are all the same outside V_γ. This implies that the number of such paths $\mathcal{P}_{\Gamma,\Gamma'}$ is upper bounded by the number of polymers containing v.

Now fix some $(\Gamma, \Gamma') \in E^*(\mathcal{M}')$ such that $(\Gamma_0, \Gamma_0') \in \mathcal{P}_{\Gamma,\Gamma'}$ and assume that $\Gamma' = \Gamma \cup \{\gamma\}$ for some polymer $\gamma \in \mathcal{C}_G$ with $v \in V_\gamma$. Then,

$$|\mathcal{P}_{\Gamma,\Gamma'}| \leq |V_\gamma| \leq M. \tag{6.2}$$

As the path $\mathcal{P}_{\Gamma,\Gamma'}$ is obtained by changing the spins vertex by vertex in the corresponding spin configurations, $f(\Gamma)$ and $f(\Gamma_0)$ differ at most $|V_\gamma|$ vertices. The condition of the theorem implies that

$$\frac{\mu_G(\Gamma)}{\mu_G(\Gamma_0)} \leq \eta^{|V_\gamma|} \leq \eta^M. \tag{6.3}$$

The update rule of the restricted Glauber dynamics gives

$$P(\Gamma_0, \Gamma_0') = \frac{1}{n} \cdot \frac{1}{q} \cdot \min\left\{1, \frac{w_G(\Gamma_0')}{w_G(\Gamma_0)}\right\} \geq \frac{1}{\eta q n} \tag{6.4}$$

and for the polymer dynamics we have

$$P'(\Gamma, \Gamma') = \frac{|V_\gamma|}{n} \cdot \frac{1}{2} \cdot w_G(\gamma) = \frac{|V_\gamma| \cdot w_G(\gamma)}{2n}. \tag{6.5}$$

Let γ_v denote a polymer on $\{v\}$ with a spin from $\{0, \ldots, q-1\} \setminus g_v$. Then, the polymer mixing condition implies that $\sum_{\gamma \nsim \gamma_v} |V_\gamma| w_G(\gamma) \leq \theta |V_{\gamma_v}| < 1$ for

some $\theta \in (0,1)$. Combining this with inequalities (6.2), (6.3), (6.4), and (6.5), we obtain that

$$\begin{aligned} A(\Gamma_0, \Gamma_0') &\leq \sum_{\gamma \,:\, v \in V_\gamma} \eta^M \cdot \eta q n \cdot \frac{|V_\gamma| \cdot w_G(\gamma)}{2n} \cdot M \\ &= \frac{1}{2} q M \eta^{M+1} \sum_{\gamma \nsim \gamma_v} |V_\gamma| \cdot w_G(\gamma) \\ &\leq \frac{1}{2} q M \eta^{M+1}. \end{aligned}$$

For the case where $\sigma_0(v) \neq g_v$ and $\sigma_0'(v) = g_v$, the proof is almost the same and we can get the same bound. Thus,

$$A = \max_{(\Gamma_0, \Gamma_0') \in E^*(\mathcal{M})} A(\Gamma_0, \Gamma_0') \leq \max\left\{q, \frac{1}{2} q M \eta^{M+1}\right\} \leq q M \eta^{M+1}.$$

The theorem then follows from Theorem 3.2.1 and Lemma 6.3.1 once we notice that $P(\Gamma, \Gamma) \geq 1/q$ and that $\mu_G(\Gamma) \geq 1/(\eta q)^n$ for all $\Gamma \in \Omega_G$. □

6.3.2 Truncated polymer model

The bound on the mixing time of the restricted Glauber dynamics in Theorem 6.3.2 is exponential in the size of the largest polymer which is in general undesirable. Here, we show that, under the polymer sampling condition, we can restrict our attention to polymers of size $O(\log n)$ in the sense that the partition function as well as the Gibbs distribution of the truncated polymer model are close to those of the original polymer model.

For $k > 0$ and $G \in \mathcal{G}$, define the truncated polymer model $(\mathcal{C}_{G,k}, w_G, J_G)$ by

$$\mathcal{C}_{G,k} = \{\gamma \in \mathcal{C}_G : |V_\gamma| \leq k\}.$$

Also, we let

$$\Omega_{G,k} = \{\Gamma \in \Omega_G \mid \Gamma \subseteq \mathcal{C}_{G,k}\} = \{\Gamma \subseteq \mathcal{C}_{G,k} \mid \forall \gamma, \gamma' \in \Gamma_G, \gamma \sim \gamma'\}$$

be the set of allowed configurations (note that $\Omega_{G,k} \subseteq \Omega_G$). The partition function of the truncated polymer model $(\mathcal{C}_{G,k}, w_G, J_G)$ is given by

$$Z_{G,k} = \sum_{\Gamma \in \Omega_{G,k}} \prod_{\gamma \in \Gamma} w_G(\gamma).$$

The corresponding Gibbs distribution on $\Omega_{G,k}$ is defined by $\mu_{G,k}(\Gamma) = \frac{\prod_{\gamma\in\Gamma} w_G(\gamma)}{Z_{G,k}}$. We remark that if the original family of polymer models $\{(\mathcal{C}_{G,k}, w_G, J_G) \mid G \in \mathcal{G}\}$ satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$, then so does the truncated polymer model, and thus Theorem 3.2.1 also applies to the truncated model.

The following lemma asserts that the Gibbs distribution and the partition function of the truncated polymer model $(\mathcal{C}_{G,k}, w_G, J_G)$ are close to those of the original model $(\mathcal{C}_G, w_G, J_G)$, provided that the polymer sampling condition holds.

Lemma 6.3.3. *Let $q \geq 2$ and $\Delta \geq 3$ be integers. Let $\mathcal{G}$ be a class of graphs and let $\{(\mathcal{C}_{G,k}, w_G, J_G) \mid G \in \mathcal{G}\}$ be a family of q-spin polymer models with degree bound Δ, that satisfy the polymer sampling condition (Definition 2.6.3) with constant $\tau \geq 5 + 3\log((q-1)\Delta)$. Let G be an n-vertex graph from $\mathcal{G}$. Then, for any $\varepsilon > 0$ and $k = \frac{3\log(2n/\varepsilon)}{2\tau}$, it follows that*

$$Z_{G,k} \leq Z_G \leq e^{\varepsilon} Z_{G,k}.$$

Moreover, the total variation distance between μ_G and $\mu_{G,k}$ is at most ε.

Proof. Note that $Z_{G,k} \leq Z_G$ follows immediately from $\Omega_{G,k} \subseteq \Omega_G$. For $\Gamma \in \Omega_{G,k}$, let $\Omega(\Gamma) = \{\Gamma' \in \Omega_G : \Gamma' \cap \mathcal{C}_{G,k} = \Gamma\}$ and let

$$\zeta(\Gamma) = \sum_{\Gamma'\in\Omega(\Gamma)} \prod_{\gamma\in\Gamma'} w_G(\gamma), \text{ so that } Z_G = \sum_{\Gamma\in\Omega_{G,k}} \zeta(\Gamma). \tag{6.6}$$

Let $\mathcal{C}^+_{G,k} = \mathcal{C}_G \setminus \mathcal{C}_{G,k} = \{\gamma \in \mathcal{C}_G \colon |V_\gamma| > k\}$ be the collection of all polymers of size greater than k. Notice that for each $\Gamma \in \Omega_{G,k}$ we have the crude bound

$$\zeta(\Gamma) \leq \prod_{\gamma\in\Gamma} w_G(\gamma) \prod_{\gamma\in\mathcal{C}^+_{G,k}} (1 + w_G(\gamma)). \tag{6.7}$$

Combining (6.6) and (6.7), we obtain that

$$Z_G \leq \prod_{\gamma\in\mathcal{C}^+_{G,k}} (1 + w_G(\gamma)) \sum_{\Gamma\in\Omega_{G,k}} \prod_{\gamma\in\Gamma} w_G(\gamma) = Z_{G,k} \prod_{\gamma\in\mathcal{C}^+_{G,k}} (1 + w_G(\gamma)). \tag{6.8}$$

Since $(1 + x) \leq e^x$ for all real x, we have that

$$\log\Bigg(\prod_{\gamma\in\mathcal{C}^+_{G,k}} (1 + w_G(\gamma))\Bigg) \leq \sum_{\gamma\in\mathcal{C}^+_{G,k}} w_G(\gamma) \leq \sum_{\gamma\in\mathcal{C}^+_{G,k}} e^{-\tau|V_\gamma|}. \tag{6.9}$$

The last inequality follows from the fact that the family of polymer models $\{(\mathcal{C}_{G,k}, w_G, J_G) \mid G \in \mathcal{G}\}$ satisfies the polymer sampling condition with constant $\tau \geq 5 + 3\log((q-1)\Delta)$. Then we deduce from Lemma 3.3.1 that

$$\sum_{\gamma \in \mathcal{C}^+_{G,k}} e^{-\tau|V_\gamma|} = \sum_{v \in V} \sum_{\ell \geq k} \sum_{\substack{\gamma \in \mathcal{C}^+_{G,k}: \\ |V_\gamma| = \ell, \\ v \in V_\gamma}} e^{-\tau\ell} \leq n \sum_{\ell \geq k} \left(\frac{e(q-1)\Delta}{e^\tau} \right)^\ell, \tag{6.10}$$

and since $\tau \geq 5 + 3\log((q-1)\Delta)$, we obtain that $e(q-1)\Delta \leq e^{\tau/3}$. It follows that

$$\sum_{\ell \geq k} \left(\frac{e(q-1)\Delta}{e^\tau} \right)^\ell \leq 2\exp\left(-\frac{2}{3}\tau k \right) = \frac{\varepsilon}{n}. \tag{6.11}$$

Combining (6.8), (6.9), (6.10), and (6.11) yields $Z_G \leq e^\varepsilon Z_{G,k}$, as needed. Finally, we bound the total variation distance between μ_G and $\mu_{G,k}$:

$$\|\mu_G - \mu_{G,k}\|_{\mathsf{TV}} = \mu_G(\Omega_G \setminus \Omega_{G,k}) = \frac{Z_G - Z_{G,k}}{Z_G} \leq 1 - e^{-\varepsilon} \leq \varepsilon,$$

where the first equality is because $\mu_G(\Gamma) > \mu_{G,k}(\Gamma)$ if and only if $\Gamma \in \Omega_G \setminus \Omega_{G,k}$, for which we have $\mu_{G,k}(\Gamma) = 0$. □

6.4 Applications

In this section, we apply the previous results to show that (spin) Glauber dynamics for the ferromagnetic Potts and hard-core models mix in polynomial time on expander graphs, when restricted to configurations close to the ground states (which, as we have already seen, constitute the main portion of the probability space at low temperatures).

