Approximation Algorithms and Inapproximability of Partition Functions of Spin Systems



Kuan Yang
St. Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2019

This thesis is dedicated to
my parents
for their encouragement, support and love.

Acknowledgements

I would like to start by expressing my deepest gratitude to my DPhil supervisors, Leslie Ann Goldberg and Andreas Galanis. Without their patient assistance and guidance during the last four years, it is not possible for me to finish this thesis. I also benefited greatly from their energy and enthusiasm for work and life, and their rigorous attitudes to research. I am grateful and feel fortunate to work with them and have them as my supervisors.

I would also like to thank my undergraduate advisor Pinyan Lu, for leading me into the area of theoretical computer science, especially my current research area of approximate counting, and hosting me as an intern at Microsoft Research Asia during 2014-2015. I enjoyed a wonderful time and collaboration working with Pinyan. In addition, I am also grateful to other brilliant collaborators: Heng Guo, Chihao Zhang and Minshen Zhu. I am pleased to have the chance to doing research with them, and some part of this thesis is based on our joint work. Besides, I must thank Jiawei Gao, Chao Liao, Yupan Liu and Jiapeng Zhang, for our inspiring discussion on different areas.

Moreover, I also want to thank the Clarendon Fund and its primary sponsor, Oxford University Press, for offering me the fully-funded scholarship during my DPhil. I feel very fortunate and honoured to be awarded by such a highly competitive scholarship. Another welfare of being a Clarendon Scholar is the membership of the Clarendon Scholars' Association, which provides many opportunities in leisure time activities and social events. I am also grateful to meet many friends in the Clarendon Scholar Association: Chon Lok Lei, Evan Hann, Binxin Ru, Qiujie Shi and Wilfred Ngia.

The algorithm and complexity theory group at Oxford has provided me with extraordinary pleasant research atmosphere. I must thank all of my office mates: Andreas Göbel, John Lapinskas, David Richerby, Jacob Focke, Andres Her-

rera Poyatos and Marc Roth. I am especially grateful to have Standa Živný and Charilaos Efthymiou as members of my thesis committee, and have Standa Živný and Varun Kanade as members of my transfer committee. I also wish to thank Elias Koutsoupias and Yufei Zhao, for the wonderful courses they taught on probability, randomised algorithms and combinatorics.

I am also very lucky to be a member of St. Hugh's College, which provides students with comfortable and warm environment for both life and study. I would like to express my gratitude to all college staff. Additionally, I must thank many nice friends in my college: Hanyang Ye, Ruoxi Chang, Yujie Duan, Fan Sha, CY Hsiao, Ding Pei, Ying Xu, Xin Yu, Qingyong Hu, Alaina Oltrogge, Marwan Mohammed, Jinghan Wu and Chaoying Wang.

Outside my research group and St. Hugh's College, I also have many excellent friends over the years: Shuxiang Cao, Zheyi Cao, Wenyu Chen, Jiarui Gan, Tong Gao, Mengyu Gu, Pan Hu, Zihang Lai, Zhenhai Li, Zhu Liang, Yuming Liu, Xiaonan Lu, Xiaoxuan Lu, Yuxin Mi, Yu Tang, Zhihua Wang, Jiaao Wei, Min Wu, Linhai Xie, Zheying Xu, Bo Yang, Mengqi Zhang, Xinxin Zhang, Yuan Zhang, Xiaoyu Zheng. I am lucky to have you as my friends and spend a lot of enjoyable time together. I must express my special thanks to Mengqi Zhang, for our wonderful time in London and Oxford. I would also like to give special thanks to Min Wu and Tong Gao, for our pleasant journey in Prague. Another special thanks go to Yuxin Mi, for the opportunity to watch exciting tennis at The Championships, Wimbledon.

Before coming to Oxford, I graduated from Zhiyuan College at Shanghai Jiao Tong University. I would like to give my gratitude to the founder of computer science program Yong Yu and my undergraduate advisor Pinyan Lu, without whom I probably would not make the decision to pursue a DPhil in theoretical computer science. I also wish to especially thank my undergraduate alumni Zheyi Cao and Zheying Xu, who graduated from Zhiyuan College as well, for teaching me punting and enjoying a great time together in Oxford.

Finally, I would like to express deep thanks to my parents, Zhigang Yang and Daqin Liu, to whom this thesis is dedicated, for their encouragement, support and love.

Abstract

Over the past 30 years, the study of counting problems has become an interesting and important work. The framework of counting problems can capture applications in diverse fields, such as graph colouring and homomorphism in graph theory, and the Ising and Potts model from statistical physics. In this thesis, we will discuss my recent research work focusing on this area.

In Chapter 1, the basic notions and definitions of counting problems will be introduced. Because of the hardness result of computation complexity for counting problems, we will mainly consider whether we have efficient approximation algorithms for counting problems.

In the algorithmic part, one of the most exciting research is to establish the connection between the uniqueness phase transition of spin systems and the approximability phase transition of computing partition functions. In Chapter 2, we first discuss some basic methods of designing approximation algorithms, including the connection between sampling and counting and the correlation decay approach. These methods will be used in Chapter 4 and 5. In Chapter 3, we will discuss the uniqueness threshold for the antiferromagnetic Potts model on d-ary trees and introduce our results for small q. Chapter 4 and 5 contain two examples of counting and sampling in the uniqueness regime. We designed approximation algorithms for counting the number of 4-colourings on cubic graphs and sampling from the Potts model on random regular graphs in the uniqueness regime.

In the complexity part, a classification result for the complexity of approximate counting Boolean Constraint Satisfaction Problems on bounded-degree instances will be introduced in Chapter 6. We focused on whether, for all Boolean constraint languages Γ, there is a "barrier" value Δ_0 such that for

all $\Delta > \Delta_0$, the counting problem of Boolean CSP is computationally hard to approximate, and completely classified the complexity of this problem.

Contents

1 Introduction **1**

1.1 Counting complexity and approximate counting 1

1.1.1 Counting problems 1

1.1.2 #P and #P-complete 2

1.1.3 Polynomial-time approximation schemes 5

1.2 Preliminaries 6

1.2.1 Spin systems 6

1.2.1.1 Two-state spin 6

1.2.1.2 Multi-state spin systems 8

1.2.2 Uniqueness of Gibbs distribution 9

1.3 Main results of this thesis 10

2 Methods **13**

2.1 From the marginal distribution to approximate counting 13

2.2 Approximate counting via correlation decay 15

2.2.1 Spatial mixing and correlation decay 16

2.2.2 Amortised analysis 18

3 Uniqueness Threshold for the q-State Antiferromagnetic Potts Model on d-ary Trees **23**

3.1 Introduction 23

3.1.1 The results 25

3.1.2 Tree recursions 25

3.2 Analysis from the two-step recursion 29

3.2.1 A simpler condition for uniqueness 29

3.2.2 Characterising the maximiser of $h_{c_1,c_2,\beta}$ . . . 32
3.2.3 Concluding the uniqueness threshold . . . 36
3.3 Uniqueness threshold for $q = 3$ . . . 38
3.3.1 Bounding the ratio of marginal probabilities by the one-step recursion 38
3.3.2 Properties of the functions f_{upper} and f_{lower} . . . 42
3.3.3 Bounding the marginal probability at the root by two sequences . . . 44
3.3.4 Bounding the maximum ratio . . . 47
3.3.5 Verifying Condition 3.2.4 . . . 57
3.4 Uniqueness threshold for $q = 4$ . . . 64
3.4.1 A generalisation of f_{upper} and f_{lower} . . . 64
3.4.2 The case in which d is small . . . 71

4 Counting in Uniqueness I : Counting Four-colourings on Cubic Graphs 75
4.1 Introduction . . . 75
4.1.1 The result . . . 76
4.1.2 Prior works . . . 77
4.2 The approximation algorithm . . . 78
4.2.1 Tree recursion and our observations . . . 78
4.2.2 The algorithm for approximately estimate marginal probabilities . 81
4.3 Bounding the marginal probability . . . 84
4.4 Verifying the correlation decay property . . . 93
4.4.1 The case $\deg_G(v) = 1$ . . . 96
4.4.2 The case $\deg_G(v) = 2$ . . . 98
4.4.2.1 A. The case $d_1 = 2$ . . . 98
4.4.2.2 B. The case $d_1 = 1$ . . . 108
4.4.3 Proof of the correlation decay property . . . 113
4.5 Proof of Theorem 4.1.1 . . . 115

5 Counting in Uniqueness II : Sampling Potts on Random Regular Graphs 118
5.1 Introduction . . . 118
5.1.1 The result . . . 119
5.1.2 Prior works . . . 120
5.2 Properties of random regular graphs . . . 121
5.3 Sampling algorithm . . . 125

5.3.1 Connection between Potts on bichromatic classes and the Ising model 125
5.3.2 Analysing an Ideal ReSample subroutine . . . 126
5.3.3 Sampling the Ising model on graphs with small average growth . . 133
5.3.4 Analysis of the Potts algorithm . . . 137
5.4 Analysing Potts on graphs with tree-like structure . . . 143
5.4.1 Analysing Potts on trees . . . 143
5.4.2 Analysing Potts on disjoint trees whose roots are connected via a path 145
5.5 Correlation decay and sampling for antiferromagnetic Ising . . . 150
5.5.1 Preliminaries . . . 150
5.5.2 Proof of Theorem 5.3.6 . . . 151
5.5.3 Proof of Lemma 5.3.7 . . . 155

6 Classification of Complexity : Counting Boolean Bounded-Degree CSP 158
6.1 Introduction . . . 158
6.1.1 Definitions and statement of our classification result . . . 159
6.1.2 Prior works . . . 161
6.1.3 CSPs and hypergraph spin systems . . . 163
6.2 Proof sketch . . . 163
6.2.1 Overview of the proof of our classification result . . . 163
6.2.2 Supporting pinning and equality . . . 166
6.2.3 Realising conditional distributions induced by pinning and equality 167
6.2.4 Simulating hard functions and inapproximability results . . . 167
6.2.5 Technical lemmas . . . 168
6.3 Properties of affine, IM_2, and symmetric functions . . . 170
6.3.1 Affine functions . . . 170
6.3.2 A characterisation of IM_2 . . . 171
6.3.3 The case where f is symmetric: extensions to the asymmetric case 171
6.4 Non-affine Boolean functions either support perfect equality or simulate a hard function . . . 173
6.4.1 A few preparatory lemmas . . . 173
6.4.2 Proof that every non-affine Boolean functions either supports perfect equality or simulates a hard function . . . 178
6.5 The case where f supports perfect equality . . . 188

6.5.1 Constraint Satisfaction Problems and implementations 188
6.5.2 The case of self-dual functions 196
6.5.3 #BIS-easiness 200
6.5.4 #BIS-hardness 201
6.5.5 NP-hardness 203
6.5.6 Proof of Theorem 6.2.10 205
6.6 Proof of Theorem 6.1.7 205

7 Conclusion and Open Problems **208**

8 Appendix: Mathematica Code **211**
8.1 Verifying inequalities in Chapter 3 211
8.1.1 Lemma 3.3.5 211
8.1.2 Lemma 3.3.13 211
8.1.3 Lemma 3.3.17 212
8.1.4 Lemma 3.3.18 212
8.1.5 Lemma 3.3.20 212
8.1.6 Lemma 3.3.21 213
8.1.7 Lemma 3.3.23 214
8.1.8 Lemma 3.3.25 215
8.1.9 Lemma 3.3.28 216
8.1.10 Lemma 3.3.36 217
8.1.11 Lemma 3.3.37 217
8.1.12 Lemma 3.4.4 219
8.1.13 Lemma 3.4.9 220
8.1.14 Lemma 3.4.10 220
8.1.15 Lemma 3.4.13 221
8.2 Verifying inequalities in Chapter 4 222
8.2.1 Initialization 222
8.2.2 Lemma 4.4.6 222
8.2.3 Lemma 4.4.8 222
8.2.4 Lemma 4.4.10, case $f_1 = (1 - f_2)/6$ 222
8.2.5 Lemma 4.4.10, case $f_1 = (3/4 - f_2)/4$ 223
8.2.6 Lemma 4.4.10, case $f_1 = 1/13$ 223

8.2.7 Lemma 4.4.11 223
8.2.8 Lemma 4.4.12 224

Bibliography **224**

List of Figures

4.1 Boundary case one 92
4.2 Boundary case two 92
4.3 Boundary case three 92

Chapter 1

Introduction

1.1 Counting complexity and approximate counting

1.1.1 Counting problems

Counting problems originated from a field of mathematics called enumerative combinatorics where the interest is in studying the number of mathematical structures with given properties. For example, one of the most well-known and beautiful formulas in enumerative combinatorics is *Cayley's formula*, which gives the number of labelled trees on n vertices. Now counting problems have a variety of applications in diverse fields, including statistics, statistical physics and computer science.

In counting problems that arise in computer science, we are interested in computing *algorithmically* the number of structures with given properties. We begin by describing some illustrative examples.

One of the most well-studied problems in computer science is the independent set problem. We will be interested in the following counting problem: given as input a graph G, compute the number of independent sets in G.[1] In general, this problem is very difficult (it is #P-complete as we will explain in Section 1.1.2), even in terms of computing an approximate answer. Nevertheless, we can do better on certain classes of graphs by utilising their combinatorial properties. One such class of graphs are trees, where the absence of cycles implies a simple polynomial-time algorithm. In the following, we describe this polynomial-time algorithm.

In particular, suppose that we want to compute the number of independent sets of a given tree T. For a vertex v of T, let $I_0(T, v)$ denote the number of the independent sets I of

[1] An independent set in a graph G is a subset of vertices which are mutually non-adjacent.

the tree T such that $v \notin I$, and $I_1(T, v)$ denote the number of the independent sets I of the tree T such that $v \in I$. Let $T_1, T_2, \ldots, T_d$ be the disjoint connected components of $T \setminus v$, and $v_1, v_2, \ldots, v_d$ be all neighbour vertices of v in the tree T, where $v_i \in T_i$ for all $1 \leqslant i \leqslant d$. We have a simple recursive algorithm for computing $I_0(T, v)$ and $I_1(T, v)$ by using the following formulas:

$$I_0(T, v) = \begin{cases} \prod_{i \in [d]} (I_0(T_i, v_i) + I_1(T_i, v_i)) & \text{if } T \text{ contains at least two vertices,} \\ 1 & \text{if } T \text{ only contains vertex } v; \end{cases}$$

and

$$I_1(T, v) = \begin{cases} \prod_{i \in [d]} I_0(T_i, v_i) & \text{if } T \text{ contains at least two vertices,} \\ 1 & \text{if } T \text{ only contains vertex } v. \end{cases}$$

Since the number of independent sets of T is $I_0(T, v) + I_1(T, v)$, we thus obtain a polynomial-time algorithm for computing the number of independent sets of trees.

Another example of a counting problem is the problem of computing the number of perfect matchings in a given graph G.[2] Analogously to the independent set problem, counting perfect matchings is in general a very hard problem, but again we can do better on certain classes of graphs such as trees. A more intricate class of graphs where we can compute the number of perfect matchings efficiently are planar graphs, using a beautiful method discovered by Kasteleyn [52]. Kasteleyn designed an algorithm for computing the number of perfect matchings in a planar graph G based on a certain orientation of the graph G and computing an appropriate determinant of the associated (directed) adjacency matrix.

Our goal in this thesis is to identify the cases where we can design polynomial-time algorithms for counting problems, or to establish their computational hardness. In the next section, we first describe the relevant complexity classes for counting problems. Subsequently, we describe the types of algorithms that we will be interested in.

1.1.2 #P and #P-complete

In this section, we describe the complexity class #P which is relevant for classifying the complexity of counting problems. We begin by first reviewing the complexity class NP for decision problems, and then describe its analogue for counting problems.

[2] A perfect matching of a graph $G = (V, E)$ is a subset $M \subseteq E$ of edges such that every vertex of G belongs to exactly one edge in M.

Recall, in a decision problem, the goal is to decide whether the input to the problem has a specific property. For example, given an input graph G and an integer k, does G contain an independent set of size k? More generally, in a decision problem $\mathcal{A}_\varphi$ we are asked to output the value of a function $\varphi : \Sigma^* \to \{0,1\}$, where Σ^* is the set of all possible inputs and an element of Σ^* represents a string whose characters are from an alphabet Σ (e.g., if $\Sigma = \{0,1\}$, the input is encoded in binary). For a string $x \in \Sigma^*$, we denote by $|x|$ the length of the string x.

The decision problem $\mathcal{A}_\varphi$ belongs to the complexity class NP if it can be decided by a nondeterministic Turing machine in time polynomial in the size of the input. An alternative but equivalent definition of NP will be more relevant for generalising to counting problems. In particular, $\mathcal{A}_\varphi$ belongs to the complexity class NP if there exist a polynomial function $p(\cdot)$ and a polynomial-time deterministic Turing Machine $\mathcal{M} : \Sigma^* \times \Sigma^* \to \{0,1\}$ such that,

1. for all $x \in \Sigma^*$ such that $\varphi(x) = 1$, there exists a *witness* $y \in \Sigma^*$ such that $|y| \leqslant p(|x|)$ and $\mathcal{M}(x,y) = 1$, i.e., $\mathcal{M}$ accepts (x,y),
2. for all $x \in \Sigma^*$ such that $\varphi(x) = 0$, for all $y \in \Sigma^*$ with $|y| \leqslant p(|x|)$ it holds that $\mathcal{M}(x,y) = 0$.

We call $\mathcal{M}$ a "witness-checking" predicate. In the decision version of the independent set problem described above, x corresponds to the input graph G and the integer k while the witness y corresponds to an independent set of size k in G (if one exists).

In counting problems, instead of binary-valued functions φ, the goal is to compute a function $f : \Sigma^* \to \mathbb{N}$. The relevant complexity class for counting problems is #P (*number*-P). Roughly, a problem in #P is the counting version of a problem in NP and it corresponds to counting the number of witnesses of an NP problem. More formally, consider a counting problem $\mathcal{A}_f$ where we want to compute the function $f : \Sigma^* \to \mathbb{N}$. Then, $\mathcal{A}_f \in$ #P if there exist a polynomial function $p(\cdot)$ and a polynomial time deterministic Turing Machine $\mathcal{M} : \Sigma^* \times \Sigma^* \to \{0,1\}$ such that for all input $x \in \Sigma^*$,

$$f(x) = |\{y \in \Sigma^* : |y| \leqslant p(|x|) \wedge \mathcal{M}(x,y) = 1\}| .$$

Just as the class NP has complete problems, the class #P also has complete problems which are called #P-complete. To introduce this notion, we first need to define reductions between counting problems (analogously to reductions for decision problems).

Let $\mathcal{A}_f$ and $\mathcal{A}_g$ be two problems in #P, where we want to compute the functions $f : \Sigma_f^* \to \mathbb{N}$ and $g : \Sigma_g^* \to \mathbb{N}$, respectively ($\Sigma_f$, Σ_g are the alphabets used to encode the inputs of $\mathcal{A}_f$ and $\mathcal{A}_g$). We say there is a *parsimonious reduction* from $\mathcal{A}_f$ to $\mathcal{A}_g$ (written as $\mathcal{A}_f \leqslant_{\mathsf{par}} \mathcal{A}_g$) if there is a function $t : \Sigma_f^* \to \Sigma_g^*$ such that

1. on input $x \in \Sigma_f^*$, $t(x)$ can be computed in time polynomial in the size of x,
2. for all $x \in \Sigma_f^*$, $f(x) = g(t(x))$.

Parsimonious reductions are often restrictive and frequently we instead use polynomial-time Turing reductions, where a counting problem A_f reduces to another counting problem A_g if there exists a polynomial time algorithm for A_f given an oracle that solves A_g.

With these definitions, we can now define #P-complete problems. A problem $\mathcal{A}_f$ is #P-complete [75] if

1. $\mathcal{A}_f$ belongs to #P;
2. Every problem in #P can be reduced to $\mathcal{A}_f$ by a polynomial-time Turing reduction.

Usually, the counting version of an NP-complete problem is #P-complete (though it is not known whether this holds in general). For example, let 3-SAT be the decision problem where, given a Boolean formula in conjunctive normal form whose clauses have at most three literals, we need to decide whether there exists a satisfying assignment of the formula. It is well known that 3-SAT is an NP-complete problem [51]. The counting version of 3-SAT, denoted by #3-SAT, is to compute the number of all satisfying assignments. As it turns out, #3-SAT is a #P-complete problem.

Theorem 1.1.1 ([75]). *#3-SAT is #P-complete under parsimonious reductions.*

Interestingly, under Turing reductions, there exist #P-complete problems whose decision version is solvable in polynomial time. The most well-known example is the problem of counting perfect matchings of bipartite graphs. Deciding whether a bipartite graph contains a perfect matching can be solved in polynomial time. However, Valiant proved that counting perfect matchings of bipartite graphs is #P-complete under Turing reductions.

Theorem 1.1.2 ([75]). *The problem of counting perfect matchings of bipartite graphs is #P-complete under Turing reductions.*

For #P-complete problems, it is generally believed in complexity theory that we cannot solve them exactly in polynomial time. So we will instead focus on approximation algorithms, which we describe in the next section.

1.1.3 Polynomial-time approximation schemes

In this section, we define the class of approximation algorithms that we will use to study counting problems. In particular, we define *fully polynomial-time approximation schemes* (FPTAS) and *fully polynomial-time randomised approximation schemes* (FPRAS).
Let $\mathcal{A}_f$ be a counting problem as in Section 1.1.2, i.e., our goal is to compute a function $f : \Sigma^* \to \mathbb{N}$, where Σ^* encodes the input to the problem (strings from an alphabet Σ).

Definition 1.1.3. An algorithm is an FPTAS for the counting problem $\mathcal{A}_f$ if, on input x and a parameter $\varepsilon \in (0, 1)$, the algorithm outputs a number $\hat{f}$ in time $\mathrm{poly}(|x|, 1/\varepsilon)$ such that

$$(1 - \varepsilon)f(x) \leqslant \hat{f} \leqslant (1 + \varepsilon)f(x).$$

An FPRAS is the randomised version of an FPTAS, and is defined as follows.

Definition 1.1.4. A randomised algorithm is an FPRAS for the counting problem $\mathcal{A}_f$ if, on input x and parameters $\varepsilon, \delta \in (0, 1)$, the algorithm outputs a number $\hat{f}$ in time $\mathrm{poly}(|x|, 1/\varepsilon, \log(1/\delta))$ such that

$$(1 - \varepsilon)f(x) \leqslant \hat{f} \leqslant (1 + \varepsilon)f(x),$$

with probability at least $1 - \delta$.

Note that if the answer to a counting problem is 0, an FPTAS/FPRAS must output 0, whereas if the answer to a counting problem is larger than 0, an FPTAS/FPRAS must output a positive number. Therefore, if a counting problem admits an FPTAS/FPRAS, then we can solve its decision problem in (randomised) polynomial time as well.
More interestingly, if we have an oracle for an NP-complete problem, then we can design an FPRAS for any counting problem in #P.

Theorem 1.1.5 ([74])**.** *Given an oracle for an* NP*-complete problem, there is an* FPRAS *for every problem in* #P.

In the rest of this thesis, we will be interested in designing an FPTAS or FPRAS for counting problems without using the assumption of having an oracle for NP.

1.2 Preliminaries

In this section, we first give some general notations which will be used in this thesis. Then we define spin systems and the associated counting problems which will be the major focus of this thesis.

1.2.1 Spin systems

1.2.1.1 Two-state spin

Spin systems play an important role in statistical physics and computer science. In statistical physics, spin systems are used to explain physical phenomena on the basis of interacting particles. In computer science, spin systems are relevant for studying well-known combinatorial problems such as the independent set and proper q-colouring problems.

We begin by describing 2-state spin systems. Roughly, a 2-state spin system is comprised of sites and each site has a spin which can be in one of two possible states, say $\{0, 1\}$. Each site interacts with its neighbouring sites and the strength of the interaction depends on the sites' states.

More specifically, we will think of the particles being arranged on a graph $G = (V, E)$ where vertices in V represent sites and edges in E represent whether two sites are neighbouring. A configuration of the spin system is an assignment $\sigma : V \to \{0, 1\}$, i.e., a configuration σ assigns a spin state to each vertex. The weight of a configuration σ is given by

$$w_G(\sigma) = \prod_{(u,v)\in E} A_{\sigma(u),\sigma(v)} \prod_{v\in V} b_{\sigma(v)},$$

where A is a 2×2 symmetric matrix with nonnegative entries and b is a 2-dimensional vector with nonnegative entries. The standard parametrisation of the matrix A and the vector b are given as follows:

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{pmatrix} = \begin{pmatrix} \beta & 1 \\ 1 & \gamma \end{pmatrix} \text{ and } b = (b_0, b_1) = (1, \lambda),$$

where β, γ, λ are nonnegative numbers (in statistical physics, λ is referred to as the absolute activity or fugacity).

The partition function of the spin system is given by

$$Z_G = \sum_{\sigma: V \to \{0,1\}} w_G(\sigma).$$

The Gibbs distribution $\Pr_G[\cdot]$ associated to the spin system is a probability distribution on the space of all configurations $\sigma : V \to \{0, 1\}$. In particular, the probability mass of a configuration σ in the Gibbs distribution is given by

$$\Pr_G[\sigma] = \frac{w_G(\sigma)}{Z_G}.$$

Now we describe two important examples of 2-state spin systems:

- The hardcore model with parameter $\lambda > 0$ corresponds to the problem of counting (weighted) independent sets of a graph $G = (V, E)$. In the hardcore model, the matrix A and the vector b are given by

  $$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } b = (1, \lambda).$$

  Note that for each configuration $\sigma : V \to \{0, 1\}$ with positive weight, it holds that the spins of adjacent vertices cannot be simultaneously equal to 1, i.e., for every $(u, v) \in E$, it holds that $\sigma(u) \cdot \sigma(v) = 0$. In particular, whenever $w_G(\sigma) > 0$, the set of vertices v such that $\sigma(v) = 1$ is an independent set of the graph G, which we denote by $I(\sigma)$, and the weight of $I(\sigma)$ is $\lambda^{|I(\sigma)|}$. We can therefore write the partition function of the hardcore model in the following equivalent way:

  $$Z_G = \sum_{\sigma: V \to \{0,1\}} w_G(\sigma) = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|},$$

  where $\mathcal{I}(G)$ denotes the set of all independent sets of the graph G. Similarly, the Gibbs distribution of the hardcore model can be equivalently interpreted as a probability distribution supported on the set $\mathcal{I}(G)$.

  Note that, in the case $\lambda = 1$, the partition function of a graph G in the hardcore model is the same as the number of independent sets of G, i.e., $Z_G = |\mathcal{I}(G)|$, and the Gibbs distribution is the uniform distribution over all independent sets.

- The Ising model with parameter $\beta > 0$ is a model of ferromagnetism in statistical physics. In the Ising model (without external field), the matrix A and the vector b are given by

  $$A = \begin{pmatrix} \beta & 1 \\ 1 & \beta \end{pmatrix} \text{ and } b = (1, 1).$$

Hence, the weight $w_G(\sigma)$ of a configuration $\sigma : V \to \{0, 1\}$ is equal to $\beta^{m(\sigma)}$, where $m(\sigma)$ is the number of edges whose endpoints have the same spin under σ.

For $0 < \beta < 1$, the Ising model is called *antiferromagnetic*, whereas for $\beta > 1$, the model is called *ferromagnetic*. Roughly, for $\beta > 1$, configurations σ with many monochromatic edges have larger weight than configurations with few monochromatic edges (and the other way around for $\beta < 1$).

Later, we will be interested in the problem of computing (approximately) the partition function of a 2-state spin system, given an input graph G.

1.2.1.2 Multi-state spin systems

We can generalise 2-state spin systems in multiple ways; in this section, we consider multi-state spin systems, where the number of spin states can potentially be larger than two.

Let $q \geqslant 2$ be an integer. We will use q to denote the number of states and $[q]$ to denote the set $\{1, 2, \ldots, q\}$ of spin states. Analogously to the 2-state spin system, in a q-state spin system on a graph $G = (V, E)$, a configuration is an assignment $\sigma : V \to [q]$. The weight of a configuration σ is given by

$$w_G(\sigma) = \prod_{(u,v)\in E} A_{\sigma(u),\sigma(v)},$$

where now A is a $q \times q$ symmetric matrix with nonnegative entries. The partition function of the q-state spin system is again given by the aggregate weight of all configurations, i.e.,

$$Z_G = \sum_{\sigma:V\to[q]} w_G(\sigma).$$

Also, in the Gibbs distribution $\Pr_G[\cdot]$, a configuration σ has probability mass given by

$$\Pr_G[\sigma] = \frac{w_G(\sigma)}{Z_G}.$$

A canonical example of a multi-state spin system is the q-state Potts model, which is a generalisation of the Ising model. In the Potts model, the parameter matrix A is given by

$$A = \begin{pmatrix} \beta & 1 & 1 & \cdots & 1 \\ 1 & \beta & 1 & \cdots & 1 \\ 1 & 1 & \beta & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & \beta \end{pmatrix}$$

where $\beta \geqslant 0$. Note that the case $q = 2$ and $\beta > 0$ is precisely the case of the Ising model. For $q \geqslant 3$, the case $\beta = 0$ is particularly interesting, since it corresponds to the proper q-colouring problem. In particular, for $\beta = 0$, we can rewrite the partition function as

$$Z_G = \sum_{\sigma: V \to [q]} \prod_{(u,v) \in E} \mathbf{1}_{\sigma(u) \neq \sigma(v)}, \text{ where } \mathbf{1}_{\sigma(u) \neq \sigma(v)} = \begin{cases} 1 & \text{if } \sigma(u) \neq \sigma(v); \\ 0 & \text{if } \sigma(u) = \sigma(v). \end{cases}$$

Observe that a configuration σ has positive weight iff every pair of adjacent vertices in G have different states, i.e., σ induces a proper q-colouring of G. Hence, in this case ($\beta = 0$), Z_G is the number of proper q-colourings of the graph G.

1.2.2 Uniqueness of Gibbs distribution

Note that in Chapter 1.2.1, the Gibbs distribution is only defined on finite graphs. For infinite graphs, we say a probability distribution $\Pr_G[\,\cdot\,]$ is a Gibbs distribution of an infinite graph G if the marginal distribution under $\Pr_G[\,\cdot\,]$ on any finite subgraph H of G, conditioned on the configuration of $(G \setminus H) \cup \partial H$ is the same to the Gibbs distribution on H conditioned on the configuration of the boundary of H. (The boundary ∂H of H is the set of those vertices in H which are adjacent to some vertex in $G \setminus H$.) Formally, let $G = (V, E)$ be an infinite graph. For any configuration $\sigma : V \to [q]$ and any subset U of V, denote by $\sigma(U)$ the restriction of the configuration σ to the vertices in U. Then $\Pr_G[\,\cdot\,]$ is a Gibbs distribution if for any finite subgraph $H \subseteq G$ and any fixed configuration $\sigma : V \to [q]$, the conditioned distribution for any configuration $\tau : V \to [q]$ satisfies

$$\Pr_G[\tau \mid \tau((G \setminus H) \cup \partial H) = \sigma((G \setminus H) \cup \partial H)] = \Pr_H[\tau \mid \tau(\partial H) = \sigma(\partial H)]\,,$$

where $\Pr_H[\,\cdot\,]$ is the Gibbs distribution on the finite graph H. It is a well known result in statistical physics that at least one such probability distribution always exists (see [40]). However, for certain values of the parameters, there might be multiple such distributions. We are interested in the case where the Gibbs distribution is unique. Given a graph $G = (V, E)$ and a set $A \subseteq V$, let $S(A, n)$ denote the set $\{v \in V \mid d(v, A) = n\}$ where $d(v, A)$ is the smallest distance from v to the set A. The uniqueness of the Gibbs distribution can be defined as follows.

Definition 1.2.1 ([71])**.** On an infinite graph $G = (V, E)$ with finite degrees we say that a Gibbs measure is *unique* if for any finite set $A \subseteq V$,

$$\limsup_{n \to \infty} \max_{\sigma_1, \sigma_2} \left\| \Pr_G[\sigma(A) = \cdot \mid \sigma(S(A, n)) = \sigma_1] - \Pr_G[\sigma(A) = \cdot \mid \sigma(S(A, n)) = \sigma_2] \right\|_{\mathrm{TV}} = 0,$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm and the maximum is taken over all boundary condition $\sigma_1, \sigma_2 \colon S(A, n) \to [q]$.

In particular, we use the following definition that formalises uniqueness instead to simplify the criteria of the uniqueness for the Potts model on d-ary trees. Let $\mathbb{T}_{d,n}$ be the rooted d-ary tree with height n (i.e., every path from the root to a leaf has n edges, and every non-leaf vertex has d children). In particular, denote by $\mathbb{T}_d$ the infinite d-ary tree. Let $\Lambda_{\mathbb{T}_{d,n}}$ be the set of leaves of $\mathbb{T}_{d,n}$ and let $v_{d,n}$ be its root.

Definition 1.2.2. The q-state Potts model with parameter β exhibits *uniqueness* on the infinite d-ary tree $\mathbb{T}_d$ if, for all colours $c \in [q]$, it holds that

$$\limsup_{n\to\infty} \max_{\tau \colon \Lambda_{\mathbb{T}_{d,n}} \to [q]} \left| \Pr_{\mathbb{T}_{d,n}} \left[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau \right] - \frac{1}{q} \right| = 0. \tag{1.1}$$

It exhibits *non-uniqueness* otherwise.

1.3 Main results of this thesis

In this thesis, we will mainly discuss the uniqueness threshold for the q-state Potts model, and then design some algorithms for approximate counting in the uniqueness regime. Finally, we will also establish the classification of complexity of approximate counting of Boolean Constraint Satisfaction Problems.

In Chapter 2, we first discuss some basic methods of designing approximation algorithms, including the connection between sampling and counting and the correlation decay approach. These methods will be used in Chapter 4 and 5.

In Chapter 3, we will discuss the uniqueness threshold for the antiferromagnetic Potts model on d-ary trees. For the ferromagnetic Potts model, the uniqueness threshold is well-studied. But for the antiferromagnetic Potts model, the threshold is much less clear, even on the regular trees. Jonasson established the uniqueness threshold in the zero-temperature case, which corresponds to the q-colourings model. In the permissive case (where the temperature is positive), the Potts model has an extra parameter $\beta \in (0, 1)$, which makes the task of analysing the uniqueness threshold even harder and much less is understood. Our work is to extend the analysis to the antiferromagnetic case. We first focus on the case $q = 3$ and give a detailed analysis of the Potts model on the tree by refining Jonasson's approach. In particular, we establish the uniqueness threshold

on the d-ary trees for all values of $d \geq 2$. When $d \geq 3$, we show that the 3-state antiferromagnetic Potts model has uniqueness for all $\beta \geq 1 - 3/(d+1)$. The case $d = 2$ is critical since it relates to the 3-colourings model on the binary tree (where $\beta = 0$), which has non-uniqueness. Nevertheless, we show that the Potts model has uniqueness for all $\beta \in (0, 1)$ on the binary tree. Both of these results are tight since it is known that uniqueness does not hold in the complementary regime. Then we consider general $q > 3$ cases. Our techniques give a criteria to prove that the Potts model has uniqueness with given parameters q and d, based on a two-step recursion on the tree. For $q = 4$, we verified this condition and establish the uniqueness threshold for $4 \leq d \leq 9$. However, it is conjectured that the Potts model has uniqueness on d-ary trees as long as $\beta \in (0, 1)$ and $\beta \geq 1 - q/(d+1)$. The exact threshold is still open.

For some particular two-state spin systems, such as the antiferromagnetic Ising model and the hard-core model, the connection between the uniqueness phase transition and the complexity of approximate counting problems has been well studied. It has been proved that an FPTAS/FPRAS exists iff the system has uniqueness. So the same picture is conjectured to hold for some general models, such as the colouring model and the general Potts model. In Chapter 4 and 5, we will introduce some approximation algorithms for counting and sampling in the uniqueness regime.

- In Chapter 4, we will discuss the counting problem of the q-colourings model, which is a special case of the antiferromagnetic Potts model. Let Δ denote the maximum degree of the graph. It is known that the q-colouring model on infinite regular trees has uniqueness iff $q \geq \Delta + 1$. This fact supports the conjecture that there is an FPTAS/FPRAS for counting the number of proper colourings as long as $q \geq \Delta + 1$. However, the conjecture remained open even for any fixed $\Delta \geq 3$.

  We refine the techniques of the correlation decay approach and design an FPTAS for counting the number of proper 4-colourings on graphs with maximum degree 3 and thus confirm the conjecture in the case of $\Delta = 3$. Our algorithm is the first one to achieve this optimal bound of $q = \Delta + 1$.

- In Chapter 5, we will consider the problem of sampling from the antiferromagnetic Potts model on random regular graphs, which is an important application of our uniqueness results. It is conjectured that polynomial-time approximate sampling is possible when the temperature of the model is in the uniqueness regime of the

regular trees, but designing algorithms remains a difficult task. In this chapter, for all integers $q \geq 3$ and $\Delta \geq 3$, we develop algorithms that produce samples within error $o(1)$ from the q-state antiferromagnetic Potts model on random Δ-regular graphs, whenever the temperature is in uniqueness.

This algorithm is based on iteratively adding the edges of the graph and resampling a bichromatic class that contains the endpoints of the newly added edge. We exploit the tree uniqueness to show that the average growth of bichromatic components is typically small, which allows us to use correlation decay algorithms for the resampling step. While the precise uniqueness threshold on the tree is still not known for general values of q and Δ in the antiferromagnetic case, our algorithm works throughout uniqueness regardless of its value.

In Chapter 6, we establish the classification for the complexity of approximate counting Boolean Constraint Satisfaction Problems (CSP) on bounded-degree instances. Let Δ be the maximum number of occurrence of any variables. For every fixed $\Delta \geq 3$, the complexity of exactly solving this problem depends on the set of allowed constraints has been completed. However, in the approximate setting, even for some widely studied case we do not know much about their complexity. For a special case which is known as the "conservative case" in the CSP literature, the complexity of approximation problem has been classified. But the "conservative case" does not include many natural problems such as the hypergraph independent sets problem. Our result extends this result in the "conservative" case. Since even the complexity of the decision version is still not completely understood, in our work, we focused on whether, for all Boolean constraint languages Γ, there is a "barrier" value Δ_0 such that for all $\Delta > \Delta_0$, the counting problem of Boolean CSP is computationally hard to approximate. In our setting we completely classified the complexity of approximately counting Boolean CSP problems.

Chapter 2

Methods

In this chapter, we review some well-known methods which are algorithmic techniques to design approximation algorithms for counting problems. We will focus on counting problems induced by spin systems: given a spin system on a graph/hypergraph, can we (approximately) compute its partition function? To illustrate the relevant ideas, throughout this chapter, we will focus on computing the partition function of the hardcore model on graphs.

2.1 From the marginal distribution to approximate counting

Consider the hardcore model (which is defined in Chapter 1.2.1.1) with parameter λ on a given graph $G = (V, E)$. For a vertex $v \in V$, we will use $\Pr_G[\sigma(v) = 1]$ to denote the marginal probability that $\sigma(v) = 1$, where σ is a random configuration sampled from the Gibbs distribution, i.e.,

$$\Pr_G[\sigma(v) = 1] = \sum_{\sigma:\sigma(v)=1} \Pr_G[\sigma]\,.$$

Interestingly, an FPTAS for computing the marginal probability $\Pr_G[\sigma(v) = 1]$ implies an FPTAS for computing the partition function. We have the following folklore lemma.

Lemma 2.1.1. *Consider the hardcore model with parameter $\lambda > 0$. Suppose that given a graph $G = (V, E)$, a vertex $v \in V$ and a parameter $\varepsilon \in (0, 1)$, we can compute in time $\mathrm{poly}(|V|, 1/\varepsilon)$ a number $\hat{p}$ such that*

$$(1 - \varepsilon)\Pr_G[\sigma(v) = 1] \leqslant \hat{p} \leqslant (1 + \varepsilon)\Pr_G[\sigma(v) = 1]\,.$$

Then, we also have an FPTAS for computing the partition function of the hardcore model with parameter λ.