6.4.1 Ferromagnetic Potts model

Let $\alpha > 0$ be a real number, let $q \geq 2$ and $\Delta \geq 3$ be integers, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. Let $\mathcal{G}_{\Delta,\alpha}$ be the class of α-expander graphs (as per Definition 2.5.2) of maximum degree at most Δ. In this section, we prove Theorem 6.1.1 for the q-color ferromagnetic Potts model. Fix a colour from $[q]$, say, red. For $M > 0$, consider the family of polymer models for the red phase of the Potts model, as defined in Section 3.5.1, defined over the class

$\mathcal{G}_{\Delta,\alpha}$. We denote this family of polymer models by $\mathcal{F}^{\text{red}}_{\mathcal{G}_\Delta,M} = \{(\mathcal{C}^{\text{red}}_{G,M}, w^{\text{red}}_G, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$. For $G \in \mathcal{G}_{\Delta,\alpha}$, let $\Omega^{\text{red}}_{G,M}$ denote the set of mutually compatible polymer configurations of the polymer model $(\mathcal{C}^{\text{red}}_{G,M}, w^{\text{red}}_G, G)$, let $\mu^{\text{red}}_{G,M}$ be its Gibbs distribution, and let $Z^{\text{red}}_{G,M}$ be its partition function.

Theorem 6.1.1. *Let $\alpha > 0$ be a real number, let $q \geq 2$ and $\Delta \geq 3$ be integers, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. Fix a colour from $[q]$, say, red. For any n-vertex α-expander graph $G = (V_G, E_G)$ of maximum degree at most Δ and any $\varepsilon \in (0,1)$, for $M = O(\log(n/\varepsilon))$ the Glauber dynamics restricted to $\Omega^{red}_{G,M}$ has mixing time $T_{mix}(\varepsilon)$ polynomial in n and $1/\varepsilon$.*

Proof. Fix an n-vertex graph $G = (V_G, E_G) \in \mathcal{G}_{\Delta,\alpha}$. From the proof of Theorem 3.2.4, we have that $\mathcal{F}^{\text{red}}_{\mathcal{G}_\Delta,n/2}$ satisfies the polymer sampling condition with constant $\tau = \alpha\beta \geq 5+3\log((q-1)\Delta)$ and hence so does the family of truncated polymer models $\mathcal{F}^{\text{red}}_{\mathcal{G}_\Delta,M}$.

The result therefore follows by applying Theorem 6.3.2, after observing that

1. the polymer model $(\mathcal{C}^{\text{red}}_{G,M}, w^{\text{red}}_G, G)$ is single-update-compatible (use DFS ordering), and

2. for a pair of polymer configurations $\Gamma, \Gamma' \in \Omega^{\text{red}}_{G,M}$ whose corresponding spin configurations σ, σ' differ at a vertex, we have that

$$\frac{\mu^{\text{red}}_{G,M}(\Gamma)}{\mu^{\text{red}}_{G,M}(\Gamma')} = \frac{e^{-\beta m_G(\sigma)}}{e^{-\beta m_G(\sigma')}} \in [1/\eta, \eta].$$

   where $\eta = \exp(\beta\Delta)$ (since G has maximum degree Δ, changing the spin of a vertex can create at most Δ new monochromatic/bichromatic edges).

□

The following lemma justifies that the set $\Omega^{\text{red}}_{G,M}$ with $M = O(\log(n/\varepsilon))$ constitutes for all but ε of the aggregate weight of colourings in the Potts distribution on G.

Lemma 6.4.1. *Let $\alpha > 0$ be a real number, let $q \geq 2$ and $\Delta \geq 3$ be integers, and let $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha}$ be a real number. Fix a colour from $[q]$, say, red. For any n-vertex α-expander graph G of maximum degree Δ and any $qe^{-n} \leq \varepsilon < 1$, for $M = \frac{3\log(4n/\varepsilon)}{2\alpha\beta}$, we have that $qZ^{red}_{G,M}$ is an ε-approximation to the partition function of the Potts model $Z_{G,q,\beta}$.*

Proof. Fix an n-vertex graph $G = (V_G, E_G) \in \mathcal{G}_{\Delta,\alpha}$. Since $\beta \geq \frac{5+3\log((q-1)\Delta)}{\alpha} > \frac{2\log(eq)}{\alpha}$ and $\varepsilon \geq qe^{-n}$, by Lemma 3.6.1 we have that $qZ^{\text{red}}_{G,n/2}$ is an $(\varepsilon/2)$-approximation to $Z_{G,q,\beta}$. From the proof of Theorem 3.2.4, we have that $\mathcal{F}^{\text{red}}_{\mathcal{G}_\Delta,n/2}$ satisfies the polymer sampling condition with constant $\tau = \alpha\beta \geq 5 + 3\log((q-1)\Delta)$ and hence so does the family of truncated polymer models $\mathcal{F}^{\text{red}}_{\mathcal{G}_\Delta,M}$. By Lemma 6.3.3, it follows that for $M = \frac{3\log(4n/\varepsilon)}{2\alpha\beta}$, we have that $qZ^{\text{red}}_{G,M}$ is an $(\varepsilon/2)$-approximation to $qZ^{\text{red}}_{G,n/2}$. Therefore, $qZ^{\text{red}}_{G,M}$ is an ε-approximation to $Z_{G,q,\beta}$. □

6.4.2 Hard core model

In this section, we state and prove the analogue of Theorem 6.1.1 for the hard-core model. In particular, let $G = (V^0_G, V^1_G, E_G)$ be an n-vertex bipartite α-expander graph (as per Definition 2.5.3) of maximum degree at most Δ, and for $i \in \{0,1\}$ and $M > 0$, let $\Omega^i_{G,M}$ denote the independent sets I whose deviations from the ground state V^i_G consists of small connected components, more precisely, $(V^i_G \setminus I) \cup (I \cap V^{1-i}_G)$ consists of connected components of size at most M. Using similar methods to Section 6.4.1, we will prove the following result.

Theorem 6.4.2. *Let α be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (6\Delta)^{3+6/\alpha}$ be a real number. For any n-vertex bipartite α-expander graph with maximum degree at most Δ and any $\varepsilon \in (0,1)$ and $i \in \{0,1\}$, with $M = O(\log(n/\varepsilon))$, the Glauber dynamics restricted to $\Omega^i_{G,M}$ has mixing time $T_{mix}(\varepsilon)$ polynomial in n and $1/\varepsilon$.*

As we shall see soon in the upcoming Lemma 6.4.5, and for λ large enough, the set $\Omega^0_{G,M} \cup \Omega^1_{G,M}$ for $M = \Theta(\log(n/\varepsilon))$ captures all but ε weight of the hard-core partition function and hence Theorem 6.4.2 can be used to obtain another polynomial time algorithm for the hard-core model on expanding graphs G in that regime.

To prove Theorem 6.4.2, it will be simpler to work with somewhat different polymer models than that of Section 3.5.2. These models were originally used in [66] (the conference version of [68]). For $i \in \{0,1\}$, and following [66], we will define a polymer model $(\mathcal{C}^i_G, w^i_G, G)$. The host graph is G and the model will capture deviations from the ground state V^i_G: a polymer $\gamma = (V_\gamma, \sigma_\gamma)$ will

be a connected set of vertices in G such that $(V_G^i \setminus I) \cup (I \cap V_G^{1-i}) = V_\gamma$, for some independent set I. Specifically, the set $\mathcal{C}_G^i$ of allowed polymers consists of all connected sets of vertices V_γ (in G) such that $|V_\gamma \cap V_G^i| \leq |V_G^i|/4$ and for any $v \in V_\gamma \cap V_G^{1-i}$, all of the neighbors of v (in G) are also in V_γ. The set of spins is $\{0,1\}$ and the ground state spin for a vertex v is $g_v = 1$ if $v \in V_G^i$, and $g_v = 0$ if $v \in V_G^{1-i}$; the spin assignment σ_γ is given by $1 - g_v$ for $v \in V_\gamma$. The weight of a polymer $\gamma \in \mathcal{C}_G^i$ is defined as

$$w_G^i(\gamma) = \frac{\lambda^{|V_\gamma \cap V_G^{1-i}|}}{\lambda^{|V_\gamma \cap V_G^i|}}.$$

The main observation behind the definition of the weight $w_G^i(\gamma)$ is that the weight of an independent set I such that $(V_G^i \setminus I) \cup (I \cap V_G^{1-i}) = V_\gamma$ is $\lambda^{|V_G^i|} w_G^i(\gamma)$.

Following again Section 6.3.2, it will be relevant to consider, for $M > 0$, the truncated polymer model $(\mathcal{C}_{G,M}^i, w_G^i, G)$ whose polymers are of size at most M; observe that the set $\Omega_{G,M}^i$ defined above is precisely the set of allowable polymer configurations in the truncated polymer model. We next verify that the polymer sampling condition (Definition 2.6.3) holds with constant $\tau \geq 5 + 3\log\Delta$, for the family of polymer models defined above, over the class of bounded-degree bipartite expanders.

Lemma 6.4.3. *Let $\alpha > 0$ be a real number, let $\Delta \geq 3$ be an integer, let $\mathcal{G}_{\Delta,\alpha}$ be the class of bipartite α-expander graphs (as per Definition 2.5.3) with maximum degree at most Δ, and let $\lambda \geq (6\Delta)^{3+6/\alpha}$ be a real number. For all $i \in \{0,1\}$, the family of polymer models defined by $\{(\mathcal{C}_G^i, w_G^i, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$ satisfies the polymer sampling condition 2.6.3 with constant $\tau = \frac{\alpha}{2+\alpha}\log\lambda \geq 5 + 3\log\Delta$.*

Proof. For $v \in V_\gamma \cap V_G^{1-i}$, we have that all of the neighbors of v are also in V_γ and hence, by the α-expansion of G, we have that $|V_\gamma \cap V_G^i| \geq (1+\alpha)|V_\gamma \cap V_G^{1-i}|$. This gives $(2+\alpha)|V_\gamma \cap V_G^i| \geq (1+\alpha)|V_\gamma|$ and therefore

$$w_\gamma^i = \frac{\lambda^{|V_\gamma \cap V_G^{1-i}|}}{\lambda^{|V_\gamma \cap V_G^i|}} \leq \frac{\lambda^{\frac{1}{1+\alpha}|V_\gamma \cap V_G^i|}}{\lambda^{|V_\gamma \cap V_G^i|}} \leq \lambda^{-\frac{\alpha}{2+\alpha}|V_\gamma|} = e^{-\tau|V_\gamma|},$$

for $\tau = \frac{\alpha}{2+\alpha}\log\lambda \geq 5 + 3\log\Delta$, where the final inequality follows from the assumed lower bound on λ. □

We now proceed to prove the main result of this section, which we restate for convenience.

Theorem 6.4.2. *Let α be a real number, let $\Delta \geq 3$ be an integer, and let $\lambda \geq (6\Delta)^{3+6/\alpha}$ be a real number. For any n-vertex bipartite α-expander graph with maximum degree at most Δ and any $\varepsilon \in (0,1)$ and $i \in \{0,1\}$, with $M = O(\log(n/\varepsilon))$, the Glauber dynamics restricted to $\Omega^i_{G,M}$ has mixing time $T_{mix}(\varepsilon)$ polynomial in n and $1/\varepsilon$.*

Proof. Consider an n-vertex graph $G \in \mathcal{G}_{\Delta,\alpha}$ and let $\mu^i_{G,M}$ denote the Gibbs distribution of the polymer model $(\mathcal{C}^i_{G,M}, w^i_G, G)$. By Lemma 6.4.3, we have that the family of polymer models $\{(\mathcal{C}^i_{G,M}, w^i_G, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$ satisfies the polymer sampling condition with consatnt $\tau = \frac{\alpha}{2+\alpha}\log\lambda$, and hence so does the family of truncated polymer models $\{(\mathcal{C}^i_{G,M}, w^i_G, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$. The result therefore follows by applying Theorem 6.3.2, after observing that

1. the polymer model $(\mathcal{C}^i_{G,M}, w^i_G, G)$ is single-update-compatible (use DFS ordering), and
2. for a pair of polymer configurations $\Gamma, \Gamma' \in \Omega^i_{G,M}$ whose corresponding independent sets I, I' differ in at most one vertex, we have

$$\frac{\mu^i_{G,M}(\Gamma)}{\mu^i_{G,M}(\Gamma')} = \frac{\lambda^{|I|}}{\lambda^{|I'|}} \in [1/\lambda, \lambda].$$

□

Finally, we justify that, for large enough λ and $M = \Theta(\log(n/\varepsilon))$, the aggregate weight of independent sets in $\Omega^0_{G,M} \cup \Omega^1_{G,M}$ captures all but ε fraction of the hard-core partition function $Z_{G,\lambda}$. Let Z^i_G denote the partition function of the polymer model $(\mathcal{C}^i_G, w^i_G, G)$ and let $Z^i_{G,M}$ denote the partition function of the polymer model $(\mathcal{C}^i_{G,M}, w^i_G, G)$. We will require the following lemma from [66].

Lemma 6.4.4 ([66, Lemmas 4.1 & 4.2])**.** *For $\lambda > \max\left\{(2e)^{\frac{8n}{\alpha n_0}}, (2e)^{\frac{8n}{\alpha n_1}}, (2e)^{(40/\alpha)}\right\}$, the number $\lambda^{n_0} Z^0_G + \lambda^{n_1} Z^1_G$ is a $(2e^{-n})$-approximation to the hard-core partition function $Z_{G,\lambda}$, where $n_i = |V^i_G|$ for $i \in \{0,1\}$.*

Lemma 6.4.5. *Let $\alpha > 0$ be a real number and let $\Delta \geq 3$ be an integer. There exists a constant $C > 0$ such that for $\lambda > (6C\Delta)^{3+6/\alpha}$, the following holds for*

all n-vertex bipartite α-expander graphs $G = (V_G^0, V_G^1, E_G)$ of maximum degree at most Δ. For all $\varepsilon \in (4e^{-n}, 1)$ and $M = \frac{3(2+\alpha)\log(4n/\varepsilon)}{2\alpha \log \lambda}$, the number

$$\hat{Z}_G = \lambda^{n_0} Z_{G,M}^0 + \lambda^{n_1} Z_{G,M}^1$$

is an ε-approximation to the hard-core partition function $Z_{G,\lambda}$, where $n_i = |V_G^i|$ for $i \in \{0, 1\}$.

Proof. Observe that $n/n_0, n/n_1 \leq 3$ (using that G is an α-expander for $\alpha \in (0, 1)$, see [66] for details). Therefore, by taking C large enough, we have that for all $\lambda > (6C\Delta)^{3+6/\alpha}$, and both Lemmas 6.4.3 and 6.4.4 apply. Let $\varepsilon \in (4e^{-n}, 1)$.

By Lemma 6.4.4, we have that $\lambda^{n_0} Z_G^0 + \lambda^{n_1} Z_G^1$ is an $(\varepsilon/2)$-approximation to $Z_{G,\lambda}$. By Lemma 6.4.3, we have that, for $i \in \{0, 1\}$, the family of polymer models $\{(\mathcal{C}_G^i, w_G^i, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$ satisfies the polymer sampling condition (Definition 2.6.3) with constant $\tau = \frac{\alpha}{2+\alpha} \log \lambda$, and hence so does the family of truncated polymer models $\{(\mathcal{C}_{G,M}^i, w_G^i, G) \mid G \in \mathcal{G}_{\Delta,\alpha}\}$. It follows by Lemma 6.3.3, for $M = \frac{3(2+\alpha)\log(4n/\varepsilon)}{2\alpha \log \lambda}$, that $Z_{G,M}^i$ is an $(\varepsilon/2)$-approximation to Z_G^i. Therefore, $\hat{Z}_G$ is an ε-approximation to $Z_{G,\lambda}$. □

Chapter 7

Sampling from the low temperature Potts model on $\mathbb{Z}^d$

Recent work of Helmuth, Perkins, and Regts [59] gave the first efficient algorithms for sampling from the low temperature Potts model and the high fugacity hard-core model. They gave efficient approximation algorithms for these models defined on a finite region of the integer lattice with boundary conditions. Their work was also the first to derive efficient low temperature algorithms from tools such as polymer models, from the statistical physics literature. Follow up work of Borgs, Chayes, Helmuth, Perkins, and Tetali [11] uses similar techniques to give efficient algorithms at all temperatures, when the number of spins is sufficiently large. What these algorithms have in common, is that they appeal to the polymer model framework in conjunction with the truncated Taylor series technique of Barvinok [2, 4]. In a similar spirit to the work of Section 3.3, we propose a Markov chain based algorithm as an alternative approach to sampling from the low temperature Potts model on the integer lattice.

7.1 Preliminaries

Consider the infinite graph on vertex set $\mathbb{Z}^d$ for $d \geq 2$, where two vertices $u, v \in \mathbb{Z}^d$ are joined by an edge if $\sum_{i=1}^{d} |u_i - v_i| = 1$. We will consider the ferromagnetic Potts model (with appropriate boundary conditions) on the subgraph induced by a finite subset of $\mathbb{Z}^d$.

7.1.1 Ferromagnetic Potts model with boundary conditions

For $u, v \in \mathbb{Z}^d$, we let $d_\infty(u,v) = \max_{i=1}^d |u_i - v_i|$. For $u \in \mathbb{Z}^d$ and $\Lambda \subset \mathbb{Z}^d$, we let $d_\infty(u,\Lambda) = \min_{v\in\Lambda} d_\infty(u,v)$ and we let $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$.