Proof. Let $G = (V, E)$ be a graph for which we want to compute its partition function Z_G with relative accuracy $\varepsilon \in (0, 1)$, i.e., we want to compute $\hat{Z}$ in time $\mathrm{poly}(|V|, 1/\varepsilon)$ such that

$$(1 - \varepsilon) Z_G \leqslant \hat{Z} \leqslant (1 + \varepsilon) Z_G. \tag{2.1}$$

Assume $V = \{v_1, \ldots, v_n\}$ and consider the configuration σ_0, where $\sigma_0(v) = 0$ for all $v \in V$. Note that σ_0 corresponds to the empty independent set and therefore $w_G(\sigma_0) = \lambda^0 = 1$. We can thus write

$$Z_G = \frac{1}{\Pr_G[\sigma_0]} = \frac{1}{\Pr_G[\bigwedge_{i=1}^{n} \sigma(v_i) = 0]} = \frac{1}{\prod_{i=1}^{n} \Pr_G\left[\sigma(v_i) = 0 \mid \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0\right]}.$$

For $1 \leqslant i \leqslant n$, let G_i be the graph induced by the set of vertices $v_i, \ldots, v_n$. The key observation is that

$$\Pr_G\left[\sigma(v_i) = 0 \mid \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0\right] = \Pr_{G_i}[\sigma(v_i) = 0]\,.$$

Thus,

$$Z_G = \prod_{i=1}^{n} \frac{1}{\Pr_{G_i}[\sigma(v_i) = 0]}. \tag{2.2}$$

By assumption, we can compute in time $\mathrm{poly}(|V|, 1/\varepsilon)$ estimates $p_1, \ldots, p_n$ such that

$$\left(1 - \frac{\varepsilon}{2\lambda n}\right) \cdot \Pr_{G_i}[\sigma(v_i) = 1] \leqslant p_i \leqslant \left(1 + \frac{\varepsilon}{2\lambda n}\right) \cdot \Pr_{G_i}[\sigma(v_i) = 1]$$

for all $i = 1, \ldots, n$. Note that, if I is an independent set containing v_i of G_i, then $I \setminus v_i$ is also an independent set of G_i, which gives that $\Pr_{G_i}[\sigma(v) = 1] \leqslant \lambda \cdot \Pr_{G_i}[\sigma(v) = 0]$. Therefore,

$$1 - p_i \leqslant 1 - \left(1 - \frac{\varepsilon}{2\lambda n}\right) \cdot \Pr_{G_i}[\sigma(v_i) = 1] \leqslant \left(1 + \frac{\varepsilon}{2n}\right) \cdot \Pr_{G_i}[\sigma(v_i) = 0]\,. \tag{2.3}$$

Analogously,

$$1 - p_i \geqslant 1 - \left(1 + \frac{\varepsilon}{2\lambda n}\right) \cdot \Pr_{G_i}[\sigma(v) = 1] \geqslant \left(1 - \frac{\varepsilon}{2n}\right) \cdot \Pr_{G_i}[\sigma(v) = 0]\,. \tag{2.4}$$

Let $\hat{Z} = \prod_{i=1}^{n} \frac{1}{1-p_i}$. We will use $\hat{Z}$ as our estimate for the partition function of G. Indeed, using (2.2), (2.3) and (2.4), it holds that

$$1 - \varepsilon \leqslant \left(1 + \frac{\varepsilon}{2n}\right)^{-n} \leqslant \frac{\hat{Z}}{Z_G} \leqslant \left(1 - \frac{\varepsilon}{2n}\right)^{-n} \leqslant 1 + \varepsilon,$$

which proves (2.1), as we wanted. □

Lemma 2.1.1 generalises in a randomised setting. Namely, an FPRAS for computing the marginal probability $\Pr_G[\sigma(v) = 1]$ can be used to design an FPRAS for the partition function. The proof is very similar to the proof of Lemma 2.1.1 and can be found in [46].

Lemma 2.1.2. *Consider the hardcore model with parameter $\lambda > 0$. Suppose that given a graph $G = (V, E)$, a vertex $v \in V$ and parameters $\varepsilon, \delta \in (0, 1)$, we can compute in time $\mathrm{poly}(|V|, 1/\varepsilon, \log(1/\delta))$ a number $\hat{p}$ such that $\hat{p}$ satisfies*

$$(1 - \varepsilon)\Pr_G[\sigma(v) = 1] \leqslant \hat{p} \leqslant (1 + \varepsilon)\Pr_G[\sigma(v) = 1],$$

with probability at least $1 - \delta$.
Then, we also have an FPRAS *for computing the partition function of the hardcore model with parameter λ.*

Lemmas 2.1.1 and 2.1.2 display that, in order to approximate the partition function, it suffices to design an algorithm for approximating marginal probabilities. But how can we approximate marginal probabilities such as $\Pr_G[\sigma(v) = 1]$? The most well-known technique is the so-called Markov chain Monte Carlo method (MCMC), which is based on designing a Markov chain to (approximately) sample from the Gibbs distribution. One can then use samples of the Gibbs distribution to estimate via rejection sampling the probability $\Pr_G[\sigma(v) = 1]$. The bottleneck of applying this method is to analyse the convergence time of the underlying Markov chain, which is frequently a very challenging task. For example, we know an MCMC based FPRAS for counting independent sets of graphs G with maximum degree $\Delta \leqslant 4$ (i.e., the hardcore model with $\lambda = 1$) [28]. It is believed that this algorithm should also work for $\Delta = 5$ but this has not been proved yet, precisely because it is not known whether the convergence time of the Markov chain is bounded by a polynomial in the size of the graph. Similarly, for counting proper q-colourings of graphs with maximum degree Δ, it is conjectured that MCMC methods yield polynomial-time algorithms for $q \geqslant \Delta + 1$, but this has only been proved for $q \geqslant 11\Delta/6$ [77].

2.2 Approximate counting via correlation decay

In this section, we will introduce an important approach to design FPTASes for computing marginal probabilities.

2.2.1 Spatial mixing and correlation decay

The correlation decay method is a new way to design approximate counting algorithms which circumvents the need to sample from the Gibbs distribution. Unlike MCMC approaches, the correlation decay method is based on computing directly marginal probabilities. Once again, we illustrate the correlation decay method by describing it in the context of computing the partition function of the hardcore model with parameter $\lambda > 0$.

Recall, for a graph $G = (V, E)$ and a vertex $v \in V$, $\Pr_G[\sigma(v) = 1]$ denotes the marginal probability that $\sigma(v) = 1$ where σ is a configuration sampled from the Gibbs distribution, i.e., $\Pr_G[\sigma(v) = 1] = \sum_{\sigma:\sigma(v)=1} \Pr_G[\sigma]$. Weitz [78] showed how to approximate $\Pr_G[\sigma(v) = 1]$ on graphs G of maximum degree Δ when the parameter λ is below a certain threshold value λ_Δ. His method has two important steps:

1. The computation of $\Pr_G[\sigma(v) = 1]$ is reduced to computing an appropriate marginal probability on a tree T. In general, the tree T has exponential size relative to the size of the initial graph G.

2. By truncating the tree T, one can obtain in polynomial time an estimate for the quantity $\Pr_G[\sigma(v) = 1]$. Of course, the truncation of the tree T introduces an error and therefore one needs to show that the error introduced by the truncation can be made small with respect to the true value.

The first step was accomplished in [78] by considering the so-called self-avoiding walk tree rooted at the vertex v. The second step turns out to be the bottleneck of the method in the sense that the error between the truncated tree and the full tree can be made sufficiently small only when λ is below the critical threshold λ_Δ.

In the following, we will focus on explaining the second step in more detail which is the crux of the correlation decay method.

Consider the hardcore model with parameter λ on a graph $G = (V, E)$ and let $v \in V$ be a vertex. In the correlation decay method, we are interested in how the probability $\Pr_G[\sigma(v) = 1]$ is influenced by fixing the configuration on a subset $\Lambda \subseteq V$ which is far away from the vertex v. The rough idea is that, if this influence is small, we can hope to compute $\Pr_G[\sigma(v) = 1]$ in good precision by just considering a small subgraph of G induced by the local neighbourhood of v.

To formalise this, for a configuration $\sigma : V \to \{0,1\}$, we denote by σ_Λ the restriction of σ to the set Λ. Let $\tau_\Lambda : \Lambda \to \{0,1\}$ be an arbitrary configuration on Λ. Then we will use $p_{G,v}^{\tau_\Lambda}$ to denote the conditional probability $\Pr_G[\sigma(v) = 1 \mid \sigma_\Lambda = \tau_\Lambda]$, i.e.,

$$p_{G,v}^{\tau_\Lambda} = \Pr_G[\sigma(v) = 1 \mid \sigma_\Lambda = \tau_\Lambda] = \frac{\sum_{\sigma:\sigma(v)=1,\sigma_\Lambda=\tau_\Lambda} \Pr_G[\sigma]}{\sum_{\sigma:\sigma_\Lambda=\tau_\Lambda} \Pr_G[\sigma]}.$$

The notion of spatial mixing captures the sensitivity of $p_{G,v}^{\tau_\Lambda}$ with respect to the configuration τ_Λ. There are two related notions of spatial mixing, weak and strong spatial mixing [78, 53, 54]. For a vertex $v \in V$ and $\Lambda \subseteq V$, we denote by $\mathrm{dist}(v, \Lambda)$ the length of the shortest path in G between v and Λ.
First we have the definition of weak spatial mixing:

Definition 2.2.1. Let $\delta : \mathbb{N} \to \mathbb{R}^+$ and $G = (V, E)$ be a finite graph. We say that the graph G exhibits weak spatial mixing with decay rate $\delta(\cdot)$ if: for every $v \in V$, for every $\Lambda \subseteq V$, and any two configurations $\tau_\Lambda, \eta_\Lambda : \Lambda \to \{0,1\}$, it holds that

$$\left| p_{G,v}^{\tau_\Lambda} - p_{G,v}^{\eta_\Lambda} \right| \leqslant \delta(\mathrm{dist}(v, \Lambda)).$$

Also we have the following definition of strong spatial mixing:

Definition 2.2.2. Let $\delta : \mathbb{N} \to \mathbb{R}^+$ and $G = (V, E)$ be a finite graph. We say that the graph G exhibits strong spatial mixing with decay rate $\delta(\cdot)$ if: for every $v \in V$, for every $\Lambda \subseteq V$, and any two configurations $\tau_\Lambda, \eta_\Lambda : \Lambda \to \{0,1\}$, it holds that

$$\left| p_{G,v}^{\tau_\Lambda} - p_{G,v}^{\eta_\Lambda} \right| \leqslant \delta(\mathrm{dist}(v, \Psi)),$$

where $\Psi \subseteq \Lambda$ stands for the subset on which σ_Λ and τ_Λ differ.

Note that if a graph G has strong spatial mixing with decay rate $\delta(\cdot)$, it also has weak spatial mixing; the reverse direction is not in general true.
For the hardcore model, we can extend Definitions 2.2.1 and 2.2.2 for infinite graphs $G = (V, E)$ as follows. An infinite graph G has weak spatial mixing with decay rate $\delta(\cdot)$ if any finite subgraph G' of G exhibits weak spatial mixing with decay rate $\delta(\cdot)$. Similarly, an infinite graph G has strong spatial mixing with decay rate $\delta(\cdot)$ if any finite subgraph G' of G exhibits strong spatial mixing with decay rate $\delta(\cdot)$.

Definition 2.2.3. Let $G = (V, E)$ be an infinite graph. We say that the graph G has the correlation decay property if G exhibits strong spatial mixing with rate $\delta(\cdot)$ which decays

exponentially, i.e., there exist $C_1, C_2 > 0$ such that $\delta(\ell) \leqslant C_1 \exp(-C_2 \cdot \ell)$ for all integers $\ell > 0$.

As a non-trivial example of these definitions, consider the infinite Δ-regular tree $\mathbb{T}_\Delta$ for an integer $\Delta \geqslant 3$. We have the following theorem, whose proof we will discuss in the following section.

Theorem 2.2.4 ([78]). *Consider the hardcore model with parameter λ. Then, for all $\Delta \geqslant 3$, for all $\lambda < \lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$, the infinite Δ-regular tree $\mathbb{T}_\Delta$ has the correlation decay property.*

To illustrate the importance of the correlation decay property in designing approximation algorithms, we state the following theorem of Weitz.

Theorem 2.2.5 ([78]). *Let $\Delta \geqslant 3, \lambda > 0$. Consider the hardcore model with parameter λ. If the infinite Δ-regular tree $\mathbb{T}_\Delta$ has the correlation decay property, then there is an* FPTAS *for computing the partition function of graphs G with maximum degree Δ.*

Theorem 2.2.5 is proved by reducing the computation of the marginal $\Pr_G[\sigma(v) = 1]$ to the computation of a marginal probability on a tree T (the self-avoiding walk tree rooted at v), which is a finite subtree of $\mathbb{T}_\Delta$. While this tree T has size exponential with respect to the size of the graph G, the correlation decay property allows us to compute a close estimate on a truncated subtree T' of polynomial size.
Note, by combining Theorems 2.2.4 and 2.2.5, we can conclude that there exists an FPTAS for computing the partition function of graphs G with maximum degree Δ whenever $\lambda < \lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$.
In the following section, we focus on the proof on Theorem 2.2.4.

2.2.2 Amortised analysis

In [78], one of the main difficulties in obtaining an FPTAS for the partition function of the hardcore model was to prove Theorem 2.2.4, i.e., to show the correlation decay property on the infinite Δ-regular tree. There, a special monotonicity property of the hardcore model was used in the analysis. Unfortunately, this method does not generalise easily to other models. In this section, we instead present a method to prove the correlation decay property for the hardcore model, which has the benefit that it generalises to other spin systems as well.

Let $\Delta \geqslant 3$ be an integer. To show the correlation decay property for the Δ-regular tree $\mathbb{T}_\Delta$, consider a finite subtree $T = (V, E)$ of $\mathbb{T}_\Delta$, a vertex $v \in V$ and a subset $\Lambda \subseteq V$. Further, consider two arbitrary configurations $\tau_\Lambda, \eta_\Lambda : \Lambda \to \{0, 1\}$ and let $\Psi \subseteq \Lambda$ denote the subset on which τ_Λ and η_Λ differ. We need to show that there exists constants $C_1, C_2 > 0$ (which do not depend on $T, \Lambda, V, \tau_\Lambda, \eta_\Lambda$) such that

$$|p_v^{\tau_\Lambda} - p_v^{\eta_\Lambda}| \leqslant C_1 \exp\big(-C_2 \cdot \operatorname{dist}(v, \Psi)\big). \tag{2.5}$$

Let $T_1 = (V_1, E_1), \ldots, T_d = (V_d, E_d)$ be the connected components of $T \setminus v$ and let $v_1, \ldots, v_d$ be the neighbours of v in $T_1, \ldots, T_d$, respectively. Further, let $\Lambda_i = \Lambda \cap V_i$ and $\tau_{\Lambda_i}, \eta_{\Lambda_i} : \Lambda_i \to \{0, 1\}$ be the restriction of τ_Λ and η_Λ to the set Λ_i, respectively. Note that, since the tree T has maximum degree at most Δ, it holds that $d \leqslant \Delta$.
Using the fact that T is a tree, it is not hard to show that we have the recursive formulas:

$$p_{T,v}^{\tau_\Lambda} = 1 - \frac{1}{1 + \lambda \prod_{i=1}^{d} (1 - p_{T_i,v_i}^{\tau_{\Lambda_i}})} \text{ and } p_{T,v}^{\eta_\Lambda} = 1 - \frac{1}{1 + \lambda \prod_{i=1}^{d} (1 - p_{T_i,v_i}^{\eta_{\Lambda_i}})}.$$

Thus, we can hope to study the difference between $p_{T,v}^{\tau_\Lambda}$ and $p_{T,v}^{\eta_\Lambda}$ recursively. In particular, for a vector $\mathbf{x} = (x_1, \ldots, x_d)$, consider the function

$$f_d(\mathbf{x}) = 1 - \frac{1}{1 + \lambda \prod_{i=1}^{d} (1 - x_i)},$$

and observe that

$$p_{T,v}^{\tau_\Lambda} = f_d(p_{T_1,v_1}^{\tau_{\Lambda_1}}, \ldots, p_{T_d,v_d}^{\tau_{\Lambda_d}}) \text{ and } p_{T,v}^{\eta_\Lambda} = f_d(p_{T_1,v_1}^{\eta_{\Lambda_1}}, \ldots, p_{T_d,v_d}^{\eta_{\Lambda_d}}). \tag{2.6}$$

We call $f_d(\mathbf{x})$ the recursion function since we can use it to compute recursively the marginal probabilities $p_{T,v}^{\tau_\Lambda}$ and $p_{T,v}^{\eta_\Lambda}$.

Lemma 2.2.6. *Let $\lambda > 0, \Delta \geqslant 3$. Consider the hardcore model with parameter λ, and let $\mathbb{T}_\Delta$ be the infinite Δ-regular tree. For $d = 1, \ldots, \Delta$, let $f_d(x_1, \ldots, x_d) = 1 - \frac{1}{1+\lambda \prod_{i=1}^{d}(1-x_i)}$. If there exists a constant $\alpha < 1$ such that*

$$\|\nabla f_d(\mathbf{x})\|_1 = \sum_{i=1}^{d} \left| \frac{\partial f_d(\mathbf{x})}{\partial x_i} \right| \leqslant \alpha$$

for all $d = 1, \ldots, \Delta$ and $\mathbf{x} \in [0, 1]^d$, then $\mathbb{T}_\Delta$ has the correlation decay property.

Proof. To prove the lemma, we will show (2.5) with $C_1 = 1, C_2 = -\ln\alpha$. Let $L = \mathrm{dist}(v, \Psi)$, so that (2.5) becomes

$$\left|p_{T,v}^{\tau_\Lambda} - p_{T,v}^{\eta_\Lambda}\right| \leqslant \alpha^L. \tag{2.7}$$

The proof is by induction on L. The base case $L = 0$ is trivial. We focus on proving the induction step. By the induction hypothesis, we have that for all $i = 1, \ldots, d$, it holds that

$$\left|p_{T_i,v_i}^{\tau_{\Lambda_i}} - p_{T_i,v_i}^{\eta_{\Lambda_i}}\right| \leqslant \alpha^{L-1}. \tag{2.8}$$

Let $\mathbf{p}^\tau := (p_{T_1,v_1}^{\tau_{\Lambda_1}}, \ldots, p_{T_d,v_d}^{\tau_{\Lambda_d}}), \mathbf{p}^\eta := (p_{T_1,v_1}^{\eta_{\Lambda_1}}, \ldots, p_{T_d,v_d}^{\eta_{\Lambda_d}})$. By (2.6), we have that

$$p_{T,v}^{\tau_\Lambda} = f_d(\mathbf{p}^\tau) \text{ and } p_{T,v}^{\eta_\Lambda} = f_d(\mathbf{p}^\eta).$$

Thus, by the Mean Value theorem, there exists $h \in [0,1]$ such that

$$p_{T,v}^{\tau_\Lambda} - p_{T,v}^{\eta_\Lambda} = \nabla f_d(\hat{\mathbf{x}}) \cdot (\mathbf{p}^\tau - \mathbf{p}^\eta), \text{ where } \hat{\mathbf{x}} = h\mathbf{p}^\tau + (1-h)\mathbf{p}^\eta.$$

Since $\mathbf{p}^\tau, \mathbf{p}^\eta \in [0,1]^d$ and $[0,1]$ is a convex set, we have $\hat{\mathbf{x}} \in [0,1]^d$. So,

$$\begin{aligned}
\left|p_{T,v}^{\tau_\Lambda} - p_{T,v}^{\eta_\Lambda}\right| &= |\nabla f_d(\hat{\mathbf{x}}) \cdot (\mathbf{p}^\tau - \mathbf{p}^\eta)| \\
&\leqslant \sum_{i=1}^{d} \left|\frac{\partial f_d(\hat{\mathbf{x}})}{\partial x_i}\right| \cdot \left|p_{T_i,v_i}^{\tau_{\Lambda_i}} - p_{T_i,v_i}^{\eta_{\Lambda_i}}\right| \\
&\leqslant \sum_{i=1}^{d} \left|\frac{\partial f_d(\hat{\mathbf{x}})}{\partial x_i}\right| \cdot \max_{1\leqslant i\leqslant d} \left|p_{T_i,v_i}^{\tau_{\Lambda_i}} - p_{T_i,v_i}^{\eta_{\Lambda_i}}\right| \\
&= \alpha \cdot \max_{1\leqslant i\leqslant d} \left|p_{T_i,v_i}^{\tau_{\Lambda_i}} - p_{T_i,v_i}^{\eta_{\Lambda_i}}\right|,
\end{aligned} \tag{2.9}$$

where in the last inequality we used the assumption of the lemma that $\|\nabla f_d(\hat{\mathbf{x}})\|_1 \leqslant \alpha$. Combining (2.8) and (2.9), we obtain (2.7), which completes the proof. □

Lemma 2.2.6 gives an analytic criterion on the recursion function which allows us to conclude the correlation decay property on the infinite regular tree. However, this criterion $\|\nabla f_d(\mathbf{x})\|_1 \leqslant \alpha < 1$ is too strong and does not actually hold even for the case of the hardcore model. Fortunately, there is an elegant workaround using amortised analysis; this was developed in [53, 54].