Let $\Lambda \subset \mathbb{Z}^d$ be finite. Let $\beta > 0$, let $q \geq 2$, and let $[q] = \{1, ..., q\}$ be a set of spins (or colours). For $i \in [q]$, we have the following set of assignments with monochromatic boundary conditions:

$$\Omega^i_\Lambda = \{\omega : \Lambda \to [q] \mid \omega(v) = i \text{ for all } v \text{ s.t } d_\infty(v, \Lambda^c) \leq 2\}.$$

The partition function of the Potts model is defined by

$$Z^i_{\Lambda,\beta} = \sum_{\omega\in\Omega^i_\Lambda} e^{-\beta m(\Lambda,\omega)}.$$

The Gibbs distribution $\mu^i_{\Lambda,\beta}$ is the probability distribution on Ω^i_Λ defined by

$$\mu^i_{\Lambda,\beta}(\omega) = \frac{e^{-\beta m(\Lambda,\omega)}}{Z^i_{\Lambda,\beta}},$$

where $m(\Lambda,\omega)$ is the number of bichromatic edges of between vertices of Λ, with respect to ω.

7.1.2 A contour model for the ferromagnetic Potts model

The following definitions and notation are based on [59, Section 3], however contour models have also been studied in [13] and [35]. We will consider the contour model for the ferromagnetic Potts model with padded monochromatic boundary conditions that was introduced in [59, Example 5]. For completeness, we also describe it here in full detail. For the Potts model, ground state configurations are the q monochromatic configurations, which we identify using the set of spins $[q]$. In general, for contour models, ground states are periodic spin configurations of maximum weight.

In general, a contour is a pair $\gamma = (V_\gamma, \omega_\gamma)$ consisting of a finite d_∞-connected vertex set $V_\gamma \subset \mathbb{Z}^d$ and a spin assignment $\omega_\gamma : V_\gamma \to [q]$. The vertex set of a contour γ partitions $\mathbb{Z}^d \setminus V_\gamma$ into a set of maximal connected components $\{A_0, A_1, ..., A_t\}$, where A_0 is infinite and $A_1, ..., A_t$ are finite. Let $\text{int}(\gamma) =$

$\bigcup_{i=1}^{t} A_i$ denote the interior of γ and let $\text{ext}(\gamma) = A_0$ denote the exterior of γ. A contour model consists of a set of contours $\mathcal{C}$, a surface energy $\|\gamma\| \in \mathbb{N}$ for each contour $\gamma \in \mathcal{C}$, and a labelling function $\text{lab}_\gamma : \{A_0, A_1, ..., A_t\} \to [q]$ for each contour $\gamma \in \mathcal{C}$.

Two contours $\gamma, \gamma' \in \mathcal{C}$ are compatible (which we write as $\gamma \sim \gamma'$) if $d_\infty(V_\gamma, V_{\gamma'}) > 1$. A set of contours $\Gamma \subseteq \mathcal{C}$ is compatible if all of its contours are pairwise compatible. A contour $\gamma \in \mathcal{C}$ is of type $\varphi \in [q]$ if its exterior is labelled with φ by lab_γ and we let

$$\text{int}_\varphi(\gamma) = \bigcup_{\substack{i \in \{1,...,t\}: \\ \text{lab}_\gamma(A_i) = \varphi}} A_i.$$

Given a set of compatible contours $\Gamma \subseteq \mathcal{C}$, a contour $\gamma \in \mathcal{C}$ is external with respect to Γ if $V_\gamma \subset \text{ext}(\gamma')$ for all $\gamma' \in \Gamma$ such that $\gamma' \neq \gamma$. The set Γ is matching and of type $\varphi \in [q]$ if all of its external contours are of type φ and either

1. $|\Gamma| = 1$, or
2. for each external contour $\gamma \in \Gamma$ and each ground state $\varphi' \in [q]$, the compatible set of contours $\{\gamma' \in \Gamma : V_{\gamma'} \subset \text{int}_{\varphi'}(\gamma)\}$ is matching and of type φ'.

Let $\mathcal{C}^\varphi \subset \mathcal{C}$ be the set of all contours of type φ. For a finite subset $\Lambda \subset \mathbb{Z}^d$, a contour $\gamma \in \mathcal{C}$ is said to be in Λ if $d_\infty(V_\gamma, \Lambda^c) > 1$. Let $\mathcal{C}_\Lambda$ be the set of all contours of $\mathcal{C}$ that are in Λ and let $\mathcal{C}^\varphi_\Lambda$ be the set of contours of $\mathcal{C}^\varphi$ that are in Λ. Finally, let $\mathcal{G}^\varphi_{\text{match}}(\Lambda)$ be the set of all sets of contours in Λ that are matching and of type φ. The partition function of a contour model is defined as follows

$$Z^\varphi_{\Lambda,z} = \sum_{\Gamma \in \mathcal{G}^\varphi_{\text{match}}(\Lambda)} \prod_{\gamma \in \Gamma} z^{\|\gamma\|}.$$

We also have the following probability distribution $\mu^\varphi_{\Lambda,z}$ on $\mathcal{G}^\varphi_{\text{match}}(\Lambda)$ defined by

$$\mu^\varphi_{\Lambda,z}(\Gamma) = \frac{\prod_{\gamma \in \Gamma} z^{\|\gamma\|}}{Z^\varphi_{\Lambda,z}}.$$

We now define the set of contours that we will use to capture the ferromagnetic Potts model on $\mathbb{Z}^d$. Let $i \in [q]$ and let $\Lambda \subset \mathbb{Z}^d$ be finite. A vertex $v \in \Lambda$

is correct with respect to $\omega \in \Omega^i_\Lambda$ if it and each of its d_∞-neighbours in Λ are mapped to the same colour by ω. All other vertices are incorrect with respect to ω. Let $\Gamma_\Lambda(\omega)$ be the set of all incorrect vertices of Λ with respect to ω. The contours of ω are the pairs $(V^1_\gamma, \omega|_{V^1_\gamma}), ..., (V^k_\gamma, \omega|_{V^k_\gamma})$, where $V^1_\gamma, ..., V^k_\gamma$ are the d_∞-connected components of $\Gamma(\omega)$ – note that each of these contours is in Λ. It is shown in [35], for each of these contours γ and each maximal d_∞-connected component A of $\mathbb{Z}^d \setminus V_\gamma$, that the set of vertices $v \in A$ such that $d_\infty(v, V_\gamma) = 1$ is d_∞-connected; thus, there is a colour $j \in [q]$ such that $\omega(v) = j$ for all such vertices. This defines the labelling function; that is, $\mathrm{lab}_\gamma(A) = j$. Note that the set of contours of ω is matching and of type i. The set $\mathcal{C}_\Lambda$ is the union, over all $\omega \in \Omega^i_\Lambda$, of the contours of ω. Finally, we define the surface energy of a contour $\gamma \in \mathcal{C}_\Lambda$ to be $m(V_\gamma, \omega_\gamma)$.

It is shown in [59] that Ω^i_Λ is in bijection with $\mathcal{G}^i_{\mathrm{match}}(\Lambda)$, therefore it follows that

$$Z^i_{\Lambda,z} = \sum_{\Gamma \in \mathcal{G}^i_{\mathrm{match}}(\Lambda)} \prod_{\gamma \in \Gamma} e^{-\beta m(V_\gamma, \omega_\gamma)} = Z^i_{\Lambda,\beta}, \tag{7.1}$$

where $z = e^{-\beta}$. We also have the following probability distribution $\mu^i_{\Lambda,z}$ on $\mathcal{G}^i_{\mathrm{match}}(\Lambda)$ defined by

$$\mu^i_{\Lambda,z}(\Gamma) = \frac{\prod_{\gamma \in \Gamma} z^{\|\gamma\|}}{Z^i_{\Lambda,z}},$$

which is equal to the probability of the corresponding Potts configuration in $\mu^i_{\Lambda,\beta}$.

7.2 Sampling from the Potts model

Let $\Lambda \subset \mathbb{Z}^d$ be finite. Let $\beta > 0$, let $q \geq 2$, let $[q] = \{1, ..., q\}$ be a set of spins (or colours), and let $i \in [q]$. We will show, for β large enough, that there is an efficient algorithm for approximately sampling from the Gibbs distribution of of the ferromagnetic Potts model with padded monochromatic boundary conditions. Consider the contour model from Section 7.1.2. The algorithm is based on a Markov chain that is rapidly mixing if the following condition is met: for all $\gamma \in \mathcal{C}_\Lambda$

$$\sum_{\gamma' \nsim \gamma} |V_{\gamma'}| e^{-\beta m(V'_\gamma, \omega_{\gamma'})} \leq \theta |V_\gamma|, \tag{7.2}$$

for some $\theta \in (0,1)$. We will then show, for β large enough, that Condition (7.2) is met and that a single update of this Markov chain can be carried out in expected linear time.