Lemma 2.2.7. *Let $\lambda > 0, \Delta \geqslant 3$. Consider the hardcore model with parameter λ, and let $\mathbb{T}_\Delta$ be the infinite Δ-regular tree. For $d = 1, \ldots, \Delta$, let $f_d(x_1, \ldots, x_d) = 1 - \frac{1}{1+\lambda\prod_{i=1}^{d}(1-x_i)}$.*

If there exists a constant $\alpha < 1$ and an increasing function $\phi : [0,1] \to \mathbb{R}$ whose derivative $\Phi = \phi'$ satisfies

$$\sum_{i=1}^{d} \frac{\Phi(f_d(\mathbf{x}))}{\Phi(x_i)} \cdot \left| \frac{\partial f_d(\mathbf{x})}{\partial x_i} \right| \leqslant \alpha$$

for all $d = 1, \ldots, \Delta$ and $\mathbf{x} \in [0,1]^d$, then $\mathbb{T}_\Delta$ has the correlation decay property.

The function $\Phi(\cdot)$ is called *the potential function* in our analysis.

Proof. Analogously to the proof of Lemma 2.2.6, our goal is to show the correlation decay property for every finite subtree T of the Δ-regular tree $\mathbb{T}_\Delta$.
Specifically, we will show (2.5) with $C_1 = (\phi(1) - \phi(0)) \cdot \max_{x \in [0,1]} \frac{1}{\Phi(x)}$ and $C_2 = -\ln \alpha$. Let $L = \operatorname{dist}(v, \Psi)$, so that (2.5) becomes

$$\left| p_{T,v}^{\tau_\Lambda} - p_{T,v}^{\eta_\Lambda} \right| \leqslant \max_{x \in [0,1]} \frac{1}{\Phi(x)} \cdot (\phi(1) - \phi(0)) \cdot \alpha^L. \tag{2.10}$$

To prove (2.10), we will first show

$$\left| \phi(p_{T,v}^{\tau_\Lambda}) - \phi(p_{T,v}^{\eta_\Lambda}) \right| \leqslant (\phi(1) - \phi(0)) \cdot \alpha^L. \tag{2.11}$$

Analogously to the proof of Lemma 2.2.6, we also prove this by induction on L. The base case $L = 0$ is trivial since $\phi(x)$ is an increasing function and $p_{T,v}^{\tau_\Lambda}, p_{T,v}^{\eta_\Lambda} \in [0,1]$. Now we focus on proving the induction step. By the induction hypothesis, we have that for all $i = 1, \ldots, d$, it holds that

$$\left| \phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}}) \right| \leqslant (\phi(1) - \phi(0)) \cdot \alpha^{L-1}. \tag{2.12}$$

Let

$$\mathbf{p}^\tau := (\phi(p_{T_1,v_1}^{\tau_{\Lambda_1}}), \ldots, \phi(p_{T_d,v_d}^{\tau_{\Lambda_d}})), \quad \mathbf{p}^\eta := (\phi(p_{T_1,v_1}^{\eta_{\Lambda_1}}), \ldots, \phi(p_{T_d,v_d}^{\eta_{\Lambda_d}})),$$

and for every $\mathbf{x} = (x_1, \ldots, x_d)$, let

$$g_d(\mathbf{x}) = \phi(f_d(\phi^{-1}(x_1), \phi^{-1}(x_2), \ldots, \phi^{-1}(x_d))).$$

By (2.6), we have that

$$\phi(p_{T,v}^{\tau_\Lambda}) = g_d(\mathbf{p}^\tau) \text{ and } \phi(p_{T,v}^{\eta_\Lambda}) = g_d(\mathbf{p}^\eta).$$

Thus, by the Mean Value Theorem, we can show that there exists some vector $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_d)$ such that

$$\hat{\mathbf{x}} = h\mathbf{p}^\tau + (1-h)\mathbf{p}^\eta \in [\phi(0), \phi(1)]^d \quad \text{for some } h \in [0,1],$$

and

$$\left|\phi(p_{T,v}^{\tau_\Lambda}) - \phi(p_{T,v}^{\eta_\Lambda})\right| \leqslant \sum_{i=1}^{d} \left|\frac{\partial g_d(\hat{\mathbf{x}})}{\partial x_i}\right| \cdot \max_{1\leqslant i\leqslant d} \left|\phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}})\right| .$$

Using the chain rule and the inverse function theorem, we further have

$$\begin{aligned}
\left|\phi(p_{T,v}^{\tau_\Lambda}) - \phi(p_{T,v}^{\eta_\Lambda})\right| &\leqslant \sum_{i=1}^{d} \left|\frac{\partial g_d(\hat{\mathbf{x}})}{\partial x_i}\right| \cdot \max_{1\leqslant i\leqslant d} \left|\phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}})\right| \\
&= \sum_{i=1}^{d} \left|\frac{\partial \phi}{\partial x}\Big|_{x=f_d(\hat{\mathbf{y}})} \cdot \frac{\partial f_d}{\partial x_i}\Big|_{\mathbf{x}=\hat{\mathbf{y}}} \cdot \frac{\partial \phi^{-1}}{\partial x}\Big|_{x=\hat{x}_i}\right| \cdot \max_{1\leqslant i\leqslant d} \left|\phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}})\right| \\
&= \sum_{i=1}^{d} \frac{\Phi(f_d(\hat{\mathbf{y}}))}{\Phi(\phi^{-1}(\hat{x}_i))} \cdot \left|\frac{\partial f_d(\hat{\mathbf{y}})}{\partial x_i}\right| \cdot \max_{1\leqslant i\leqslant d} \left|\phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}})\right| \\
&\leqslant \alpha \cdot \max_{1\leqslant i\leqslant d} \left|\phi(p_{T_i,v_i}^{\tau_{\Lambda_i}}) - \phi(p_{T_i,v_i}^{\eta_{\Lambda_i}})\right| , \qquad (2.13)
\end{aligned}$$

where $\hat{\mathbf{y}} = (\phi^{-1}(\hat{x}_1), \ldots, \phi^{-1}(\hat{x}_d)) \in [0,1]^d$. The last inequality holds by setting $\mathbf{x} = \hat{\mathbf{y}}$ in the assumption that $\sum_{i=1}^{d} \frac{\Phi(f_d(\mathbf{x}))}{\Phi(x_i)} \cdot \left|\frac{\partial f_d(\mathbf{x})}{\partial x_i}\right| \leqslant \alpha$.
The inequalities (2.12) and (2.13) yield (2.11). Moreover, again by the Mean Value Theorem, there exists some number $\tilde{x} \in [0,1]$ such that

$$\left|p_{T,v}^{\tau_\Lambda} - p_{T,v}^{\eta_\Lambda}\right| = \frac{1}{\Phi(\tilde{x})} \left|\phi(p_{T,v}^{\tau_\Lambda}) - \phi(p_{T,v}^{\eta_\Lambda})\right| . \qquad (2.14)$$

Combining (2.11) and (2.14) we obtain (2.10), which completes the proof. □

The analytic criterion in Lemma 2.2.7 is easier to apply than Lemma 2.2.6. In particular, Li, Lu and Yin [54] proved the following.

Lemma 2.2.8. *For all λ, Δ such that $\lambda < \lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$, the condition in Lemma 2.2.7 holds with the function $\phi(x) = \int_0^x \frac{1-t}{\sqrt{t}}\,\mathrm{d}t$ for $x \in [0,1]$.*

Combining Lemmas 2.2.7 and 2.2.8, we can conclude Theorem 2.2.4, i.e., that the infinite tree $\mathbb{T}_\Delta$ has the correlation decay property for all $\lambda < \lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$.

Chapter 3

Uniqueness Threshold for the q-State Antiferromagnetic Potts Model on d-ary Trees

In this chapter, Section 3.2 and 3.3 are based on the joint work with Andreas Galanis and Leslie Goldberg [37], Section 3.4 is based my own work. We will discuss the uniqueness of the q-state antiferromagnetic Potts model on the d-ary trees and prove the uniqueness threshold when $q = 3$.

3.1 Introduction

For spin systems like the Ising model and the Potts model, one of the most well-studied subjects in statistical physics is the uniqueness phase transition on lattice graphs, such as the grid or the regular tree. Note that uniqueness is a strong property, which guarantees that the effect of fixing an arbitrary boundary configuration eventually dies out. As an example, for the antiferromagnetic Ising model on the d-ary tree it is well-known that uniqueness holds iff $\beta \geq \frac{d-1}{d+1}$; the value $\frac{d-1}{d+1}$ is the *uniqueness threshold* because it is the point at which the uniqueness phase transition occurs.

The uniqueness phase transition plays a prominent role in connecting the efficiency of algorithms for sampling from the Gibbs distribution to the properties of the Gibbs distribution itself. One of the first examples of such a connection is in the analysis of the Gibbs sampler Markov chain for the Ising model on the 2-dimensional lattice, where the uniqueness phase transition marks the critical value of β where the mixing time switches from polynomial to exponential (see [58, 59, 73]).

From a computational complexity perspective, it is the uniqueness phase transition on the regular tree which is particularly important. For many 2-state spin models, including the antiferromagnetic Ising model and the hard-core model, it has been proved [70, 72, 38, 54] that the uniqueness phase transition on the tree coincides with a more general computational transition in the complexity of approximating the partition function or sampling from the Gibbs distribution. In the case of the antiferromagnetic Ising model for example, the problem of approximating the partition function on $(d+1)$-regular graphs undergoes a computational transition at the tree uniqueness threshold: it admits a polynomial-time algorithm when $\beta \in (\frac{d-1}{d+1}, 1)$ and it is NP-hard for $\beta \in (0, \frac{d-1}{d+1})$. This connection has been established in full generality for antiferromagnetic 2-state systems. For antiferromagnetic multi-state systems, the situation is much less clear and, in fact, even understanding the uniqueness phase transition on the tree poses major challenges. One of the key reasons behind these difficulties is that certain monotonicities that hold for two-state systems simply do not hold in the multi-state setting, which therefore necessitates far more elaborate techniques. For analysing the uniqueness threshold on the tree, this difficulty has already been illustrated in the case of the q-colourings model, where Jonasson [49], building upon work of Brightwell and Winkler [10], established via a painstaking method that the model is in uniqueness on the d-ary tree iff $q > d+1$. The goal of this chapter is to extend this analysis to the Potts model (beyond the zero-temperature case).

There are several reasons for focusing on establishing uniqueness on the tree. For the colouring model and the antiferromagnetic Potts model, it is widely conjectured that the uniqueness phase transition on the d-ary tree captures the complexity of approximating the partition function on graphs with maximum degree $d+1$, as is the case for antiferromagnetic 2-state models. It has been known since the 80s that non-uniqueness holds for the colourings model when $q \leq d+1$ and for the Potts model when $\beta < 1 - q/(d+1)$, see [67]. More recently, it was shown in [32] that the problem of approximating the partition function is NP-hard when $q < d+1$ for the colourings model and when $\beta < 1 - q/(d+1)$ for the Potts model (for q even). It is not known however whether efficient algorithms can be designed in the complementary regime; for correlation decay algorithms in particular (see [39, 57, 56]), it has been difficult to capture the uniqueness threshold in the analysis — this becomes even harder in the case of the Potts model where uniqueness is not known. For a more direct algorithmic consequence of uniqueness, it has been demonstrated that,

on sparse random graphs, sampling algorithms for the Gibbs distribution can be designed by exploiting the underlying tree-like structure and the decay properties on the tree guaranteed by uniqueness. In particular, in the $G(n, d/n)$ random graph, Efthymiou [29] developed a sampling algorithm for q-colourings when $q > (1+\epsilon)d$, based on Jonasson's uniqueness result. Related results on $G(n, d/n)$ appear in [79, 30, 69, 63]. Also, in Chapter 5, we will describe an application on random regular graphs.

3.1.1 The results

Our results proved the uniqueness threshold for $q = 3, d \geq 2$ and $q = 4, d \in \{4, 5, 6, 7, 8, 9\}$.

Theorem 3.1.1. *Let $q = 3$. When $d \geq 3$, the 3-state Potts model on the d-ary tree has uniqueness for all $\beta \in [1 - 3/(d+1), 1)$. When $d = 2$, the 3-state Potts model on the binary tree has uniqueness for all $\beta \in (0, 1)$.*

Theorem 3.1.2. *Let d be an integer in $\{4, 5, 6, 7, 8, 9\}$. The 4-state Potts model on the d-ary tree has uniqueness for all $\beta \in [1 - 4/(d+1), 1)$.*

Theorem 3.1.1 and 3.1.2 precisely pinpoints the uniqueness threshold for the 3-state Potts model (and 4-state Potts model with $4 \leq d \leq 9$) since it is known that the model is in non-uniqueness in the complementary regime. When $d \geq 3$, non-uniqueness for $\beta < 1 - \frac{q}{d+1}$ follows from the existence of multiple semi-translation-invariant Gibbs measures. When $d = 2$, the 3-state Potts model for $\beta = 0$ corresponds to the 3-colouring model, and non-uniqueness holds in this case because of the existence of so-called *frozen* 3-colourings; in these colourings, the configuration on the leaves determines uniquely the colour of the root, see [10].
The proof of Theorem 3.1.1 and Theorem 3.1.2 will be presented in Chapter 3.3 and Chapter 3.4, respectively. In the next section, we will introduce some basic amount and functions in our proof approach.

3.1.2 Tree recursions

To prove the uniqueness, we are interested in the ratio of marginal probabilities, which is defined as follows:

$$\gamma(q, d, \beta, n) = \max_{\substack{\tau:\Lambda_{\mathbb{T}_{d,n}} \to [q] \\ c_1, c_2 \in [q]}} \frac{\Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c_1 \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau]}{\Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c_2 \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau]}. \tag{3.1}$$

Note that if $\beta > 0$ and $n > 0$, then for every $\tau\colon \Lambda_{\mathbb{T}_{d,n}} \to [q]$ and every $c \in [q]$, $\Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] > 0$. So $\gamma(q,d,\beta,n)$ is well-defined. Suppose, for fixed q, β and d, that $\lim_{n\to\infty}\gamma(q,d,\beta,n) = 1$. This implies that the limsup in the uniqueness definition (Definition 1.2.2) is zero. Thus, this amount is an important tool for us to study the uniqueness threshold.

In order to track the ratio $\gamma(q,d,\beta,n)$, it is a natural idea to establish a recursion among ratios of marginal probability. In this section, we will formulate two appropriate recursions on the infinite d-ary tree which can help us to analyse the limit of ratio $\gamma(q,d,\beta,n)$.

The first one is the one-step recursion.

Proposition 3.1.3 (One-step recursion)**.** *Suppose $q \geq 3$ and $\beta \in (0,1)$. Let T be a tree with root v and leaves Λ. Let $\tau\colon \Lambda \to [q]$ be an arbitrary configuration and $v_1,\dots,v_d$ be the children of v in T. For $i \in [d]$, let T_i be the subtree of T rooted at v_i and let Λ_i denote the set of leaves of the subtree T_i. Then, for any colour $c \in [q]$, it holds that*

$$\Pr_T[\sigma(v) = c \mid \sigma(\Lambda) = \tau] = \frac{\prod_{i=1}^{d}(1-(1-\beta)\Pr_{T_i}[\sigma(v_i) = c \mid \sigma(\Lambda_i) = \tau(\Lambda_i)])}{\sum_{c'=1}^{q}\prod_{i=1}^{d}(1-(1-\beta)\Pr_{T_i}[\sigma(v_i) = c' \mid \sigma(\Lambda_i) = \tau(\Lambda_i)])}. \tag{3.2}$$

Proof. For any graph G (which may be a sub-tree of T), we use $V(G)$ to denote the vertex set of G and we let $w_G(\sigma(v) = c) = \sum_{\sigma'\colon V(G)\to[q],\sigma'(v)=c} w_G(\sigma')$. We have

$$\Pr_T[\sigma(v) = c \mid \sigma(\Lambda) = \tau] = \frac{\Pr_T[\sigma(v) = c \wedge \sigma(\Lambda) = \tau]}{\Pr_T[\sigma(\Lambda) = \tau]} = \frac{w_T(\sigma(v) = c \wedge \sigma(\Lambda) = \tau)}{w_T(\sigma(\Lambda) = \tau)}. \tag{3.3}$$

Now we compute $w_T(\sigma(v) = c \wedge \sigma(\Lambda) = \tau)$:

$$\begin{aligned}
&w_T(\sigma(v) = c \wedge \sigma(\Lambda) = \tau)\\
&\quad= \prod_{i=1}^{d}\left(w_{T_i}(\sigma(v_i) \neq c \wedge \sigma(\Lambda_i) = \tau(\Lambda_i)) + \beta w_{T_i}(\sigma(v_i) = c \wedge \sigma(\Lambda_i) = \tau(\Lambda_i))\right)\\
&\quad= \prod_{i=1}^{d} w_{T_i}(\sigma(\Lambda_i) = \tau(\Lambda_i))\left(1-(1-\beta)\frac{w_{T_i}(\sigma(v_i) = c \wedge \sigma(\Lambda_i) = \tau(\Lambda_i))}{w_{T_i}(\sigma(\Lambda_i) = \tau(\Lambda_i))}\right)\\
&\quad= \prod_{i=1}^{d} w_{T_i}(\sigma(\Lambda_i) = \tau(\Lambda_i))(1-(1-\beta)\Pr_{T_i}[\sigma(v_i) = c \mid \sigma(\Lambda_i) = \tau(\Lambda_i)]).
\end{aligned} \tag{3.4}$$

Also, we have $w_T(\sigma(\Lambda) = \tau) = \sum_{c'=1}^{q} w_T(\sigma(v) = c' \wedge \sigma(\Lambda) = \tau)$. Combining this with (3.3) and (3.4), we complete the proof of (3.2). □

Next we will introduce the two-step recursion. We denote the set of q-dimensional probability vectors by $\triangle$, i.e.,

$$\triangle = \{(p_1, p_2, \ldots, p_q) \colon 0 \leq p_1, p_2, \ldots, p_q \leq 1 \ \wedge \ p_1 + p_2 + \cdots + p_q = 1\}.$$

Suppose that c_1 and c_2 are two colours in $[q]$. We define two functions $g_{c_1,c_2,\beta}$ and $h_{c_1,c_2,\beta}$, indexed by these colours. The argument of each of these functions is a tuple $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)})$ where, for each $j \in [d]$, $\mathbf{p}^{(j)} \in \triangle$. The functions are defined as follows.

$$\begin{aligned} g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) &:= \prod_{k=1}^{d} \left(1 - \frac{(1-\beta)(p_{c_1}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + \sum_{c \neq c_2} p_c^{(k)}}\right). \\ h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) &:= 1 + \frac{(1-\beta)(1 - g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}))}{\beta + \sum_{c \neq c_2} g_{c,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)})}. \end{aligned} \tag{3.5}$$

Note that the functions $g_{c_1,c_2,\beta}$ and $h_{c_1,c_2,\beta}$ are well-defined when $\beta \in (0,1)$ and all of $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}$ have non-negative entries; they are also well-defined when $\beta = 0$ and all of $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}$ have positive entries.

One feature of the functions $g_{c_1,c_2,\beta}$ and $h_{c_1,c_2,\beta}$ which will be important shortly is that they are scale-free. This means that we can multiply each of their arguments by some constant without changing their value, i.e., for scalars $t_1, \ldots, t_d > 0$ it holds that

$$\begin{aligned} g_{c_1,c_2,\beta}(t_1\mathbf{p}^{(1)}, \ldots, t_d\mathbf{p}^{(d)}) &= g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}), \\ h_{c_1,c_2,\beta}(t_1\mathbf{p}^{(1)}, \ldots, t_d\mathbf{p}^{(d)}) &= h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}). \end{aligned} \tag{3.6}$$

The following proposition of two-step tree recursion shows the relevance of these functions for analysing the tree.

Proposition 3.1.4 (Two-step recursion)**.** *Suppose $q \geq 3$, $d \geq 2$ and $\beta \in (0,1)$. For an integer $n \geq 2$, let T be the tree $\mathbb{T}_{d,n}$ with root $z = v_{d,n}$ and leaves $\Lambda = \Lambda_{\mathbb{T}_{d,n}}$. Let $\tau \colon \Lambda \to [q]$ be an arbitrary configuration.*

Let $z_1, \ldots, z_d$ be the children of z in T and, for $i \in [d]$, let $\{z_{i,j}\}_{j \in [d]}$ be the children of z_i. Denote by $T_{i,j}$ the subtree of T rooted at $z_{i,j}$ and by $\Lambda_{i,j}$ the set of leaves of $T_{i,j}$. For $i \in [d]$, $j \in [d]$ and $c \in [q]$, let $r_c^{(i,j)} := \Pr_{T_{i,j}}[\sigma(z_{i,j}) = c \mid \sigma(\Lambda_{i,j}) = \tau(\Lambda_{i,j})]$, and denote by $\mathbf{r}^{(i,j)}$ the vector $\mathbf{r}^{(i,j)} = (r_1^{(i,j)}, \ldots, r_q^{(i,j)})$. Then for any colours $c_1 \in [q]$ and $c_2 \in [q]$ we have

$$\frac{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_1 \mid \sigma(\Lambda) = \tau]}{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_2 \mid \sigma(\Lambda) = \tau]} = \prod_{k=1}^{d} h_{c_1,c_2,\beta}(\mathbf{r}^{(k,1)}, \ldots, \mathbf{r}^{(k,d)}).$$

Proof. For $i \in [d]$ and $c \in [q]$, let T_i be the subtree of T rooted at z_i with leaves Λ_{T_i} and let $r_c^{(i)} = \Pr_{T_i}[\sigma(z_i) = c \mid \sigma(\Lambda_{T_i}) = \tau(\Lambda_{T_i})]$; let $\mathbf{r}^{(i)}$ be the vector $(r_1^{(i)}, \dots, r_q^{(i)})$.
For every $i \in [d]$ and $c \in [q]$ we can apply Proposition 3.1.3 to T_i to obtain

$$r_c^{(i)} = \frac{\prod_{j=1}^{d}(1-(1-\beta)r_c^{(i,j)})}{\sum_{c'=1}^{q}\prod_{j=1}^{d}(1-(1-\beta)r_{c'}^{(i,j)})}.$$

We have $r_c^{(i)} > 0$ for every $c \in [q]$, hence we can apply Proposition 3.1.3 to T to obtain

$$\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c \mid \sigma(\Lambda) = \tau] = \frac{\prod_{i=1}^{d}(1-(1-\beta)r_c^{(i)})}{\sum_{c'=1}^{q}\prod_{i=1}^{d}(1-(1-\beta)r_{c'}^{(i)})}.$$

Thus, for every $i \in [d]$ and $c_1, c_2 \in [q]$,

$$\frac{r_{c_1}^{(i)}}{r_{c_2}^{(i)}} = \prod_{j=1}^{d}\frac{1-(1-\beta)r_{c_1}^{(i,j)}}{1-(1-\beta)r_{c_2}^{(i,j)}} = g_{c_1,c_2,\beta}(\mathbf{r}^{(i,1)}, \dots, \mathbf{r}^{(i,d)}). \tag{3.7}$$

Analogously, for every $c_1, c_2 \in [q]$, we have

$$\frac{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_1 \mid \sigma(\Lambda) = \tau]}{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_2 \mid \sigma(\Lambda) = \tau]} = \prod_{k=1}^{d}\frac{1-(1-\beta)r_{c_1}^{(k)}}{1-(1-\beta)r_{c_2}^{(k)}} = \prod_{k=1}^{d}\left(1 + \frac{(1-\beta)(1-r_{c_1}^{(k)}/r_{c_2}^{(k)})}{\beta + \sum_{c\neq c_2} r_c^{(k)}/r_{c_2}^{(k)}}\right) \tag{3.8}$$

Plugging (3.7) into (3.8), and using the definition of $h_{c_1,c_2,\beta}$ from (3.5), we obtain the statement of the lemma. □

The two-step recursion will allow us to iteratively bootstrap our bounds on the ratio $\gamma(q, d, \beta, n)$. To formalise this, we will use the following definition.

Definition 3.1.5. Suppose $q \geq 3$, $d \geq 2$ and $\beta \in [0, 1)$. For any $\alpha > 1$, let

$$\triangle_\alpha = \{(p_1, \dots, p_q) \in \triangle \colon \max_{i\in[q]} p_i \leq \alpha \min_{j\in[q]} p_j\}.$$

(Note that every vector in $\triangle_\alpha$ has strictly positive entries.) For colours $c_1 \in [q]$ and $c_2 \in [q]$, let

$$M_{\alpha,c_1,c_2,\beta} = \max_{(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})\in\triangle_\alpha^d} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}). \tag{3.9}$$

Since $\triangle_\alpha$ is compact and $h_{c_1,c_2,\beta}$ is continuous, the maximisation in (3.9) is well-defined.

Definition 3.1.5 ensures that $\triangle_\alpha$ is the subset of $\triangle$ induced by probability vectors whose entries are within a factor of $\alpha > 1$ of each other. $M_{\alpha,c_1,c_2,\beta}$ is the maximum of the two-step recursion function $h_{c_1,c_2,\beta}$ when each of its arguments are from $\triangle_\alpha$.

Proposition 3.1.6. *Let $q \geq 3$, $d \geq 2$, $\beta \in (0,1)$ and $\alpha_0 = \max\{\gamma(q,d,\beta,n_0-1), \gamma(q,d,\beta,n_0)\}$ for some integer $n_0 > 0$. Suppose that for all $1 < \alpha \leq \alpha_0$ and any colours $c_1, c_2 \in [q]$, it holds that*

$$M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}.$$

Then, it holds that $\gamma(q,d,\beta,n) \to 1$ as $n \to \infty$, i.e., the q-state Potts model with parameter β has uniqueness on the d-ary tree.

We will prove this key proposition in Section 3.2.3.

3.2 Analysis from the two-step recursion

In this chapter, we will show how to simplify the condition in Proposition 3.1.6 and how to conclude the uniqueness.

3.2.1 A simpler condition for uniqueness

Proposition 3.1.6 gives a sufficient condition on the two-step recursion that is sufficient for establishing uniqueness based on the maximisation of $h_{c_1,c_2,\beta}$. Due to the many variables involved in the maximisation, this is rather complicated for any direct verification. We will simplify this maximisation significantly by showing that it suffices to consider very special vectors whose entries are either equal to α or 1. We start with the following definition of "extremal tuples".

Definition 3.2.1. Let $\alpha > 1$, and consider a colour $c \in [q]$. A tuple $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)})$ is called (α, c)-*extremal* iff for all $k \in [d]$,

- for all $c' \in [q]$, either $p^{(k)}_{c'} = p^{(k)}_{c}$, or $p^{(k)}_{c'} = \alpha \cdot p^{(k)}_{c}$;
- there exists $c' \in [q]$ such that $p^{(k)}_{c'} = \alpha \cdot p^{(k)}_{c}$.

Lemma 3.2.2. *Let $q \geq 3$, $d \geq 2$ and $\beta \in [0,1)$. For any colours $c_1, c_2 \in [q]$, there is an (α, c_2)-extremal tuple which achieves the maximum in $\max_{(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \in \triangle_\alpha^d} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)})$ (cf. (3.9)).*

One of the consequences of Lemma 3.2.2 is that the validity of the inequality in Proposition 3.1.6 is monotone with respect to β. In particular, assuming Lemma 3.2.2 (it will be proved in Section 3.2.2), we have the following lemma.

Lemma 3.2.3. *Let $q \geq 3$, $d \geq 2$ and $\beta', \beta'' \in [0,1)$ with $\beta' \leq \beta''$. Then, for all $\alpha > 1$ and any colours $c_1, c_2 \in [q]$, it holds that*

$$M_{\alpha,c_1,c_2,\beta''} \leq M_{\alpha,c_1,c_2,\beta'}.$$

Proof. Fix $\alpha > 1$ and arbitrary colours $c_1, c_2 \in [q]$. Note that the set of (α, c_2)-extremal tuples does not depend on the parameter β, and for each $\beta \in (0,1)$ there exists by Lemma 3.2.2 an (α, c_2)-extremal tuple which achieves the maximum in

$$\max_{(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \in \triangle_\alpha^d} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}).$$

Therefore, the inequality will follow by showing that

$$h_{c_1,c_2,\beta''}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \leq h_{c_1,c_2,\beta'}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}), \tag{3.10}$$

where $(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})$ is an arbitrary (α, c_2)-extremal tuple. Using the extremality of the tuple $(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})$, we have that $p_{c_2}^{(k)} \leq p_c^{(k)}$ for all colours $c \in [q]$ and every $k \in [d]$. Hence, using the definition (3.5) of $g_{c_1,c_2,\beta}$ and that $\beta' \leq \beta''$, we have

$$1 \geq g_{c,c_2,\beta''}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \geq g_{c,c_2,\beta'}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \text{ for all } c \in [q].$$

In turn, using the definition (3.5) of $h_{c_1,c_2,\beta}$, we obtain from this that (3.10) holds, as wanted. □

Another consequence of Lemma 3.2.2 is that, combined with the scale-free property, it reduces the verification of the condition in Proposition 3.1.6 to the verification of single-variable inequalities in α. These inequalities are obtained by trying all d-tuples of q-dimensional vectors whose entries are as follows.

$$\mathrm{Ex}_c(\alpha) = \big\{(p_1,\dots,p_q) \in \{1,\alpha\}^q \mid p_c = 1 \wedge \exists\, c' \in [q] \text{ such that } p_{c'} = \alpha\big\}. \tag{3.11}$$

The following simplified condition will be our main focus henceforth.

Condition 3.2.4. Let $q \geq 3$, $d \geq 2$. Set $\beta_* := \max\big\{1 - \frac{q}{d+1}, 0\big\}$. For $\alpha > 1$, let $\mathcal{C}(\alpha)$ be the condition

$$\mathcal{C}(\alpha): \quad \forall\, c_1, c_2 \in [q],\ h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) < \alpha^{1/d} \text{ for all } \mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha).$$

If $\mathcal{C}(\alpha)$ holds, we say that the pair (q,d) satisfies Condition 3.2.4 for α.

Now, to verify the inequality in Proposition 3.1.6, we will show shortly that it suffices only to establish Condition 3.2.4 for all $\alpha > 1$, which turns out to be a much more feasible task because of the very explicit form of the set $\mathrm{Ex}_{c_2}(\alpha)$. In the next section, we discuss how to do this in detail, but for now let us state a proposition which asserts that this is indeed sufficient.

Proposition 3.2.5. *Let $q \geq 3$, $d \geq 2$ and $\alpha_0 = \max\{\gamma(q,d,\beta,n_0-1), \gamma(q,d,\beta,n_0)\}$ for some integer $n_0 > 0$. Suppose that the pair (q,d) satisfies Condition 3.2.4 for all $1 < \alpha \leq \alpha_0$. Let $\beta_* = \max\left\{1 - \frac{q}{d+1}, 0\right\}$. Then, the q-state Potts model on the d-ary tree exhibits uniqueness for all $\beta \in (0,1)$ satisfying $\beta \geq \beta_*$.*

Remark. Note that, when $\beta_* > 0$, the conclusion of Proposition 3.2.5 asserts uniqueness in the half-open interval $[\beta_*, 1)$; when $\beta_* = 0$, it instead asserts uniqueness in the open interval $(0,1)$. The case $\beta_* = 0$ is intricate. Brightwell and Winkler [10] showed that q-colourings of the d-ary tree exhibit nonuniqueness if $q \leq d$. So it is known that when $\beta_* = 0$, the q-state Potts model on the d-ary tree exhibit nonuniqueness if $q \leq d$ and $\beta = 0$.

Proof of Proposition 3.2.5. We consider first the case where $\beta_* > 0$. We will show that for all colours $c_1, c_2 \in [q]$, it holds that

$$M_{\alpha,c_1,c_2,\beta_*} < \alpha^{1/d} \text{ for all } \alpha > 1. \tag{3.12}$$

Then by Lemma 3.2.3, we obtain that, for all $\beta \in [\beta_*, 1)$ it holds that $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$ as well for all $\alpha > 1$ and therefore, by Proposition 3.1.6, the Potts model has uniqueness for all such β.

To prove (3.12), consider an arbitrary $\alpha > 1$ and colours $c_1, c_2 \in [q]$. By Lemma 3.2.2, there exists an (α, c_2)-extremal tuple $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)})$ such that

$$M_{\alpha,c_1,c_2,\beta_*} = h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}). \tag{3.13}$$

For $c \in [q]$ and $k \in [d]$, denote by $p_c^{(k)}$ the entry of $\mathbf{p}^{(k)}$ corresponding to colour c and let $\hat{\mathbf{p}}^{(k)}$ be the vector $t_k \mathbf{p}^{(k)}$ where $t_k = 1/p_{c_2}^{(k)}$. By the definition of an (α, c_2)-extremal tuple, we have that

$$\hat{\mathbf{p}}^{(1)}, \ldots, \hat{\mathbf{p}}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha).$$

Moreover, by the scale-free property (3.6) we have that

$$h_{c_1,c_2,\beta_*}(\hat{\mathbf{p}}^{(1)}, \ldots, \hat{\mathbf{p}}^{(d)}) = h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}). \tag{3.14}$$

Finally, since the pair (q, d) satisfies Condition 3.2.4 for all $\alpha > 1$, we have that

$$h_{c_1,c_2,\beta_*}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)}) < \alpha^{1/d}. \tag{3.15}$$

Combining (3.13), (3.14), and (3.15) yields (3.12), as needed.

The case $\beta_* = 0$ is analogous. Now, we need to show that we have uniqueness for all $\beta \in (0, 1)$ assuming that Condition 3.2.4 holds for all $\alpha > 1$. Just as before, we obtain that $M_{\alpha,c_1,c_2,\beta_*} < \alpha^{1/d}$ for all $\alpha > 1$ and hence by Lemma 3.2.3, we have that $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$ for all $\alpha > 1$ and $\beta \in (0, 1)$. Uniqueness for $\beta \in (0, 1)$ therefore follows from applying Proposition 3.1.6. □

3.2.2 Characterising the maximiser of $h_{c_1,c_2,\beta}$

In this section, we prove Lemma 3.2.2 from Section 3.2.1. Recall that

$$\begin{aligned} g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) &:= \prod_{k=1}^{d} \left(1 - \frac{(1-\beta)(p_{c_1}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + \sum_{c \neq c_2} p_c^{(k)}}\right). \\ h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) &:= 1 + \frac{(1-\beta)(1 - g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}))}{\beta + \sum_{c \neq c_2} g_{c,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)})}. \end{aligned} \tag{3.5}$$

To prove Lemma 3.2.2, it will be helpful in this section to consider the set of maximisers of $h_{c_1,c_2,\beta}$.

Definition 3.2.6. Suppose $q \geq 3$, $d \geq 2$ and $\beta \in [0, 1)$. For colours $c_1, c_2 \in [q]$ and $\alpha > 1$, let

$$\mathcal{M}_{\alpha,c_1,c_2,\beta} = \arg\max_{(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) \in \triangle_\alpha^d} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}). \tag{3.16}$$

The following lemmas give properties of the maximisers in $\mathcal{M}_{\alpha,c_1,c_2,\beta}$.

Lemma 3.2.7. *Fix $\alpha > 1$ and $\beta \in [0, 1)$ and colours $c_1, c_2 \in [q]$. Then for any vector $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$, we have $g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \leq 1$.*

Proof. Assume for the sake of contradiction that $g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) > 1$. Then, we have that $h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) < 1$, which contradicts the fact that $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$ since $h_{c_1,c_2,\beta}$ can take the value 1 by setting all of its arguments to be equal to the uniform vector $(1/q, \dots, 1/q) \in \triangle_\alpha$. □

Lemma 3.2.8. *Fix $\alpha > 1$ and $\beta \in [0, 1)$ and any two distinct colours $c_1 \in [q]$ and $c_2 \in [q]$. Then for any vector $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$ and any $k \in [d]$, we have $p_{c_2}^{(k)} = \min_{c \in [q]} \{p_c^{(k)}\}$.*

Proof. Assume for the sake of contradiction that there is $k \in [d]$ and $c \in [q]$ such that $p_{c_2}^{(k)} > p_c^{(k)}$. Define $(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) \in \triangle_\alpha^d$ as follows.

- If $j \neq k$, then $\tilde{\mathbf{p}}^{(j)} = \mathbf{p}^{(j)}$.
- If $c' \notin \{c, c_2\}$, then $\tilde{p}_{c'}^{(k)} = p_{c'}^{(k)}$.
- $\tilde{p}_{c_2}^{(k)} = p_c^{(k)}$.
- $\tilde{p}_c^{(k)} = p_{c_2}^{(k)}$.

The definition of g_{c',c_2} ensures that, for all $c' \neq c_2$, we have

$$g_{c',c_2,\beta}(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) < g_{c',c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}),$$

since the k-th factor in the definition (3.5) of $g_{c_1,c_2,\beta}$ became larger (by switching $\mathbf{p}^{(k)}$ to $\tilde{\mathbf{p}}^{(k)}$). The definition of $h_{c_1,c_2,\beta}$, together with the fact that c_1 and c_2 are distinct and Lemma 3.2.7, implies

$$h_{c_1,c_2,\beta}(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) > h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}),$$

which contradicts the fact that $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$. □

Lemma 3.2.9. *Fix $\alpha > 1$ and $\beta \in [0, 1)$ and any two distinct colours $c_1 \in [q]$ and $c_2 \in [q]$. Then for any vector $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$ and any $k \in [d]$, we have $p_{c_2}^{(k)} < \max_{c\in[q]}\{p_c^{(k)}\}$.*

Proof. For the sake of contradiction, suppose that there is $k \in [d]$ such that $p_{c_2}^{(k)} \geq \max_{c\in[q]}\{p_c^{(k)}\}$. By Lemma 3.2.8, $p_{c_2}^{(k)}$ is the minimum entry of the vector $\mathbf{p}^{(k)}$, so all entries of $\mathbf{p}^{(k)}$ must be equal (and hence, equal to $1/q$). Define the vector $(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) \in \triangle_\alpha^d$ as follows.

- If $j \neq k$, then $\tilde{\mathbf{p}}^{(j)} = \mathbf{p}^{(j)}$.
- $\tilde{p}_{c_1}^{(k)} = \alpha/(\alpha + q - 1)$.
- If $c \neq c_1$, then $\tilde{p}_c^{(k)} = 1/(\alpha + q - 1)$.

The definition of $g_{c,c_2,\beta}$ together with Lemma 3.2.7 ensure that

$$g_{c_1,c_2,\beta}(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) < g_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) = 1$$

and that for every $c \neq c_1$, $g_{c,c_2,\beta}(\tilde{\mathbf{p}}^{(1)}, \ldots, \tilde{\mathbf{p}}^{(d)}) = g_{c,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)})$. The definition of $h_{c_1,c_2,\beta}$ therefore implies that

$$h_{c_1,c_2,\beta}(\tilde{\mathbf{p}}^{(1)}, \ldots, \tilde{\mathbf{p}}^{(d)}) > h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}),$$

which contradicts the fact that $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$. □

To proceed, we will need the following technical fact.

Lemma 3.2.10. *Let A_0, B_0, A_1 and B_1 be real numbers. Let a and b be real numbers satisfying $a \leq b$ such that, for all $x \in [a, b]$, $A_1 + B_1 x \neq 0$. Let $\mathcal{L}(x) = (A_0 + B_0 x)/(A_1 + B_1 x)$. Then*

$$\max_{x \in [a,b]} \mathcal{L}(x) = \max\{\mathcal{L}(a), \mathcal{L}(b)\}.$$

Proof. Since $A_1 + B_1 x \neq 0$ and

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}x} = \frac{A_1 B_0 - A_0 B_1}{(A_1 + B_1 x)^2},$$

$\mathcal{L}(x)$ is a monotone function on $[a, b]$. Thus, $\max_{x \in [a,b]} \mathcal{L}(x) = \max\{\mathcal{L}(a), \mathcal{L}(b)\}$. □

Lemma 3.2.11. *Fix $\alpha > 1$ and $\beta \in [0, 1)$ and two distinct colours c_1 and c_2 in $[q]$. Suppose that $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$. Then for every $k \in [d]$, there exists $\tilde{\mathbf{p}} \in \triangle_\alpha$ such that*

$$(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k-1)}, \tilde{\mathbf{p}}, \mathbf{p}^{(k+1)}, \ldots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$$

and, for all $c \in [q]$, $\tilde{p}_c / \tilde{p}_{c_2} \in \{1, \alpha\}$.