For $v \in \Lambda$, let $\mathcal{A}(v)$ be the set of contours containing v. Let ν_v be the probability distribution on $\mathcal{A}(v) \cup \emptyset$ defined by

$$\nu_v(\gamma) = \frac{e^{-\beta m(V_\gamma, \omega_\gamma)}}{q} \qquad \text{and} \qquad \nu_v(\emptyset) = 1 - \sum_{\gamma \in \mathcal{A}(v)} \nu_v(\gamma).$$

This defines a probability distribution since $\sum_{\gamma \in \mathcal{A}(v)} \nu_v(\gamma) < 1$. To see this, let γ' be the smallest contour that contains v. Applying (7.2) to this contour yields

$$\sum_{\gamma \in \mathcal{A}(v)} \nu_v(\gamma) \leq \sum_{\gamma'' \nsim \gamma'} \frac{|V_{\gamma''}|}{|V_{\gamma'}|} e^{-\beta m(V_{\gamma''}, \omega_{\gamma''})} \leq \theta < 1.$$

7.2.1 Contour Markov chain

Let $\Gamma \subset \mathcal{C}_\Lambda$ be a compatible set of contours and consider $\gamma \in \Gamma$. The children of γ in Γ are the contours $\gamma' \in \Gamma$ such that $V_{\gamma'} \subset \text{int}(\gamma)$ and there is no $\gamma'' \in \Gamma$ such that $V_{\gamma''} \subset \text{int}(\gamma)$ and $V_{\gamma'} \subset \text{int}(\gamma'')$. The parent of γ in Γ is the unique contour $\gamma' \in \Gamma$ such that γ is a child of γ'. If γ is external in Γ then it has no parent. Consider $\gamma \in \mathcal{C}_\Lambda$. Let $\partial_{\text{ext}}(\gamma)$ be the set of vertices $v \in V_\gamma$ such that $d_\infty(v, \text{ext}(\gamma)) = 1$. Let $\partial^j_{\text{in}}(\gamma)$ be the set of vertices $v \in V_\gamma$ such that $d_\infty(v, \text{int}_j(\gamma)) = 1$ and let $\partial_{\text{in}}(\gamma)$ be the set of vertices $v \in V_\gamma$ such that $d_\infty(v, \text{int}(\gamma)) = 1$. Recall from Section 7.1.2 that $\partial_{\text{ext}}(\gamma)$ is coloured with a single colour, and $\partial^j_{\text{in}}(\gamma)$ is coloured with j.

Consider the following Markov chain $\mathcal{M}^i_{\Lambda,\beta}$ on state space $\mathcal{G}^i_{\text{match}}(\Lambda)$ which transitions from Γ_t to Γ_{t+1} according to the following rules.

1. Choose $v \in \Lambda$ uniformly at random and let γ_v be the unique contour in Γ_t that contains v, if such a contour exists; otherwise, let $\gamma_v = \emptyset$.

2. Mutually exclusively, each with probability $1/2$, do the following.

   (a) Let j be the colour of $\partial_{\text{ext}}(\gamma_v)$ and let $\Gamma = \{\gamma \in \Gamma_t : V_\gamma \subset \text{int}(\gamma_v)\}$. Let $\Gamma_{t+1} = ((\Gamma_t \setminus \Gamma) \setminus \{\gamma_v\}) \cup \textbf{Recolour}(\Gamma, j)$.

(b) Sample $\boldsymbol{\gamma}$ from ν_v and let $\Gamma_{t+1} = \Gamma_t$ if $\Gamma_t \cup \{\boldsymbol{\gamma}\}$ is not compatible. If $\boldsymbol{\gamma}$ is external in $\Gamma_t \cup \{\boldsymbol{\gamma}\}$ let $j = i$; otherwise, let γ' be the parent of $\boldsymbol{\gamma}$ in $\Gamma_t \cup \{\boldsymbol{\gamma}\}$ and let $j \in [q]$ be such that $\boldsymbol{\gamma} \in \mathrm{int}_j(\gamma')$. Let $\Gamma = \{\gamma \in \Gamma_t : V_\gamma \subset \mathrm{int}(\boldsymbol{\gamma})\} \cup \{\boldsymbol{\gamma}\}$ and let $\Gamma_{t+1} = (\Gamma_t \setminus \Gamma) \cup \mathbf{Recolour}(\Gamma, j)$.

The recolouring procedure takes as input a colour $j \in [q]$ and a compatible contour configuration Γ, such that either

- $\Gamma \in \mathcal{G}^{j'}_{\mathrm{match}}(\Lambda)$ for some $j' \in [q]$, or
- $\Gamma = \Gamma'' \cup \{\gamma\}$ where $\Gamma'' \in \mathcal{G}^{j'}_{\mathrm{match}}(\Lambda)$ for some $j' \in [q]$ and γ is external in Γ.

It produces as output a contour configuration $\Gamma' \in \mathcal{G}^{j}_{\mathrm{match}}(\Lambda)$, according to the following deterministic algorithm.

Algorithm 6: Recursively recolouring a contour configuration.

```
Recolour (Γ, j)
    Γ' = ∅;
    for γ ∈ Γ external do
        γ = RecolourContour(γ, j);
        Γ' = Γ' ∪ {γ};
        for j' ∈ [q] do
            Γ_{j'} = {γ' ∈ Γ : V_{γ'} ⊂ int_{j'}(γ)};
            Γ' = Γ' ∪ Recolour(Γ_{j'}, j');
    return Γ';
```

In order to recolour an individual contour, the above algorithm makes a call to the following deterministic algorithm. It takes as input a contour $\gamma \in \mathcal{C}_\Lambda$ and a colour $j \in [q]$, and produces as output a contour $\gamma' \in \mathcal{C}_\Lambda$ on the same vertex subset, such that $\partial_{\mathrm{ext}}(\gamma')$ is coloured with j. Note that $m(V_\gamma, \omega_\gamma) = m(V_{\gamma'}, \omega_{\gamma'})$.

Algorithm 7: Recolouring an individual contour.

```
RecolourContour (γ, j)
    ω_{γ'} = ω_γ;
    j' = colour of ∂_ext(γ) under ω_γ;
    for v ∈ V_γ do
        if ω_γ(v) = j then
            ω_{γ'}(v) = j';
        else if ω_γ(v) = j' then
            ω_{γ'}(v) = j;
    return (V_γ, ω_{γ'});
```

The advantage of recolouring is that an update of $\mathcal{M}^i_{\Lambda,\beta}$ is only ever rejected if we attempt to add a contour that is incompatible with a contour in the current configuration. That is, a move is never rejected because the colouring of the configuration that we arrive at is not matching (because recolouring ensures it is).

The following lemma shows that the unique stationary distribution of the Markov chain is the Gibbs distribution of the contour model.

Lemma 7.2.1. *The unique stationary distribution of $\mathcal{M}^i_{\Lambda,\beta}$ is $\mu^i_{\Lambda,\beta}$.*

Proof. To see that $\mu^i_{\Lambda,\beta}$ is a stationary distribution of $\mathcal{M}^i_{\Lambda,\beta}$ we check detailed balance. Suppose that $\Gamma, \Gamma' \in \mathcal{G}^i_{\text{match}}(\Lambda)$ are such that $\Gamma \neq \Gamma'$ and $\Pr(\Gamma, \Gamma') > 0$, then we may assume that

1. $\{V_{\gamma'} : \gamma' \in \Gamma'\} = \{V_{\gamma'} : \gamma' \in \Gamma\} \cup \{V_\gamma\}$ for some contour $\gamma \in \mathcal{C}_\Lambda$,
2. $\{\gamma' \in \Gamma : V_{\gamma'} \subset \text{ext}(\gamma)\} = \{\gamma' \in \Gamma : V_{\gamma'} \subset \text{ext}(\gamma)\}$, and
3. $\prod_{\substack{\gamma' \in \Gamma: \\ V_{\gamma'} \subset \text{int}(\gamma)}} e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})} = \prod_{\substack{\gamma' \in \Gamma': \\ V_{\gamma'} \subset \text{int}(\gamma)}} e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})}$.

To arrive at Γ' from Γ, our uniform vertex choice must be some $v \in V_\gamma$ and then we must sample a contour γ_v which when recoloured by Algorithm 7 gives γ. After these choices are made, the remainder of the recolouring process is deterministic and so we are guaranteed to arrive at Γ', since $\Pr(\Gamma, \Gamma') > 0$. The probability of making this choice of vertex v is equal to $|V_\gamma|/|\Lambda|$. Conditioned on this choice of vertex, the probability of sampling a contour γ_v which, when

recoloured by Algorithm 7, gives γ is $q \cdot \nu_v(\gamma)$. This is because there are precisely q such contours, and they all have the same weight since the recolouring procedure preserves the number of bichromatic edges. We therefore have that $\Pr(\Gamma, \Gamma') = q \cdot \nu_v(\gamma)|V_\gamma|/2|\Lambda| = e^{-\beta m(V_\gamma, \omega_\gamma)}|V_\gamma|/2|\Lambda|$. On the other hand, if we are at configuration Γ' and our uniformly chosen vertex is a vertex of V_γ, then we arrive back at Γ after removing γ. It therefore follows that $\Pr(\Gamma', \Gamma) = |V_\gamma|/2|\Lambda|$. By item 3, we have that

$$\frac{\mu^i_{\Lambda,\beta}(\Gamma')}{\mu^i_{\Lambda,\beta}(\Gamma)} = e^{-\beta m(V_\gamma, \omega_\gamma)} = \frac{e^{-\beta m(V_\gamma, \omega_\gamma)}|V_\gamma|/2|\Lambda|}{|V_\gamma|/2|\Lambda|} = \frac{q \cdot \nu_v(\gamma)|V_\gamma|/2|\Lambda|}{|V_\gamma|/2|\Lambda|} = \frac{\Pr(\Gamma, \Gamma')}{\Pr(\Gamma', \Gamma)}.$$

Beginning with $\Gamma, \Gamma' \in \mathcal{G}^i_{\text{match}}(\Lambda)$ such that $\Pr(\Gamma', \Gamma) > 0$, a similar argument shows that $\Pr(\Gamma, \Gamma') > 0$ and that detailed balance is also satisfied.