Proof. Fix a tuple $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$. Fix $k \in [d]$. Given any $\hat{\mathbf{p}} \in \triangle_\alpha$, we will be interested in the quantity $h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k-1)}, \hat{\mathbf{p}}, \mathbf{p}^{(k+1)}, \ldots, \mathbf{p}^{(d)})$. Given any $c' \neq c_2$, define

$$P_{c'} := \prod_{j \neq k} \left(1 - \frac{(1-\beta)(p_{c'}^{(j)} - p_{c_2}^{(j)})}{\beta p_{c_2}^{(j)} + \sum_{c \neq c_2} p_c^{(j)}}\right).$$

It will be helpful to re-parameterise the elements of $\hat{\mathbf{p}}$. Recall that the definition of $\triangle_\alpha$ implies that $\hat{p}_c > 0$ for every $c \in [q]$. For every $c \in [q]$, let $\mu_c(\hat{\mathbf{p}}) = \hat{p}_c / \hat{p}_{c_2}$. Let $\boldsymbol{\mu}(\hat{\mathbf{p}})$ be the tuple $\boldsymbol{\mu}(\hat{\mathbf{p}}) = (\mu_1(\hat{\mathbf{p}}), \ldots, \mu_q(\hat{\mathbf{p}}))$. Going the other direction from a tuple $\boldsymbol{\mu}$ with entries in $[1, \alpha]$, for every $c \in [q]$, let $p_c(\boldsymbol{\mu}) = \mu_c / \sum_{c' \in [q]} \mu_{c'}$ and let $\mathbf{p}(\boldsymbol{\mu})$ be the tuple $(p_1(\boldsymbol{\mu}), \ldots, p_q(\boldsymbol{\mu}))$.

It is going to be important to note that the re-parametrisation is without loss of information, so, to this end, let $\Omega_\alpha = \{\boldsymbol{\mu} \in [1, \alpha]^q \mid \mu_{c_2} = 1\}$. The definition of $\triangle_\alpha$

and Lemma 3.2.8 ensure that, for every $(\tilde{\mathbf{p}}^{(1)}, \dots, \tilde{\mathbf{p}}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$ and any $j \in [d]$, $\boldsymbol{\mu}(\tilde{\mathbf{p}}^{(j)}) \in \Omega_\alpha$. Also, given any $\boldsymbol{\mu} \in \Omega_\alpha$, the vector $\mathbf{p}(\boldsymbol{\mu})$ is in $\triangle_\alpha$.

Given a tuple $\boldsymbol{\mu} \in \Omega_\alpha$, we will simplify notation by letting $m(\boldsymbol{\mu}) := \beta + \sum_{c \neq c_2} \mu_c$. Then we can write g_{c',c_2} as

$$\begin{aligned} g_{c',c_2}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k-1)}, \hat{\mathbf{p}}, \mathbf{p}^{(k+1)}, \dots, \mathbf{p}^{(d)}) &= P_{c'} \left(1 - \frac{(1-\beta)(\hat{p}_{c'} - \hat{p}_{c_2})}{\beta \hat{p}_{c_2} + \sum_{c \neq c_2} \hat{p}_c}\right) \\ &= P_{c'} \left(1 - \frac{(1-\beta)(\mu_{c'}(\hat{\mathbf{p}}) - 1)}{m(\boldsymbol{\mu}(\hat{\mathbf{p}}))}\right). \end{aligned}$$

Given the (fixed) values of $\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k-1)}$ and $\mathbf{p}^{(k+1)}, \dots, \mathbf{p}^{(d)}$, let

$$h(\boldsymbol{\mu}) := \frac{m(\boldsymbol{\mu}) - m(\boldsymbol{\mu}) P_{c_1} + P_{c_1}(1-\beta)(\mu_{c_1} - 1)}{\beta m(\boldsymbol{\mu}) + m(\boldsymbol{\mu}) \sum_{c' \neq c_2} P_{c'} - \sum_{c' \neq c_2} P_{c'}(1-\beta)(\mu_{c'} - 1)}.$$

Then we can write $h_{c_1,c_2,\beta}$ as

$$\begin{aligned} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k-1)}, \hat{\mathbf{p}}, \mathbf{p}^{(k+1)}, \dots, \mathbf{p}^{(d)}) &= 1 + \frac{(1-\beta)\left(1 - P_{c_1}\left(1 - \frac{(1-\beta)(\mu_{c_1}(\hat{\mathbf{p}})-1)}{m(\boldsymbol{\mu}(\hat{\mathbf{p}}))}\right)\right)}{\beta + \sum_{c' \neq c_2} P_{c'}\left(1 - \frac{(1-\beta)(\mu_{c'}(\hat{\mathbf{p}})-1)}{m(\boldsymbol{\mu}(\hat{\mathbf{p}}))}\right)} \\ &= 1 + (1-\beta) h(\boldsymbol{\mu}(\hat{\mathbf{p}})). \end{aligned}$$

Since $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}$, taking $\hat{\mathbf{p}} = \mathbf{p}^{(k)}$ maximises

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k-1)}, \hat{\mathbf{p}}, \mathbf{p}^{(k+1)}, \dots, \mathbf{p}^{(d)})$$

over $\triangle_\alpha^d$. Thus, for any maximiser $\boldsymbol{\mu}$ of $h(\boldsymbol{\mu})$ over Ω_α, we have

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k-1)}, \boldsymbol{p}(\boldsymbol{\mu}), \mathbf{p}^{(k+1)}, \dots, \mathbf{p}^{(d)}) \in \mathcal{M}_{\alpha,c_1,c_2,\beta}.$$

So to prove the lemma (taking $\tilde{\mathbf{p}} = \mathbf{p}(\boldsymbol{\mu})$) it suffices to find a maximiser $\boldsymbol{\mu}$ of $h(\boldsymbol{\mu})$ over Ω_α such that for all $c \in [q]$, $\mu_c \in \{1, \alpha\}$. This is what we will do in the rest of the proof. The definition of Ω_α guarantees that $\mu_{c_2} = 1$.

Fix any $c \neq c_2$. For any fixed values $\mu_1, \dots, \mu_{c-1}$ and $\mu_{c+1}, \dots, \mu_q$, all in $[1, \alpha]$, satisfying $\mu_{c_2} = 1$, consider $h(\boldsymbol{\mu})$ as a function of μ_c. Note that both the numerator and denominator of $h(\boldsymbol{\mu})$ are linear in μ_c. We will argue that the denominator is not zero when $\mu_c \in [1, \alpha]$. Using this, Lemma 3.2.10 guarantees that, given $\mu_1, \dots, \mu_{c-1}, \mu_{c+1}, \dots, \mu_q$, $h(\boldsymbol{\mu})$ is maximised by setting $\mu_c \in \{1, \alpha\}$. Going through the colours c one-by-one, we find the desired maximiser $\boldsymbol{\mu}$.

To complete the proof, we just need to show that the denominator of $h(\boldsymbol{\mu})$ is not zero when $\boldsymbol{\mu} \in \Omega_\alpha$. This follows easily, since, for all $c' \neq c_2$, $P_{c'} > 0$ and $m(\boldsymbol{\mu}) \geq (1-\beta)(\mu_{c'} - 1)$. $\square$

Lemmas 3.2.9 and 3.2.11 yield Lemma 3.2.2 as an immediate corollary.

Lemma 3.2.2. *Let $q \geq 3$, $d \geq 2$ and $\beta \in [0,1)$. For any colours $c_1, c_2 \in [q]$, there is an (α, c_2)-extremal tuple which achieves the maximum in $\max_{(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})\in\triangle_\alpha^d} h_{c_1,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})$ (cf. (3.9)).*

3.2.3 Concluding the uniqueness threshold

In this section, we will give the proof of Proposition 3.1.6. We first introduce the following proposition:

Proposition 3.2.12. *Let $q \geq 3$, $d \geq 2$ and $\beta \in (0,1)$. Suppose that, for some integer $n \geq 3$ and some $\alpha > 1$, it holds that $\gamma(q,d,\beta,n-2) = \alpha$ and $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$ for all colours $c_1, c_2 \in [q]$. Then $\gamma(q,d,\beta,n) \leq (M_{\alpha,c_1,c_2,\beta})^d < \gamma(q,d,\beta,n-2)$.*

Proof. Consider the tree $\mathbb{T}_{d,n}$ with root $z = v_{d,n}$ and leaves $\Lambda = \Lambda_{\mathbb{T}_{d,n}}$. Let $\tau : \Lambda_{\mathbb{T}_{d,n}} \to [q]$ be an arbitrary configuration. As in Proposition 3.1.4, let $\{z_{i,j}\}_{i,j\in[d]}$ denote the grandchildren of the root, let $T_{i,j}$ be the subtree of T rooted at $z_{i,j}$, and let $\Lambda_{i,j}$ be the set of leaves of $T_{i,j}$. Further, let $\mathbf{r}^{(i,j)}$ be the marginal distribution at $z_{i,j}$ in the subtree $T_{i,j}$, conditioned on the configuration $\tau(\Lambda_{i,j})$.
By the assumption $\gamma(q,d,\beta,n-2) = \alpha$ and the definition (3.1) of the ratio $\gamma(q,d,\beta,n-2)$, we have that $\mathbf{r}^{(i,j)} \in \triangle_\alpha$ for all $i,j \in [d]$. Proposition 3.1.4 also guarantees that for colours $c_1 \in [q]$ and $c_2 \in [q]$ we have

$$\frac{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_1 \mid \sigma(\Lambda) = \tau]}{\Pr_{\mathbb{T}_{d,n}}[\sigma(z) = c_2 \mid \sigma(\Lambda) = \tau]} = \prod_{k=1}^{d} h_{c_1,c_2,\beta}(\mathbf{r}^{(k,1)},\dots,\mathbf{r}^{(k,d)}) \leq (M_{\alpha,c_1,c_2,\beta})^d < \alpha,$$

where the strict inequality follows by the assumption that $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$. Since τ was an arbitrary configuration on the leaves Λ, we obtain that $\gamma(q,d,\beta,n) < \gamma(q,d,\beta,n-2)$ as needed. □

Now we prove Proposition 3.1.6, which we restate here for convenience.

Proposition 3.1.6. *Let $q \geq 3$, $d \geq 2$, $\beta \in (0,1)$ and $\alpha_0 = \max\{\gamma(q,d,\beta,n_0-1), \gamma(q,d,\beta,n_0)\}$ for some integer $n_0 > 0$. Suppose that for all $1 < \alpha \leq \alpha_0$ and any colours $c_1, c_2 \in [q]$, it holds that*

$$M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}.$$

Then, it holds that $\gamma(q,d,\beta,n) \to 1$ as $n \to \infty$, i.e., the q-state Potts model with parameter β has uniqueness on the d-ary tree.

Proof. Fix q, d and β as in the statement. For all $n \geq 1$, let $\alpha_n := \gamma(q,d,\beta,n_0+n)$. We may assume that $\alpha_0 > 1$ (otherwise, uniqueness follows trivially by applying Proposition 3.1.3 repeatedly to show $\alpha_n = 1$ for all $n \geq n_0$.)
Using Proposition 3.2.12 and the assumption that $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$ for all $1 < \alpha \leq \alpha_0$ and colours $c_1, c_2 \in [q]$, we obtain that

$$1 < \alpha_n < \alpha_{n-2} \leq \alpha_0. \tag{3.17}$$

This implies that both of the sequences $\{\alpha_{2n}\}$ and $\{\alpha_{2n+1}\}$ are decreasing. Since both of these sequences is bounded below by 1, we obtain that for $n \to \infty$ it holds that[1]

$$\alpha_{2n} \downarrow \alpha_{\mathsf{ev}}, \quad \alpha_{2n+1} \downarrow \alpha_{\mathsf{od}}$$

for some $\alpha_{\mathsf{ev}}, \alpha_{\mathsf{od}} \geq 1$. We claim that in fact both of $\alpha_{\mathsf{ev}}, \alpha_{\mathsf{od}}$ are equal to 1, which proves that $\gamma(q,d,\beta,n) \to 1$ as $n \to \infty$.
Suppose for the sake of contradiction that $\alpha_{\mathsf{ev}} > 1$ (a similar argument applies for α_{od}). Let $\mathbf{m}_{2n} = (\mathbf{p}_{2n}^{(1)}, \ldots, \mathbf{p}_{2n}^{(d)})$ achieve the maximum in (3.9) for $\alpha = \alpha_{2n}$, i.e.,

$$M_{\alpha_{2n},c_1,c_2,\beta} = h_{c_1,c_2,\beta}(\mathbf{p}_{2n}^{(1)}, \ldots, \mathbf{p}_{2n}^{(d)}).$$

Note that for all $n \geq 1$ we have that

$$h_{c_1,c_2,\beta}(\mathbf{p}_{2n}^{(1)}, \ldots, \mathbf{p}_{2n}^{(d)}) \geq (\alpha_{\mathsf{ev}})^{1/d}; \tag{3.18}$$

otherwise, we would have that $M_{\alpha_{2n},c_1,c_2,\beta} < (\alpha_{\mathsf{ev}})^{1/d}$ and hence, by Proposition 3.2.12, we would have that $\alpha_{2n+2} < \alpha_{\mathsf{ev}}$, contradicting that $\alpha_{2n} \downarrow \alpha_{\mathsf{ev}}$.
Moreover, observe that $\mathbf{m}_{2n}$ belongs to the compact space $\triangle^d$ for all $n \geq 1$ and therefore there exists a subsequence $\{n_k\}_{k\geq 1}$ and $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) \in \triangle^d$ such that

$$\mathbf{m}_{2n_k} \to (\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}),$$

In fact, since $\{\alpha_{2n_k}\}_{k\geq 1}$ is a subsequence of the convergent sequence $\{\alpha_{2n}\}_{n\geq 1}$, we have that the sequence $\{\alpha_{2n_k}\}$ converges to α_{ev} as well. From $\mathbf{m}_{2n_k} \in \triangle^d_{\alpha_{2n_k}}$, we

[1] The notation $\alpha_{2n} \downarrow \alpha_{\mathsf{ev}}$ means that the sequence α_{2n} converges to α_{ev} by decreasing monotonically.

therefore obtain that $(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \in \triangle^d_{\alpha_{\mathsf{ev}}}$. Applying the assumption $M_{\alpha,c_1,c_2,\beta} < \alpha^{1/d}$ for $\alpha = \alpha_{\mathsf{ev}}$, we therefore have that

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) < (\alpha_{\mathsf{ev}})^{1/d}. \tag{3.19}$$

Since the function $h_{c_1,c_2,\beta}$ is continuous on $\triangle^d$ for all $\beta \in (0,1)$, we have that as $k \to \infty$

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}_{2n_k}, \dots, \mathbf{p}^{(d)}_{2n_k}) \to h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}).$$

This contradicts (3.18) and (3.19). Therefore, $\alpha_{\mathsf{ev}} = 1$, and similarly $\alpha_{\mathsf{od}} = 1$, completing the proof. □

3.3 Uniqueness threshold for $q = 3$

In this chapter, we will introduce our proof of Theorem 3.1.1, which establishes the uniqueness threshold for the 3-state antiferromagnetic Potts model.
The proof is divided into two parts. First, in Chapter 3.3.1, 3.3.2, 3.3.3 and 3.3.4, we use the one-step recursion to give an upper bound α^* for the ratio between marginal probabilities. Then, in Chapter 3.3.5, we fix $c_1, c_2 \in [q]$ and for all $1 < \alpha \leq \alpha^*$ verify Condition 3.2.4 for all (α, c_2)-extremal tuples. Using Lemmas 3.2.2, 3.2.3 and Proposition 3.1.6 we can conclude our result.

3.3.1 Bounding the ratio of marginal probabilities by the one-step recursion

First, we use the one-step recursion to give an upper bound for $\gamma(3, d, \beta, n)$. In particular, we prove the following lemma.

Lemma 3.3.1. *Let $q = 3$ and $c \in [3]$ be an arbitrary colour. For $d \geq 2$, consider the d-ary tree $\mathbb{T}_{d,n}$ with height n and let $\tau : \Lambda_{\mathbb{T}_{d,n}} \to [3]$ be an arbitrary configuration on the leaves. When $d = 2$, for all $\beta \in (0,1)$, for all sufficiently large n it holds that*

$$\frac{459}{2000} \leq \Pr_{\mathbb{T}_{2,n}}[\sigma(v_{2,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{2,n}}) = \tau] \leq \frac{1107}{2500}.$$

When $d \geq 3$, for all $\beta \in [1 - \frac{3}{d+1}, 1)$, there exist sequences $\{L_n\}$ and $\{U_n\}$ (depending on d and β) such that for all sufficiently large n

$$L_n \leq \Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] \leq U_n \text{ and } U_n/L_n \leq 53/27.$$

The following corollary is an immediate consequence of Lemma 3.3.1.

Corollary 3.3.2. *Let $q = 3$, $d \geq 2$, $\beta \in (0,1)$ and $\beta \geq \beta_* = 1 - q/(d+1)$. Then there exists an integer n_0 such that, for all $n \geq n_0$, $\gamma(q,d,\beta,n) \leq 53/27$.*

We begin by giving an upper and a lower bound for the marginal probability that the root is assigned a colour c via the one-step recursion (see the upcoming Lemma 3.3.4).
First we define two functions. Let

$$f_{\mathsf{upper}}(d,\beta,x,y) = \frac{\big(1-(1-\beta)y\big)^d}{\big(1-(1-\beta)y\big)^d + 2\big(1-(1-\beta)x\big)^{d/2}\big(1-(1-\beta)(1-x-y)\big)^{d/2}}$$

and

$$f_{\mathsf{lower}}(d,\beta,x,y) = \frac{\big(1-(1-\beta)x\big)^d}{\big(1-(1-\beta)x\big)^d + \big(1-(1-\beta)y\big)^d + \big(1-(1-\beta)(1-x-y)\big)^d}.$$

We will use the following lemma.

Lemma 3.3.3. *Let f be a convex function on an interval $I = [a,b]$.*

1. *Given $\rho \in [2a, a+b]$, the function $g(x) = f(x) + f(\rho - x)$ is decreasing on $J = [a, \rho/2]$.*
2. *Given $\rho \in [a+b, 2b]$, the function $g(x) = f(x) + f(\rho - x)$ is increasing on $J = [\rho/2, b]$.*

Proof. We first prove Item 1. Suppose $y, z \in J$ satisfy $y < z$, we will show that $g(z) \leq g(y)$. We have $y < z \leq \rho - z < \rho - y$ and $y \geq a$, $\rho - y \leq b$ (using $\rho \leq a + b$). It follows that all of $y, z, \rho - y, \rho - z$ belong to I. Moreover, by the convexity of f on I, we conclude that the slope of f in the interval $[\rho - z, \rho - y]$ is greater or equal to the slope of f in the interval $[y, z]$, i.e., $\frac{f(\rho-y)-f(\rho-z)}{(\rho-y)-(\rho-z)} \geq \frac{f(z)-f(y)}{z-y}$. Re-arranging, we obtain $g(z) \leq g(y)$.
The proof of Item 2 is analogous. For $y, z \in J$ satisfying $y < z$, we have that $\rho - z < \rho - y \leq y < z$ and all of $y, z, \rho - y, \rho - z$ belong to I (using $\rho \geq a + b$). By the convexity of f on I, we conclude that the slope of f in the interval $[\rho - z, \rho - y]$ is less or equal to the slope of f in the interval $[y, z]$, which gives that $g(z) \geq g(y)$. □

The following lemma gives recursively-generated bounds on the probability that the root of $\mathbb{T}_{d,n}$ is a given colour.

Lemma 3.3.4. *Suppose $q = 3$, $d \geq 2$ and $\beta \in [0,1]$. For any $n \geq 0$, let $T = \mathbb{T}_{d,n}$ with root $v = v_T$ and leaves $\Lambda = \Lambda_T$. Let $v_1, \ldots, v_d$ be the children of v in T, T_i be the subtree of T*

rooted at v_i and Λ_i denote the set of leaves of the subtree T_i. Consider any configuration $\tau\colon \Lambda \to [q]$ and any real numbers $L, U \in [0,1]$ such that, for all $i \in [d]$ and all $j \in [3]$ we have

$$L \leq \Pr_{T_i}[\sigma(v_i) = j \mid \sigma(\Lambda_i) = \tau(\Lambda_i)] \leq U.$$

Then, for all colours $c \in [q]$, we also have

$$f_{\mathsf{lower}}(d, \beta, U, L) \leq \Pr_T[\sigma(v) = c \mid \sigma(\Lambda) = \tau] \leq f_{\mathsf{upper}}(d, \beta, U, L).$$

Proof. By symmetry between the colours, we may assume that $c = 1$. Let $p = \Pr_T[\sigma(v) = 1 \mid \sigma(\Lambda) = \tau]$. For any colour $c' \in [3]$ and any child $i \in [d]$, let $p_{i,c'} = \Pr_{T_i}[\sigma(v_i) = c' \mid \sigma(\Lambda_i) = \tau(\Lambda_i)]$.

For convenience, let $\hat{\beta} := 1 - \beta \in [0,1]$. By Proposition 3.1.3, with $q = 3$ and $c = 1$, we have that

$$p = \frac{1}{1+R}, \text{ where } R := \frac{\sum_{c'=2}^{3} \prod_{i=1}^{d} \left(1 - \hat{\beta} p_{i,c'}\right)}{\prod_{i=1}^{d} \left(1 - \hat{\beta} p_{i,1}\right)}. \tag{3.20}$$

We first show that $p \geq f_{\mathsf{lower}}(d, \beta, U, L)$. Since $p_{i,1} \leq U$ for every $i \in [d]$, we obtain that

$$R \leq \frac{\sum_{c'=2}^{3} \prod_{i=1}^{d} \left(1 - \hat{\beta} p_{i,c'}\right)}{\left(1 - \hat{\beta} U\right)^d}. \tag{3.21}$$

For $c' \in [3]$, let $\bar{p}_{c'}$ denote the mean $\frac{1}{d}\sum_{i \in [d]} p_{i,c'}$. The function $f(x) = \ln(1 - \hat{\beta} x)$ is concave on the interval $[0,1]$ so by Jensen's inequality

$$\frac{1}{d}\sum_{i=1}^{d} \ln(1 - \hat{\beta} p_{i,c'}) \leq \ln(1 - \hat{\beta}\bar{p}_{c'}).$$

Thus, for $c' \in \{2,3\}$ we have $\prod_{i=1}^{d}\left(1 - \hat{\beta} p_{i,c'}\right) \leq \left(1 - \hat{\beta}\bar{p}_{c'}\right)^d$ which implies

$$\sum_{c'=2}^{3} \prod_{i=1}^{d} \left(1 - \hat{\beta} p_{i,c'}\right) \leq \left(1 - \hat{\beta}\bar{p}_2\right)^d + \left(1 - \hat{\beta}\bar{p}_3\right)^d. \tag{3.22}$$

Let $f(x) = (1 - \hat{\beta} x)^d$. Let $a = 0, b = 1, \rho = \bar{p}_2 + \bar{p}_3$ and consider the interval $I = [a,b]$. Since f is convex on the interval I and $\rho \in [2a, a+b]$, Item 1 of Lemma 3.3.3 implies that the function $g(x) = f(x) + f(\rho - x)$ is decreasing on $J = [a, \rho/2] = [0, \rho/2]$. Since $L \leq \bar{p}_2$ and $L \leq \bar{p}_3$, the values L and $\min\{\bar{p}_2, \bar{p}_3\}$ are in J and $\min\{\bar{p}_2, \bar{p}_3\} \geq L$, so $g(\min\{\bar{p}_2, \bar{p}_3\}) \leq g(L)$, i.e.,

$$\left(1 - \hat{\beta}\bar{p}_2\right)^d + \left(1 - \hat{\beta}\bar{p}_3\right)^d \leq \left(1 - \hat{\beta} L\right)^d + \left(1 - \hat{\beta}(\bar{p}_2 + \bar{p}_3 - L)\right)^d.$$

Since, for every $i \in [d]$, $p_{i,2} + p_{i,3} = 1 - p_{i,1} \geq 1 - U$, we have $\bar{p}_2 + \bar{p}_3 \geq 1 - U$, so

$$\left(1 - \hat{\beta}\bar{p}_2\right)^d + \left(1 - \hat{\beta}\bar{p}_3\right)^d \leq \left(1 - \hat{\beta}L\right)^d + \left(1 - \hat{\beta}(1 - U - L)\right)^d.$$

Plugging this into (3.22) and then into (3.21), we obtain that

$$R \leq \frac{\left(1 - \hat{\beta}L\right)^d + \left(1 - \hat{\beta}(1 - U - L)\right)^d}{\left(1 - \hat{\beta}U\right)^d}.$$

Therefore, using (3.20), we obtain the lower bound $p \geq f_{\mathsf{lower}}(d, \beta, U, L)$.
Next, we show that $p \leq f_{\mathsf{upper}}(d, \beta, U, L)$. To give an upper bound on p, it suffices to lower bound R. Since $p_{i,1} \geq L$ for every $i \in [d]$, we obtain the lower bound

$$R \geq \frac{\sum_{c'=2}^{3} \prod_{i=1}^{d} \left(1 - \hat{\beta}p_{i,c'}\right)}{\left(1 - \hat{\beta}L\right)^d}. \tag{3.23}$$

Using the arithmetic-mean geometric-mean inequality we have

$$\sum_{c'=2}^{3} \prod_{i=1}^{d} \left(1 - \hat{\beta}p_{i,c'}\right) \geq 2 \prod_{i=1}^{d} \left(\left(1 - \hat{\beta}p_{i,2}\right)\left(1 - \hat{\beta}p_{i,3}\right)\right)^{1/2}. \tag{3.24}$$

Now let $f(x) = -\ln(1 - \hat{\beta}x)$ and consider an arbitrary $i \in [d]$. Let $a = -1, b = 1, \rho = p_{i,2} + p_{i,3}$ and consider the interval $I = [a, b]$. Since f is convex on I and $\rho \in [a + b, 2b]$, Item 2 of Lemma 3.3.3 implies that the function $g(x) = f(x) + f(\rho - x)$ is increasing on the interval $J = [\rho/2, b] = [\rho/2, 1]$. Let $x = U$ and $y = \max\{p_{i,2}, p_{i,3}\}$. Since $p_{i,2}$ and $p_{i,3}$ are at most U, x and y are in J and satisfy $x \geq y$. Therefore, $g(x) \geq g(y)$, which gives that

$$\ln\left(1 - \hat{\beta}y\right) + \ln\left(1 - \hat{\beta}(\rho - y)\right) \geq \ln\left(1 - \hat{\beta}x\right) + \ln\left(1 - \hat{\beta}(\rho - x)\right).$$

Thus, by substituting in the values of x, y and ρ and exponentiating, we have

$$\left(1 - \hat{\beta}p_{i,2}\right)\left(1 - \hat{\beta}p_{i,3}\right) \geq \left(1 - \hat{\beta}U\right)\left(1 - \hat{\beta}(p_{i,2} + p_{i,3} - U)\right).$$

Using the inequality $p_{i,2} + p_{i,3} = 1 - p_{i,1} \leq 1 - L$ in the right-hand side, we get

$$\left(1 - \hat{\beta}p_{i,2}\right)\left(1 - \hat{\beta}p_{i,3}\right) \geq \left(1 - \hat{\beta}U\right)\left(1 - \hat{\beta}(1 - U - L)\right).$$

Plugging this into (3.24) for each $i \in [d]$ and then into (3.23), we obtain that

$$R \geq \frac{2\left(1 - \hat{\beta}U\right)^{d/2}\left(1 - \hat{\beta}(1 - U - L)\right)^{d/2}}{\left(1 - \hat{\beta}L\right)^d}.$$

Therefore, using (3.20), we obtain the upper bound $p \leq f_{\mathsf{upper}}(d, \beta, U, L)$. □

3.3.2 Properties of the functions f_{upper} and f_{lower}

In this section, we establish useful monotonicity properties of the functions f_{upper} and f_{lower} that will be relevant later.

Lemma 3.3.5. *For any fixed $d \geq 2$ and any fixed x and y satisfying $0 \leq y \leq x \leq 1$ and $2y + x \leq 1 \leq 2x + y$,*

- *$f_{\mathsf{upper}}(d, \beta, x, y)$ is a decreasing function of β on the interval $(0, 1)$ and*
- *$f_{\mathsf{lower}}(d, \beta, x, y)$ is an increasing function of β on the interval $(0, 1)$.*

Proof. Let $\hat{\beta} := 1-\beta$, and $W = 1-3y+\hat{\beta}\big((3y-1)+(1-x-2y)+x(2x+y-1)+y(x-y)\big)$. The derivative of f_{upper} with respect to β is given by

$$\frac{\partial f_{\mathsf{upper}}}{\partial \beta} = -{f_{\mathsf{upper}}}^2 \frac{d(1-\hat{\beta}x)^{d/2}(1-\hat{\beta}(1-x-y))^{d/2}W}{(1-\hat{\beta}y)^{d+1}(1-\hat{\beta}x)(1-\hat{\beta}(1-x-y))}.$$

(Obviously, this can be checked directly, but the reader may prefer to use the Mathematica code in Section 8.1.1 to check this and the derivative of f_{lower} with respect to β, which appears below.) Using the conditions on x and y in the statement of the lemma, we find that

$$W \geq 1 - 3y + \hat{\beta}(3y - 1) = \beta(1 - 3y) \geq 0.$$

We conclude that $\frac{\partial f_{\mathsf{upper}}}{\partial \beta} \leq 0$ so $f_{\mathsf{upper}}(d, \beta, x, y)$ is a decreasing function of β on the interval $[0, 1]$. Similarly,

$$\frac{\partial f_{\mathsf{lower}}}{\partial \beta} = {f_{\mathsf{lower}}}^2 \frac{d\left((2x + y - 1)(1 - \hat{\beta}(1 - x - y))^{d-1} + (x - y)(1 - \hat{\beta}y)^{d-1}\right)}{(1 - \hat{\beta}x)^{d+1}} \geq 0,$$

so $f_{\mathsf{lower}}(d, \beta, x, y)$ is an increasing function of β on the interval $(0, 1)$. □

Lemma 3.3.6. *For any fixed $d \geq 2$ and $0 < \beta \leq 1$,*

1. *$f_{\mathsf{lower}}(d, \beta, x, y)$ is a decreasing function of x when $x, y \in [0, 1]$, and*
2. *$f_{\mathsf{lower}}(d, \beta, x, y)$ is an increasing function of y when $x + 2y \leq 1$ and $x, y \in [0, 1]$.*

Proof. Let

$$R := \frac{\big(1 - (1 - \beta)y\big)^d + \big(1 - (1 - \beta)(1 - x - y)\big)^d}{\big(1 - (1 - \beta)x\big)^d}, \text{ so that } f_{\mathsf{lower}}(d, \beta, x, y) = \frac{1}{1 + R}.$$

We first prove Item 2. Let $a = 0$, $b = 1$, $\rho = 1 - x$ and consider the interval $I = [a, b]$. Since $f(y) = (1 - (1 - \beta)y)^d$ is convex on I and $\rho \in [2a, a + b]$, Item 1 of Lemma 3.3.3 yields that $g(y) = f(y) + f(\rho - y)$ is decreasing on $J = [a, \rho/2] = [0, \rho/2]$. It follows that, for fixed x, R is a decreasing function of y on J and therefore f_{lower} is increasing in y. It remains to observe that, for $x \in [0, 1]$, the condition $y \in J$ is equivalent to the condition $x + 2y \leq 1$ and $y \in [0, 1]$ in the statement.

For Item 1, note that $1 - (1 - \beta)(1 - x - y)$ is an increasing nonnegative function of x and $1 - (1 - \beta)x$ is a decreasing nonnegative function of x, so R is an increasing function of x. Thus, f_{lower} is a decreasing function of x. □

Lemma 3.3.7. *For any fixed $d \geq 2$ and $0 < \beta \leq 1$,*

1. *$f_{\mathsf{upper}}(d, \beta, x, y)$ is an increasing function of x when $1 \leq 2x + y$ and $x, y \in [0, 1]$, and*
2. *$f_{\mathsf{upper}}(d, \beta, x, y)$ is a decreasing function of y when $x, y \in [0, 1]$.*

Proof. The proof is analogous to that of Lemma 3.3.6. Let

$$R := \frac{2\big(1 - (1 - \beta)x\big)^{d/2}\big(1 - (1 - \beta)(1 - x - y)\big)^{d/2}}{\big(1 - (1 - \beta)y\big)^d}, \text{ so that } f_{\mathsf{upper}}(d, \beta, x, y) = \frac{1}{1 + R}.$$

We first prove Item 1. Let $a = -y$, $b = 1$, $\rho = 1 - y$ and consider the interval $I = [a, b]$. Since the function $f(x) = -\ln\big(1-(1-\beta)x\big)$ is convex on the interval I and $\rho \in [a+b, 2b]$, by Item 2 of Lemma 3.3.3, the function $g(x) = f(x)+f(\rho-x)$ is increasing on the interval $J = [\rho/2, b]$. It follows that the function

$$\exp(-g(x)) = \big(1 - (1 - \beta)x\big)\big(1 - (1 - \beta)(1 - x - y)\big)$$

is a decreasing function of x on J, and therefore R has the same property as well. Thus, f_{upper} is an increasing function of x on the interval J. It remains to observe that, for $y \in [0, 1]$, the condition $x \in J$ is equivalent to the condition $1 \leq 2x + y$ and $x \in [0, 1]$ in the statement.

For Item 2, note that $1 - (1 - \beta)(1 - x - y)$ is an increasing nonnegative function of y and $1 - (1 - \beta)y$ is a decreasing nonnegative function of y, so R is an increasing function of y. Thus, f_{upper} is a decreasing function of y. □

Lemma 3.3.8. *For any fixed $d \geq 2$, $0 < \beta < 1$ and $0 \leq y \leq x \leq 1$ such that $2y + x \leq 1 \leq 2x + y$, we have*

$$2f_{\mathsf{lower}}(d, \beta, x, y) + f_{\mathsf{upper}}(d, \beta, x, y) \leq 1 \leq 2f_{\mathsf{upper}}(d, \beta, x, y) + f_{\mathsf{lower}}(d, \beta, x, y).$$

Proof. Since $2y + x \le 1 \le 2x + y$, we obtain that $y \le 1 - x - y \le x$. Further, by the AM-GM inequality,

$$\big(1-(1-\beta)x\big)^d + \big(1-(1-\beta)(1-x-y)\big)^d \ge 2\big(1-(1-\beta)x\big)^{d/2}\big(1-(1-\beta)(1-x-y)\big)^{d/2}.$$

So the denominator in the definition of $f_{\mathsf{lower}}(d,\beta,x,y)$ is at least as big as the denominator in the definition of $f_{\mathsf{upper}}(d,\beta,x,y)$. We conclude that

$$2f_{\mathsf{lower}} + f_{\mathsf{upper}} \le \frac{2\big(1-(1-\beta)x\big)^d + \big(1-(1-\beta)y\big)^d}{\big(1-(1-\beta)y\big)^d + 2\big(1-(1-\beta)x\big)^{d/2}\big(1-(1-\beta)(1-x-y)\big)^{d/2}} \le 1,$$

and

$$2f_{\mathsf{upper}} + f_{\mathsf{lower}} \ge \frac{2\big(1-(1-\beta)y\big)^d + \big(1-(1-\beta)x\big)^d}{\big(1-(1-\beta)x\big)^d + \big(1-(1-\beta)y\big)^d + \big(1-(1-\beta)(1-x-y)\big)^d} \ge 1.$$

□

Lemma 3.3.9. *For any fixed $d \ge 2$, $0 < \beta < 1$ and $0 \le y_1 \le y_2 \le x_2 \le x_1 \le 1$ such that $2y_1 + x_1 \le 1 \le 2x_1 + y_1$ and $2y_2 + x_2 \le 1 \le 2x_2 + y_2$, we have*

$$f_{\mathsf{upper}}(d,\beta,x_2,y_2) \le f_{\mathsf{upper}}(d,\beta,x_1,y_1) \quad \textit{and} \quad f_{\mathsf{lower}}(d,\beta,x_2,y_2) \ge f_{\mathsf{lower}}(d,\beta,x_1,y_1).$$

Proof. Using the assumptions in the statement of the lemma, we obtain $1 \le 2x_1 + y_2$ and $2y_1 + x_2 \le 1$. Therefore, by Lemmas 3.3.6 and 3.3.7, we obtain

$$\begin{aligned} &f_{\mathsf{upper}}(d,\beta,x_2,y_2) \le f_{\mathsf{upper}}(d,\beta,x_1,y_2) \le f_{\mathsf{upper}}(d,\beta,x_1,y_1), \text{ and} \\ &f_{\mathsf{lower}}(d,\beta,x_2,y_2) \ge f_{\mathsf{lower}}(d,\beta,x_2,y_1) \ge f_{\mathsf{lower}}(d,\beta,x_1,y_1). \end{aligned}$$

□

3.3.3 Bounding the marginal probability at the root by two sequences

For any $\beta > 0$ and $d \ge 2$, we define two sequences:

$$\begin{cases} u_0(d,\beta) = 1, \\ \ell_0(d,\beta) = 0, \end{cases}$$

and for every non-negative integer n,

$$\begin{cases} u_{n+1}(d,\beta) = f_{\mathsf{upper}}(d,\beta,u_n(d,\beta),\ell_n(d,\beta)), \text{and} \\ \ell_{n+1}(d,\beta) = f_{\mathsf{lower}}(d,\beta,u_n(d,\beta),\ell_n(d,\beta)). \end{cases} \tag{3.25}$$

Our interest in the sequences $u_n(d,\beta)$ and $\ell_n(d,\beta)$ is that they give upper and lower bounds on the probability $\Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c]$, respectively (subject to any boundary configuration at the leaves).

Lemma 3.3.10. *Suppose that $q = 3$, $d \geq 2$, and $\beta \in (0,1)$. For any $n \geq 0$, for the d-ary tree $\mathbb{T}_{d,n}$ with depth n and root $v_{d,n}$, for any configuration $\tau\colon \Lambda_{\mathbb{T}_{d,n}} \to [q]$ on the leaves and any colour $c \in [q]$, it holds that*

$$\ell_n(d,\beta) \leq \Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] \leq u_n(d,\beta).$$

Proof. We prove the lemma by induction on n. For the base case $n = 0$, note that $\mathbb{T}_{d,n}$ has a single vertex. Thus, for every $c \in [q]$ and every τ assigning a colour to this vertex,

$$\ell_0(d,\beta) = 0 \leq \Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] \leq 1 = u_0(d,\beta).$$

For the inductive step, suppose $n > 0$. For convenience, denote by T the tree $\mathbb{T}_{d,n}$, by v the root $v_{d,n}$ and by Λ the leaves $\Lambda_{\mathbb{T}_{d,n}}$. Let $v_1, \ldots, v_d$ be the children of v in T and let $\Lambda_i = \Lambda_{T[v_i]}$ denote the set of leaves of the subtree $T[v_i]$. Consider any configuration $\tau\colon \Lambda \to [q]$ and any colour c. By the induction hypothesis, for every $i \in [d]$ and $j \in [q]$, we have

$$\ell_{n-1}(d,\beta) \leq \Pr_{T[v_i]}[\sigma(v_i) = j \mid \sigma(\Lambda_i) = \tau(\Lambda_i)] \leq u_{n-1}(d,\beta).$$

By Lemma 3.3.4 and (3.25), we conclude that $\ell_n(d,\beta) \leq \Pr_T[\sigma(v) = c \mid \sigma(\Lambda) = \tau] \leq u_n(d,\beta)$. □

The following lemma will be used to show that the sequences $u_n(d,\beta)$ and $\ell_n(d,\beta)$ converge.

Lemma 3.3.11. *For any fixed $d \geq 2$, $0 < \beta < 1$ and $n \in \mathbb{N}$, we have*

1. $u_n(d,\beta) \geq u_{n+1}(d,\beta)$,
2. $\ell_n(d,\beta) \leq \ell_{n+1}(d,\beta)$, *and*
3. $2\ell_n(d,\beta) + u_n(d,\beta) \leq 1 \leq 2u_n(d,\beta) + \ell_n(d,\beta)$.