Finally, $\mathcal{M}^i_{\Lambda,\beta}$ is aperiodic since it has self-loops and it is irreducible since we can move from any $\Gamma \in \mathcal{G}^i_{\text{match}}(\Lambda)$ to any $\Gamma' \in \mathcal{G}^i_{\text{match}}(\Lambda)$; for example, by removing external contours until we reach the empty set, then adding external contours until we reach Γ'. □

The following lemma shows that when Condition (7.2) is met, the Markov chain mixes rapidly.

Lemma 7.2.2. *If Condition* (7.2) *holds then* $\mathcal{M}^i_{\Lambda,\beta}$ *has mixing time* $T_{mix}(\varepsilon) = O(|\Lambda| \log(|\Lambda|/\varepsilon))$.

Proof. We proceed by path coupling. Let $D(\cdot, \cdot)$ be a metric on $\mathcal{G}^i_{\text{match}}(\Lambda)$ which we define by setting $D(\Gamma, \Gamma') = 1$ for contour configurations $\Gamma, \Gamma' \in \mathcal{G}^i_{\text{match}}(\Lambda)$ such that $\Gamma \triangle \Gamma' = \{\gamma\}$ for some contour $\gamma \in \mathcal{C}_\Lambda$. This can be extended to a shortest path metric – the symmetric difference.

Let (X_t, Y_t) be a coupling for $\mathcal{M}^i_{\Lambda,\beta}$ where both chains make the same random choices. Suppose that $D(X_t, Y_t) = 1$ and that $X_t = Y_t \cup \{\gamma\}$ for some contour $\gamma \in \mathcal{C}_\Lambda$. We obtain (X_{t+1}, Y_{t+1}) from (X_t, Y_t) and examine the various possibilities for $D(X_{t+1}, Y_{t+1})$ as follows.

1. $D(X_{t+1}, Y_{t+1}) = 0$ if we remove γ from both configurations. This occurs with probability $|V_\gamma|/2|\Lambda|$.

2. $D(X_{t+1}, Y_{t+1}) \leq 2$ if we attempt to add a contour $\gamma' \in \mathcal{C}_\Lambda$ to both configurations, and the move is accepted in Y_t but rejected in X_t due to the presence of γ. Since updates are never rejected due to colour conflicts, this happens precisely when $d_\infty(V_\gamma, V_{\gamma'}) \leq 1$. The probability that this occurs is therefore at most

$$\frac{1}{2|\Lambda|} \sum_{\gamma' \nsim \gamma} |V_{\gamma'}| e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})}.$$

Combining these two cases, we obtain that

$$\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq 1 + \frac{1}{2|\Lambda|} \left(-|V_\gamma| + \sum_{\gamma' \nsim \gamma} |V_{\gamma'}| e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})} \right),$$

and therefore by Condition (7.2) that

$$\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq 1 - \frac{|V_\gamma|(1-\theta)}{2|\Lambda|} \leq 1 - \frac{1-\theta}{2|\Lambda|}.$$

Let W be the diameter of $\mathcal{G}^i_{\text{match}}(\Lambda)$ with respect to the metric $D(\cdot, \cdot)$ and note that $W \leq 2|\Lambda|$. By the path coupling technique of [30, Section 6], it follows that the mixing time of $\mathcal{M}^i_{\Lambda,\beta}$ is at most $2|\Lambda| \log(W/\varepsilon)/(1-\theta)$. Since θ is a constant strictly less than 1, the mixing time is $O(|\Lambda| \log(|\Lambda|/\varepsilon))$. □

What remains to be shown is that we can efficiently perform a single update of $\mathcal{M}^i_{\Lambda,\beta}$. The next result shows that this is possible subject to the following condition. There exists $\tau \geq 5 + 3\log(q \cdot 3^d)$ such that all $\gamma \in \mathcal{C}_\Lambda$ satisfy

$$\frac{e^{-\beta m(V_\gamma, \omega_\gamma)}}{q} \leq e^{-\tau |V_\gamma|}. \tag{7.3}$$

Lemma 7.2.3. *If Condition* (7.3) *holds, then the expected time taken to perform a single update of* $\mathcal{M}^i_{\Lambda,\beta}$ *is* $O(|\Lambda|)$.

Proof. The bottleneck of performing an update is the recolouring procedure. We can execute a call to **Recolour**(Γ, j) in $O(|\Lambda|)$ time, for any compatible set of contours Γ and any colour $j \in [q]$.

In order to sample a contour from ν_v, we appeal to the 'single polymer sampler' of [22, Section 2.3]. With only minor modification, this algorithm can be used to sample a contour from ν_v in expected constant time. Condition (7.3) is analogous to the polymer sampling condition of [22], except that

1. the maximum degree of the 'host graph' is 3^d in this setting[1], and
2. $q-1$ is replaced by q since a contour can be coloured with q colours whereas a polymer can be coloured with at most $q-1$ (see [22, Section 2] for more details).

To apply the single polymer sampler in the setting of [22], the polymer model also needs to be 'computationally feasible'. A polymer model is computationally feasible if it is possible to determine whether a connected (in the host graph) set of vertices and a spin assignment to it is a polymer, and compute its weight if it is, in time polynomial in the size of the polymer. In our setting, this translates to being able to determine whether a d_∞-connected set of vertices and a spin assignment to it is a contour $\gamma \in \mathcal{C}_\Lambda$, and compute $e^{-\beta m(V_\gamma,\omega_\gamma)}/q$ in $\text{poly}(|V_\gamma|)$ time, if it is. We can do this in $O(|V_\gamma|)$ time by checking that every vertex is incorrect, that the interior and exterior boundaries of the vertex set are appropriately coloured, and counting the bichromatic edges. □

Finally, we have the following result.

Theorem 7.2.4. *For all integers $d \geq 2$, all integers $q \geq 2$, all $i \in [q]$, and all $\beta \geq 2 \cdot 3^d(5 + 3\log(q \cdot 3^d))$, there is an algorithm which, given $\Lambda \subset \mathbb{Z}^d$ and $\varepsilon \in (0,1)$ as input, produces an ε-approximate sample from $\mu^i_{\Lambda,\beta}$. Its running time is $O(|\Lambda|^2 \log(|\Lambda|/\varepsilon)\log(1/\varepsilon))$.*

Proof. Subject to Conditions (7.2) and (7.3), Lemmas 7.2.1, 7.2.2, and 7.2.3 give an algorithm which produces an ε-approximate sample from $\mu^i_{\Lambda,\beta}$, and has an expected running time of $O(|\Lambda|^2 \log(|\Lambda|/\varepsilon))$. By terminating this algorithm and returning the empty configuration after $O(|\Lambda|^2 \log(|\Lambda|/\varepsilon))$ steps, it is shown in the proof of Theorem 3.2.2 how to give an upper bound of $O(|\Lambda|^2 \log(|\Lambda|/\varepsilon)\log(1/\varepsilon))$ on the worst-case running time of the algorithm.

We now show that Conditions (7.2) and (7.3) are met. It is shown in [59, Example 5], that $m(V_\gamma, w_\gamma) \geq |V_\gamma|/(2 \cdot 3^d)$ for all $\gamma \in \mathcal{C}_\Lambda$. Thus,

$$e^{-\beta m(V_\gamma,w_\gamma)} \leq e^{-\beta|V_\gamma|/(2\cdot 3^d)} \leq e^{-\tau|V_\gamma|}, \tag{7.4}$$

for $\tau = 5 + 3\log(q \cdot 3^d)$, and Condition (7.3) holds.

[1]To see this, fix a point in u in $\mathbb{Z}^d$ and consider all of the other points w in $\mathbb{Z}^d$ for which $d_\infty(u,w) = \max_{i=1}^d |u_i - w_i| = 1$.

Finally, by a calculation similar to the one given in [22, Section 2.1], we show that Condition (7.2) also holds. For all $\gamma \in \mathcal{C}_\Lambda$, we have that

$$\sum_{\gamma' \nsim \gamma} |V_{\gamma'}| e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})} \leq \sum_{\gamma' \nsim \gamma} e^{|V_{\gamma'}|} e^{-\beta m(V_{\gamma'}, \omega_{\gamma'})} \leq \sum_{\substack{v \in \Lambda: \\ d_\infty(v, V_\gamma) \leq 1}} \sum_{k \geq 1} \sum_{\substack{\gamma' \in \mathcal{C}_\Lambda: \\ |V'_\gamma| = k, v \in V'_\gamma}} e^k e^{-\tau k},$$

where the final inequality follows from (7.4). Lemma 2.1 of [12] implies that there are at most $(e \cdot 3^d)^{k-1}$ d_∞-connected vertex subsets of size k containing a fixed vertex, which gives us an upper bound of $(e \cdot 3^d)^{k-1} q^k$ on the number of contours of size k containing a fixed vertex. Thus,

$$\begin{aligned}
\sum_{\gamma' \nsim \gamma} |V'_\gamma| e^{-\beta m(V'_\gamma, \omega_{\gamma'})} &\leq |V_\gamma| (3^d + 1) \sum_{k \geq 1} (e \cdot 3^d)^{k-1} q^k e^k e^{-\tau k} \\
&\leq \frac{|V_\gamma| (3^d + 1)}{e \cdot 3^d} \sum_{k \geq 1} e^{-3k} \\
&\leq |V_\gamma|,
\end{aligned}$$

where the penultimate inequality follows from the fact that $\tau = 5 + 3 \log(q \cdot 3^d)$. □

Chapter 8

Conclusions and open problems

In this thesis, we have developed novel algorithmic approaches to sampling from low temperature spin systems, with applications in a variety of different settings. The starting point for the work of this thesis is the observation that, at low temperatures, the mixing times of single-update Markov chains such as the Glauber dynamics undergo an exponential slowdown on bounded degree graphs, which rules out their straightforward application as sampling algorithms for low temperature spin systems. In spite of the fact that this phenomenon persists on random regular graphs, we have shown how Markov chains can still be used as an effective tool for sampling from low temperature spin systems, by presenting the polymer dynamics of Chapter 3 – a simple and efficient algorithm for sampling from the Gibbs distribution of a polymer model. With this new approach, we saw how to obtain efficient algorithms for the low temperature Potts model and the high-fugacity hard-core model, on random regular and bipartite random regular graphs, respectively. Whilst doing this, we significantly improved the running times of existing algorithms for the same problems.

We have also demonstrated a wide range of interesting applications of the polymer model framework and its associated sampling algorithms. In Chapter 4, we generalised the approach of Chapter 3 to any general spin system on a random regular bipartite graph. In Chapter 5, we revised the polymer model framework, and developed efficient algorithms for sampling from the low temperature Potts model on sparse random graphs of unbounded maximum degree. This is in contrast to almost all known low temperature algorithms, for which the bounded degree assumption is crucial. Finally, in Chapter 7, we saw

how similar techniques to those of Chapter 3 can be employed in order to sample from the Gibbs distribution of a contour model. As an application of this method, we obtained efficient algorithms for the low temperature Potts model on finite regions of $\mathbb{Z}^d$ with appropriate boundary conditions.

The essential aim of the program of research that this thesis is a part of, is to develop efficient algorithms for the broadest possible ranges of parameters and input instances. Whilst our results have significantly extended the range of application of low temperature algorithms, a number of interesting open problems still remain.

8.1 Improving the range of parameters

Due to the nature of the polymer approach, it is highly unlikely that the range of parameters in which existing algorithms apply, is the optimal one. For the q-colour low temperature Potts model on bounded-degree expanders, an interesting question to investigate would therefore be: for what range of Δ, q, and β can we efficiently sample from the Gibbs distribution of the Potts model? In recent work, Blanca, Cannon, and Perkins [6] make progress on this question by giving an efficient algorithm for sampling from the low temperature Potts model on bounded-degree expanders, which applies for a greater range of β that the results of, for example, [68, 74], and Chapter 3. Whilst these algorithms apply for all values of $q \geq 3$, another approach is to consider values of q sufficiently large as a function of Δ, and ask for which values of β can we efficiently sample from the Gibbs distribution of the ferromagnetic Potts model? Recent work of Helmuth, Jenssen, and Perkins answers this question for the random Δ-regular graph (which is an expander, with high probability). Similar 'all temperature' results have also been proven for finite regions of $\mathbb{Z}^d$, when q is sufficiently large with respect to d. One interested question is whether similar results to these can be proven for random graphs of bounded average degree as in Chapter 5, or for other spin models such as the hard-core model.

8.2 Low temperature algorithms on $\mathcal{G}(n, d/n)$

The work of Chapter 5 extends the polymer model framework to classes of sparse random graphs with a given degree sequence. In order to ensure strong enough expansion properties of such graphs, one condition required of the degree sequence is that the degree of every vertex should be at least 3. Unfortunately, this latter restriction rules out perhaps the most canonical example of a sparse random graph of unbounded maximum degree – the supercritical Erdős-Rényi random graph, also known as $\mathcal{G}(n, d/n)$, where $d > 1$ is a constant independent of n. The difficulty in obtaining algorithms for low temperature spin systems on $\mathcal{G}(n, d/n)$ using existing approaches, such as the polymer model framework, stems from the fact that it lacks the necessary small set expansion properties. Due to the high connectivity of $\mathcal{G}(n, d/n)$, it can indeed be shown that the 'ground state' phenomenon of, for example, the low temperature Potts model, persists on $\mathcal{G}(n, d/n)$. In spite of this, a naïve attempt to capture the deviations that configurations take from these ground states as polymers, will fail. This is due to the fact that we can no longer appeal to the expansion properties of polymers, in order to ensure that their weights decay exponentially and conditions such as the polymer mixing condition hold. For the bipartite Erdős-Rényi model, the analogous open problem remains for the high fugacity hard-core model. In general, it is unclear as to how we might exploit the ground state phenomenon in order to develop efficient low temperature algorithms, in settings where small set expansion properties do not hold.

8.3 Unrestricted Glauber dynamics within a phase

On random regular graphs, the Potts Glauber dynamics mixes exponentially slowly at low temperatures. This rules out a naïve application of the Glauber dynamics as an efficient sampling algorithm, and is indeed one of the main pieces of motivation for much of the work in this thesis. The reason for this is that, at low temperatures, the state space is effectively partitioned into a number of phases separated by bottlenecks which single-update Markov chains struggle to cross.

In Chapter 6, we saw how the Glauber dynamics can be used as the basis for an efficient sampling algorithm for the low temperature Potts model. This is achieved by initialising it in one of q ground states, and explicitly rejecting any moves which take it too far out of phase. In addition to this, we further restrict the dynamics to run only on configurations that induce polymers of at most logarithmic size. It is unclear whether these two restrictions are necessary, however, and it remains an open problem to establish the mixing time[1] of the *unrestricted* Glauber dynamics, to the Gibbs distribution of the low temperature Potts model restricted to the red portion of the state space. This is a problem that had previously been suggested in both [59] and [68].

Subsequently to the work of Chapter 6, more partial progress has been made by Gheissari and Sinclair [48]. They consider the unrestricted Ising Glauber dynamics on the random regular graph, and prove that it comes close (in total variation distance) to the Gibbs distribution of the plus phase of the Ising model, in $O(n \log n)$ steps. In order to show this, they appeal to a property they introduce as 'weak spatial mixing within a phase'. Whilst they remark that the absence of certain monotonicity properties for the Potts model rules out the extension of their methods, one model which does exhibit the required monotonicity properties is the random cluster model. One interesting possible direction for progress, would be to understand whether the analogous notion of weak spatial mixing within a phase can be established for the random cluster model, perhaps by appealing to the polymer representation of the random cluster model introduced in [58]. Recently, results of this nature have also been proven for the random cluster dynamics on finite subsets of $\mathbb{Z}^d$ [49].

[1]Strictly speaking, we are not interested in the mixing time, but rather the number of steps taken to become close (in total variation distance) to the Gibbs distribution of the red phase. One might expect the Glauber dynamics to stay here for some number of steps, before again moving away.

Bibliography

[1] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing*, (0):FOCS20–1, 2021. 4

[2] Alexander Barvinok. *Combinatorics and Complexity of Partition Functions*. Algorithms and Combinatorics. Springer International Publishing, 2017. 9, 32, 80, 125

[3] Alexander Barvinok and Guus Regts. Weighted counting of solutions to sparse systems of equations. *Combinatorics, Probability and Computing*, 28(5):696–719, 2019. 80

[4] Alexander Barvinok and Pablo Soberón. Computing the partition function for graph homomorphisms with multiplicities. *Journal of Combinatorial Theory, Series A*, 137:1–26, 2016. 125

[5] Ivona Bezáková, Daniel Štefankovič, Vijay V. Vazirani, and Eric Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM Journal on Computing*, 37(5):1429–1454, 2008. 42, 43

[6] Antonio Blanca, Sarah Cannon, and Will Perkins. Fast and Perfect Sampling of Subgraphs and Polymer Systems. In Amit Chakrabarti and Chaitanya Swamy, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, volume 245 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:18, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 35, 138

[7] Béla Bollobás. *The art of mathematics: Coffee time in Memphis*. Cambridge University Press, 2006. 85

[8] Magnus Bordewich, Catherine Greenhill, and Viresh Patel. Mixing of the Glauber dynamics for the ferromagnetic potts model. *Random Structures & Algorithms*, 48(1):21–52, 2016. 4, 6, 111

[9] Christian Borgs. Absence of zeros for the chromatic polynomial on bounded degree graphs. *Combinatorics, Probability and Computing*, 15(1-2):63–74, 2006. 7

[10] Christian Borgs, Jennifer Chayes, Alan Frieze, Jeong Han Kim, Prasad Tetali, Eric Vigoda, and Van Ha Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 218–229, 1999. 32

[11] Christian Borgs, Jennifer Chayes, Tyler Helmuth, Will Perkins, and Prasad Tetali. Efficient sampling and counting algorithms for the Potts model on $\mathbb{Z}^d$ at all temperatures. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 738–751, 2020. 80, 81, 125

[12] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. Left and right convergence of graphs with bounded degree. *Random Structures & Algorithms*, 42(1):1–28, 2013. 36, 135

[13] Christian Borgs and John Z. Imbrie. A unified approach to phase diagrams in field theory and statistical mechanics. *Communications in mathematical physics*, 123(2):305–328, 1989. 7, 33, 126

[14] Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and applications. *Combinatorics, Probability and Computing*, 31(2):229–267, 2022. 59, 61

[15] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231. IEEE, 1997. 21

[16] Andrei Bulatov and Martin Grohe. The complexity of partition functions. *Theoretical Computer Science*, 348(2):148–186, 2005. 59

[17] Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, Mark Jerrum, Daneil Štefankovič, and Eric Vigoda. #BIS-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *Journal of Computer and System Sciences*, 82(5):690–711, 2016. 60

[18] Sarah Cannon and Will Perkins. Counting independent sets in unbalanced bipartite graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1456–1466, 2020. 80, 81

[19] Charles Carlson, Ewan Davies, and Alexandra Kolla. Efficient algorithms for the Potts model on small-set expanders. *arXiv preprint arXiv:2003.01154*, 2020. 80, 81

[20] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. *arXiv preprint arXiv:2203.07771*, 2022. 4

[21] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 7, 31, 109

[22] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. *Random Struct. Algorithms*, 58(2):294–321, 2021. Theorems 5 and 6 from arxiv.org/abs/1901.0665. 7, 31, 58, 80, 81, 93, 109, 133, 134, 135

[23] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1537–1550, 2021. 4

[24] Stefaan De Winter, Jeroen Schillewaert, and Jacques Verstraete. Large incidence-free sets in geometries. *The electronic journal of Combinatorics*, 19(4):P24, 2012. 24

[25] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993. 113

[26] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993. 113

[27] Roland L. Dobrushin. Estimates of semi-invariants for the Ising model at low temperatures. *Translations of the American Mathematical Society-Series 2*, 177:59–82, 1996. 33

[28] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004. 67

[29] Martin Dyer, Leslie Ann Goldberg, Mark Jerrum, and Russell Martin. Markov chain comparison. *Probab. Surveys*, 3:89–111, 2006. 113, 115

[30] Martin Dyer and Catherine Greenhill. Random walks on combinatorial objects. *London Mathematical Society Lecture Note Series*, pages 101–136, 1999. 90, 133

[31] Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Structures & Algorithms*, 17(3-4):260–289, 2000. 59

[32] Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000. 32

[33] Charilaos Efthymiou, Thomas P. Hayes, Daniel Štefankovič, Eric Vigoda, and Yitong Yin. Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 704–713, 2016. 32

[34] Nikolaos Fountoulakis and Bruce A. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures & Algorithms*, 33(1):68–86, 2008. 99, 101, 102, 103

[35] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017. 126, 128

[36] Joel Friedman. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724, 2003. 27

[37] Tobias Friedrich, Andreas Göbel, Martin S Krejca, and Marcus Pappik. A spectral independence view on hard spheres via block dynamics. *SIAM Journal on Discrete Mathematics*, 36(3):2282–2322, 2022. 81

[38] Andreas Galanis, Leslie Goldberg, and Mark Jerrum. Approximately counting H-colorings is #BIS-hard. *SIAM Journal on Computing*, 45(3):680–711, 2016. 59, 67

[39] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. In Javier Esparza and Daniel Kráľ, editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. 7, 57

[40] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast algorithms for general spin systems on bipartite expanders. *ACM Transactions on Computation Theory (TOCT)*, 13(4):1–18, 2021. 7, 57, 80, 81

[41] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. In Mary Wootters and Laura Sanità, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, volume 207 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:13, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 8, 79

[42] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. Fast mixing via polymers for random graphs with unbounded degree. *Information and Computation*, page 104894, 2022. 8, 79

[43] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016. 3, 5

[44] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures & Algorithms*, 54(1):82–147, 2019. 110

[45] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6), 2015. 60

[46] Andreas Galanis, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016. 3, 6, 59, 63

[47] David Galvin and Prasad Tetali. Slow mixing of Glauber dynamics for the hard-core model on regular bipartite graphs. *Random Structures & Algorithms*, 28(4):427–443, 2006. 32, 110

[48] Reza Gheissari and Alistair Sinclair. Low-temperature Ising dynamics with random initializations. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1445–1458, 2022. 140

[49] Reza Gheissari and Alistair Sinclair. Spatial mixing and the random-cluster dynamics on lattices. *arXiv preprint arXiv:2207.11195*, 2022. 140

[50] Roy J. Glauber. Time-dependent statistics of the Ising model. *Journal of mathematical physics*, 4(2):294–307, 1963. 1

[51] Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic Potts model. *J. ACM*, 59(5), 2012. 59

[52] Leslie Ann Goldberg and Mark Jerrum. The complexity of approximately counting tree homomorphisms. *ACM Trans. Comput. Theory*, 6(2):8:1–8:31, 2014. 59

[53] Leslie Ann Goldberg, Steven Kelk, and Mike Paterson. The complexity of choosing an H-coloring (nearly) uniformly at random. *SIAM Journal on Computing*, 33(2):416–432, 2004. 59

[54] Artem Govorov, Jin-Yi Cai, and Martin Dyer. A dichotomy for bounded degree graph homomorphisms with nonnegative weights. *Journal of Computer and System Sciences*, 132:1–15, 2023. 59

[55] Catherine Greenhill. The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. *computational complexity*, 9(1):52–72, 2000. 3

[56] C. Gruber and H. Kunz. General properties of polymer systems. *Communications in Mathematical Physics*, 22(2):133–161, 1971. 7, 27

[57] Willem H. Haemers. Interlacing eigenvalues and graphs. *Linear Algebra and its applications*, 226(228):593–616, 1995. 24

[58] Tyler Helmuth, Matthew Jenssen, and Will Perkins. Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs. *arXiv preprint arXiv:2006.11580*, 2020. 80, 81, 140

[59] Tyler Helmuth, Will Perkins, and Guus Regts. Algorithmic Pirogov–Sinai theory. *Probability Theory and Related Fields*, 176(3):851–895, 2020. 7, 8, 9, 12, 32, 35, 36, 41, 59, 80, 81, 110, 125, 126, 128, 134, 140

[60] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006. 24, 25, 59

[61] Mark Huber. Approximation algorithms for the normalizing constant of Gibbs distributions. *The Annals of Applied Probability*, 25(2):974–985, 2015. 42, 43

[62] Jeroen Huijben, Viresh Patel, and Guus Regts. Sampling from the low temperature Potts model through a Markov chain on flows. *Random Structures & Algorithms*, 2021. 80, 81, 83

[63] Ernst Ising. *Beitrag zur theorie des ferro-und paramagnetismus*. PhD thesis, Grefe & Tiedemann, 1924. 1

[64] Svante Janson. The probability that a random multigraph is simple. II. *Journal of Applied Probability*, 51(A):123–137, 2014. 23

[65] Matthew Jenssen and Peter Keevash. Homomorphisms from the torus. *arXiv preprint arXiv:2009.08315*, 2020. 81

[66] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2235–2247, 2019. 49, 50, 52, 53, 54, 59, 62, 67, 70, 121, 123, 124

[67] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *arXiv preprint*, 1807.04804v2, 2019. 52, 54

[68] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020. 6, 7, 8, 9, 10, 11, 12, 27, 32, 34, 35, 36, 48, 52, 54, 58, 80, 81, 95, 110, 121, 138, 140

[69] 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. 16

[70] Nabil Kahale. Eigenvalues and expansion of regular graphs. *J. ACM*, 42(5):1091–1106, 1995. 25

[71] Roman Koteckỳ and David Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986. 7, 36, 84

[72] Lahoussine Laanait, Alain Messager, Salvador Miracle-Solé, Jean Ruiz, and Senya Shlosman. Interfaces in the Potts model I: Pirogov-Sinai theory of the Fortuin-Kasteleyn representation. *Communications in Mathematical Physics*, 140(1):81–91, 1991. 7

[73] Wilhelm Lenz. Beitršge zum verstšndnis der magnetischen eigenschaften in festen kšrpern. *Physikalische Z*, 21:613–615, 1920. 1

[74] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. Counting independent sets and colorings on random regular bipartite graphs. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:12, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 7, 8, 9, 10, 32, 35, 48, 54, 55, 59, 80, 81, 138

[75] Jingchen Liu and Pinyan Lu. FPTAS for #BIS with degree bounds on one side. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC)*, pages 549–556, 2015. 32

[76] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017. 15, 19, 20

[77] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3):401–439, 2009. 5, 32

[78] Viresh Patel and Guus Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM Journal on Computing*, 46(6):1893–1919, 2017. 39, 91

[79] S. A. Pirogov and Ya. G. Sinai. Phase diagrams of classical lattice systems. *Teoret. Mat. Fiz.*, 25(3):358–369, 1975. 33

[80] Dana Randall and Prasad Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.*, 41(3):1598–1615, 2000. Probabilistic techniques in equilibrium and nonequilibrium statistical physics. 113

[81] Alistair Sinclair. *Algorithms for random generation and counting: a Markov chain approach*. Springer Science & Business Media, 2012. 3

[82] Alan Sly and Nike Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6):2383–2416, 11 2014. 3, 60

[83] Allan Sly. Computational transition at the uniqueness threshold. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 287–296. IEEE, 2010. 3, 5

[84] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 361–369. IEEE, 2012. 5

[85] Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *Journal of the ACM*, 56(3):18, 2009. 42, 43

[86] R. Michael Tanner. Explicit concentrators from generalized N-gons. *SIAM Journal on Algebraic Discrete Methods*, 5(3):287–293, 1984. 25

[87] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM Journal on Computing*, 20(5):865–877, 1991. 14

[88] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical computer science*, 8(2):189–201, 1979. 14

[89] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge university press, 2016. 23

[90] Eric Vigoda. A note on the Glauber dynamics for sampling independent sets. *the electronic journal of combinatorics*, 8(1):R8, 2001. 4

[91] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149, 2006. 3, 4, 32