Proof. We prove this lemma by induction on n. Since d and β are fixed, we simplify the notation by writing u_n for $u_n(d,\beta)$, writing ℓ_n for $\ell_n(d,\beta)$, writing $f_{\mathsf{upper}}(x,y)$ for $f_{\mathsf{upper}}(d,\beta,x,y)$ and writing $f_{\mathsf{lower}}(x,y)$ for $f_{\mathsf{lower}}(d,\beta,x,y)$.
For the base case $n = 0$, we have $u_0 = 1$ and $\ell_0 = 0$, so Item 3 holds since $2\ell_0 + u_0 = 1 < 2u_0 + \ell_0$. Items 1 and 2 follow from

$$u_1 = f_{\mathsf{upper}}(1,0) = \frac{1}{1 + 2\beta^{d/2}} < 1 = u_0, \text{ and } \ell_1 = f_{\mathsf{lower}}(1,0) = \frac{\beta^d}{\beta^d + 2} > 0 = \ell_0.$$

For the inductive step, suppose $n > 0$. Item 3 follows (using the induction hypothesis) from Lemma 3.3.8 with $x = u_{n-1}$ and $y = \ell_{n-1}$. We now obtain Items 1 and 2. By the induction hypothesis,

$$0 \leq \ell_{n-1} \leq \ell_n \leq u_n \leq u_{n-1} \leq 1, \text{ and } 2\ell_{n-1} + u_{n-1} \leq 1 \leq 2u_{n-1} + \ell_{n-1}.$$

Using Lemma 3.3.9 (with these facts and with item 3), we obtain

$$u_{n+1} = f_{\mathsf{upper}}(u_n, \ell_n) \leq f_{\mathsf{upper}}(u_{n-1}, \ell_{n-1}) = u_n,$$

proving Item 1. Similarly, we also obtain that $\ell_{n+1} = f_{\mathsf{lower}}(u_n, \ell_n) \geq f_{\mathsf{lower}}(u_{n-1}, \ell_{n-1}) = \ell_n$, proving Item 2. □

By Lemma 3.3.11, we have that the sequences $\{u_n(d, \beta\})$ and $\{\ell_n(d, \beta)\}$ are bounded and monotonic, so they both converge. Let

$$u_\infty(d, \beta) := \lim_{n\to\infty} u_n(d, \beta), \text{ and } \ell_\infty(d, \beta) := \lim_{n\to\infty} \ell_n(d, \beta). \tag{3.26}$$

We have the following characterisation of the limits $u_\infty(d, \beta), \ell_\infty(d, \beta)$.

Lemma 3.3.12. *For any $d \geq 2$ and $0 < \beta \leq 1$, $(x, y) = (u_\infty(d, \beta), \ell_\infty(d, \beta))$ is a solution to the system of equations*

$$\begin{cases} f_{\mathsf{upper}}(d, \beta, x, y) = x \\ f_{\mathsf{lower}}(d, \beta, x, y) = y \end{cases}$$

satisfying $0 < y \leq 1 - x - y \leq x < 1$.

Proof. Since d and β are fixed, we simplify the notation by writing u_∞ for $u(d, \beta)$ and ℓ_∞ for $\ell(d, \beta)$. We also drop d and β as parameters of u_n, ℓ_n, f_{upper} and f_{lower} (as in the proof of Lemma 3.3.11). By Lemma 3.3.11, we have $\ell_\infty \geq \ell_1 = \beta^d/(\beta^d + 2) > 0$ and $u_\infty \leq u_1 = 1/(1 + 2\beta^{d/2}) < 1$. Also, for every non-negative integer n, we have $\ell_n \leq 1 - u_n - \ell_n \leq u_n$, which implies, by applying limits, that $\ell_\infty \leq 1 - u_\infty - \ell_\infty \leq u_\infty$. Recall that, for $n \geq 0$, $u_{n+1} = f_{\mathsf{upper}}(u_n, \ell_n)$ and $\ell_{n+1} = f_{\mathsf{lower}}(u_n, \ell_n)$. Using these definitions and the continuity of the functions $f_{\mathsf{upper}}(x, y)$ and $f_{\mathsf{lower}}(x, y)$ with respect to x and y (in the third equality below), we have

$$u_\infty = \lim_{n\to\infty} u_n = \lim_{n\to\infty} f_{\mathsf{upper}}(u_{n-1}, \ell_{n-1}) = f_{\mathsf{upper}}(\lim_{n\to\infty} u_{n-1}, \lim_{n\to\infty} \ell_{n-1}) = f_{\mathsf{upper}}(u_\infty, \ell_\infty).$$

Similarly, $\ell_\infty = f_{\mathsf{lower}}(u_\infty, \ell_\infty)$. □

3.3.4 Bounding the maximum ratio

In this section, we place the final pieces for the proof of Lemma 3.3.1. The first lemma accounts for the $d = 2$ case of Lemma 3.3.1.

Lemma 3.3.13. *Suppose $q = 3$ and $\beta \in (0, 1)$. Then there is a positive integer n_0 such that for every $n \geq n_0$, every $c \in [q]$, and every configuration $\tau \colon \Lambda_{\mathbb{T}_{2,n}} \to [q]$,*

$$\frac{459}{2000} \leq \Pr_{\mathbb{T}_{2,n}}[\sigma(v_{2,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{2,n}}) = \tau] \leq \frac{1107}{2500}.$$

Proof. For any $0 < \beta \leq 1$, by Lemma 3.3.12, $(x, y) = (u_\infty(2, \beta), \ell_\infty(d, \beta))$ is a solution to the system of equations

$$\begin{cases} f_{\mathsf{upper}}(2, \beta, x, y) = x \\ f_{\mathsf{lower}}(2, \beta, x, y) = y \end{cases} \tag{3.27}$$

satisfying $0 < y \leq 1 - x - y \leq x < 1$. In Section 8.1.2, we use the `Resolve` function of Mathematica to show rigorously that there is no solution to (3.27) satisfying

$$0 < y \leq \frac{1}{3} \text{ and } \frac{1106}{2500} \leq x < 1$$

and there is no solution satisfying

$$0 < y \leq \frac{460}{2000} \text{ and } \frac{1}{3} \leq x < 1.$$

If $0 < y \leq 1 - x - y \leq x < 1$ then $0 < y \leq 1/3$ and $1/3 \leq x < 1$. So any solution to (3.27) which satisfies $0 < y \leq 1 - x - y \leq x < 1$ must also satisfy $y > 460/2000$ and $x < 1106/2500$. We conclude that $\ell_\infty(d, \beta) > 460/2000$ and $u_\infty(d, \beta) < 1106/2500$.
Since $\ell_\infty(2, \beta)$ and $u_\infty(2, \beta)$ are the limits of the sequences $\ell_n(2, \beta)$ and $u_n(2, \beta)$, respectively, there is a positive integer n_0 such that, for all $n \geq n_0$, $\ell_n(2, \beta) \geq 459/2000$ and $u_n(2, \beta) \leq 1107/2500$. Thus by Lemma 3.3.10, for every $n \geq n_0$ and every $\tau \colon \Lambda_{\mathbb{T}_{2,n}} \to [3]$,

$$\frac{459}{2000} \leq \ell_n(2, \beta) \leq \Pr_{\mathbb{T}_{2,n}}[\sigma(v_{2,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{2,n}}) = \tau] \leq u_n(2, \beta) \leq \frac{1107}{2500}. \qquad \square$$

Definition 3.3.14. For any $d \geq 3$, define the critical parameter $\beta_*(d)$ by $\beta_*(d) = 1 - \dfrac{3}{d+1}$.

Note that $\beta_*(d) > 0$. The following lemma shows that $u_n(d, \beta)$ and $\ell_n(d, \beta)$ are bounded by the values corresponding to the critical parameter.

Lemma 3.3.15. *Fix any $d \geq 3$. For any β in the range $\beta_*(d) \leq \beta < 1$ and any non-negative integer n, we have $u_n(d, \beta_*(d)) \geq u_n(d, \beta)$ and $\ell_n(d, \beta_*(d)) \leq \ell_n(d, \beta)$.*

Proof. We prove the lemma by induction on n. Since d is fixed, we simplify the notation by writing β_* for $\beta_*(d)$. We also drop the argument d from $u_n(d,\beta)$, $\ell_n(d,\beta)$, $f_{\mathsf{upper}}(d,\beta,x,y)$ and $f_{\mathsf{lower}}(d,\beta,x,y)$.
For the base case $n=0$, note that for every β, it holds that $u_0(\beta)=u_0(\beta_*)=1$ and $\ell_0(\beta)=\ell_0(\beta_*)=0$.
For the inductive step, suppose $n>0$. By Lemma 3.3.11, we have

$$2\ell_{n-1}(\beta)+u_{n-1}(\beta)\le 1\le 2u_{n-1}(\beta)+\ell_{n-1}(\beta), \text{ and}$$
$$2\ell_{n-1}(\beta_*)+u_{n-1}(\beta_*)\le 1\le 2u_{n-1}(\beta_*)+\ell_{n-1}(\beta_*).$$

By Lemma 3.3.11 and the induction hypothesis, we have

$$0\le \ell_{n-1}(\beta_*)\le \ell_{n-1}(\beta)\le u_{n-1}(\beta)\le u_{n-1}(\beta_*)\le 1.$$

Now, using the definitions and Lemma 3.3.5, we get

$$u_n(\beta_*)=f_{\mathsf{upper}}(\beta_*,u_{n-1}(\beta_*),\ell_{n-1}(\beta_*))\ge f_{\mathsf{upper}}(\beta,u_{n-1}(\beta_*),\ell_{n-1}(\beta_*)).$$

Then using Lemma 3.3.9, we continue with

$$f_{\mathsf{upper}}(\beta,u_{n-1}(\beta_*),\ell_{n-1}(\beta_*))\ge f_{\mathsf{upper}}(\beta,u_{n-1}(\beta),\ell_{n-1}(\beta))=u_n(\beta).$$

Similarly, again using Lemma 3.3.5 and then Lemma 3.3.9, we have that

$$f_{\mathsf{lower}}(\beta_*,u_{n-1}(\beta_*),\ell_{n-1}(\beta_*))\le f_{\mathsf{lower}}(\beta,u_{n-1}(\beta_*),\ell_{n-1}(\beta_*))\le f_{\mathsf{lower}}(\beta,u_{n-1}(\beta),\ell_{n-1}(\beta)),$$

which gives that $\ell_n(\beta_*)\le\ell_n(\beta)$. □

Our next goal is to prove Lemma 3.3.26 below, which will help us to obtain an upper bound on the ratio $u_\infty(d,\beta_*(d))/\ell_\infty(d,\beta_*(d))$ when d is sufficiently large. In order to do this, we first define some useful re-parametrisations of f_{upper} and f_{lower}, and establish some properties of these.

Definition 3.3.16. Let $g_u(d,\mu,y)=f_{\mathsf{upper}}(d,\beta_*(d),\mu\cdot y,y)-\mu\cdot y$ and $g_\ell(d,\mu,y)=f_{\mathsf{lower}}(d,\beta_*(d),\mu\cdot y,y)-y$.

Note that the argument μ in g_u,g_ℓ corresponds to the ratio x/y of the arguments x,y of $f_{\mathsf{upper}},f_{\mathsf{lower}}$.

Lemma 3.3.17. *For every $d \geq 5$ and $\mu \geq 1$, $g_u(d, \mu, y)$ is a decreasing function of y in the range $1/(2\mu+1) \leq y \leq 1/(\mu+2)$.*

Proof. Let $A = (1 - 3y/(d+1))^d$ and

$$B = \left(1 - \frac{3(1-y(\mu+1))}{d+1}\right)^{d/2} \left(1 - \frac{3\mu y}{d+1}\right)^{d/2}.$$

Since $y \leq 1/(\mu+2) \leq 1/3$ and $d \geq 5$, we have $3y < d+1$, so $A > 0$. Also, $3\mu y < d+1$ and $3(1-y(\mu+1)) < d+1$, so $B > 0$. Let $W = AB/(A+2B)^2 > 0$. The derivative of g_u with respect to y (see Section 8.1.3 for Mathematica assistance) is given by the following.

$$\frac{\partial g_u}{\partial y} = -\mu + \left(\frac{9dW}{3y(\mu+1)+d-2}\right)\left(\frac{\mu(2\mu y+y-1)}{1+d-3\mu y} - \frac{2\mu y+y+d-1}{1+d-3y}\right).$$

The upper bound on y yields (crudely) $2\mu y + y - 1 < 1$ and $1 + d - 3\mu y > d - 2$. Similarly, the lower bound on y yields $2\mu y + y + d - 1 \geq d$ and (since y is non-negative) $1 + d - 3y < 1 + d$. Plugging these in, we obtain

$$\frac{\partial g_u}{\partial y} < -\mu + \left(\frac{9dW}{3y(\mu+1)+d-2}\right)\left(\frac{\mu}{d-2} - \frac{d}{1+d}\right). \tag{3.28}$$

Note that $3y(\mu+1) > 0$ and $d > 2$ so the expression $3y(\mu+1)+d-2$ in the denominator is positive. We now consider two cases.

If $\mu/(d-2) - d/(1+d) \leq 0$ then recall that $W > 0$, so (3.28) gives that $\frac{\partial g_u}{\partial y} < -\mu < 0$. Otherwise, $\mu/(d-2) - d/(1+d) > 0$. In this case, note that, for any A and B, $(A-2B)^2 \geq 0$, so $8AB \leq A^2 + 4AB + 4B^2 = (A+2B)^2$. From the definition of W, this ensures that $W \leq 1/8$. So, (3.28) gives that

$$\frac{\partial g_u}{\partial y} < -\mu + \frac{\frac{9d}{8}\left(\frac{\mu}{d-2} - \frac{d}{1+d}\right)}{3y(\mu+1)+d-2} < -\mu + \frac{9d}{8(d-2)}\left(\frac{\mu}{d-2} - \frac{d}{1+d}\right) < \mu\left(\frac{9d}{8(d-2)^2} - 1\right) < 0,$$

where the final inequality uses $d \geq 5$. □

The following lemma is analogous to Lemma 3.3.17, but for the function g_ℓ.

Lemma 3.3.18. *For every $d \geq 3$ and $\mu \geq 1$, $g_\ell(d, \mu, y)$ is a decreasing function of y in the range $1/(2\mu+1) \leq y \leq 1/(\mu+2)$.*

Proof. Let

$$W = \frac{(\mu(2d-1)+d+1)\left(\frac{3y(\mu+1)+d-2}{d+1}\right)^d}{(1+d-3\mu y)(d-2+3y(1+\mu))} + \frac{(\mu-1)(d+1)\left(1-\frac{3y}{d+1}\right)^d}{(1+d-3y)(1+d-3\mu y)}.$$

Since $\mu \geq 1$ and $d \geq 3$ and $y \leq 1/(\mu+2)$ all of the factors in W are positive, so $W > 0$. The derivative of g_ℓ with respect to y (see Section 8.1.4 for Mathematica assistance) is given by the following.

$$\frac{\partial g_\ell}{\partial y} = -\frac{3d\left(1-\frac{3\mu y}{d+1}\right)^d W}{\left(\left(\frac{3(\mu+1)y+d-2}{d+1}\right)^d + \left(1-\frac{3\mu y}{d+1}\right)^d + \left(1-\frac{3y}{d+1}\right)^d\right)^2} - 1.$$

We've already seen that $W > 0$ and the denominator is greater than 0 since it is a square. Since $3\mu y < 3 < d+1$, the remaining term is also positive, so $\frac{\partial g_\ell}{\partial y} < 0$, as required. □

Next, we will identify a value y_μ so that, when μ and d are sufficiently large, $g_u(d,\mu,y_\mu) < 0$ and $g_\ell(d,\mu,y_\mu) > 0$.

Definition 3.3.19. Define the quantity y_μ as follows.

$$y_\mu = \begin{cases} \frac{7}{10\mu+12} + \frac{3}{500} & \text{if } \mu < 32, \\ \frac{7}{10\mu+12} & \text{if } \mu \geq 32. \end{cases}$$

Let $x_\mu = \mu y_\mu$. Now define the functions h_u and h_ℓ as $h_u(d,\mu) = g_u(d,\mu,y_\mu)$ and $h_\ell(d,\mu) = g_\ell(d,\mu,y_\mu)$.

Then we have the following lemmas.

Lemma 3.3.20. *If $\mu \geq 157/80$ then $0 < y_\mu < 1 - x_\mu - y_\mu < \frac{1}{3} < x_\mu < 1 - y_\mu$.*

Proof. The inequalities follow directly from Definition 3.3.19. Mathematica code is given in Chapter 8. □

Lemma 3.3.21. *Suppose $d \geq 23$. If $157/80 \leq \mu < 32$ or $32 < \mu$ then $\frac{\partial h_u}{\partial \mu} < 0$.*

Proof. Since d is fixed in the proof of this lemma, we will drop it as an argument of β_*, h_u, g_u. We will use $\hat{\beta}_*$ to denote $1 - \beta_* = 3/(d+1)$. We will drop d and β_* as an argument of f_{upper}. So, plugging in Definitions 3.3.16 and 3.3.19, we get $h_u(\mu) = g_u(\mu, y_\mu) = f_{\text{upper}}(x_\mu, y_\mu) - x_\mu$. We have

$$\frac{\partial h_u(\mu)}{\partial \mu} = \frac{\partial f_{\text{upper}}(x_\mu, y_\mu)}{\partial x_\mu} \cdot \frac{\partial x_\mu}{\partial \mu} + \frac{\partial f_{\text{upper}}(x_\mu, y_\mu)}{\partial y_\mu} \cdot \frac{\partial y_\mu}{\partial \mu} - \frac{\partial x_\mu}{\partial \mu}. \tag{3.29}$$

Let

$$R(x,y) = \frac{(1-\hat{\beta}_* x)^{d/2}(1-\hat{\beta}_*(1-x-y))^{d/2}}{(1-\hat{\beta}_* y)^d}.$$

The derivatives of $f_{\mathsf{upper}}(x,y)$ with respect to x and y are as follows (see Section 8.1.6 for Mathematica assistance).

$$\begin{aligned}\frac{\partial f_{\mathsf{upper}}(x,y)}{\partial x} &= \left(\frac{f_{\mathsf{upper}}(x,y)^2 R(x,y) d\hat{\beta}_*}{1-\hat{\beta}_*(1-x-y)}\right)\left(\frac{\hat{\beta}_*(2x+y-1)}{1-\hat{\beta}_* x}\right), \text{ and} \\ \frac{\partial f_{\mathsf{upper}}(x,y)}{\partial y} &= -\left(\frac{f_{\mathsf{upper}}(x,y)^2 R(x,y) d\hat{\beta}_*}{1-\hat{\beta}_*(1-x-y)}\right)\left(\frac{3+\hat{\beta}_*(2x+y-2)}{1-\hat{\beta}_* y}\right).\end{aligned} \tag{3.30}$$

If $0 \le x \le 1$, $0 \le y \le 1$, $0 \le 1-x-y \le 1$ and $2x+y > 1$ then all of the factors are positive, so by Lemma 3.3.20, $\frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial x_\mu} > 0$ and $\frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial y_\mu} < 0$. Let

$$z = -\frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial x_\mu} \Big/ \frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial y_\mu} \tag{3.31}$$

and note that z is positive. Using (3.29) and (3.31), we can express $\frac{\partial h_u(\mu)}{\partial \mu}$ as

$$\frac{\partial h_u(\mu)}{\partial \mu} = \left(-z \cdot \frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial y_\mu} - 1\right)\frac{\partial x_\mu}{\partial \mu} + \frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial y_\mu} \cdot \frac{\partial y_\mu}{\partial \mu}.$$

From the definition of y_μ (Definition 3.3.19), $\frac{\partial y_\mu}{\partial \mu} = -\frac{35}{2(5\mu+6)^2} < 0$ for all $\mu \neq 32$. If $\mu < 32$ then $\frac{\partial x_\mu}{\partial \mu} = \frac{21}{(5\mu+6)^2} + \frac{3}{500} > 0$. If $\mu > 32$ then $\frac{\partial x_\mu}{\partial \mu} = \frac{21}{(5\mu+6)^2} > 0$. Thus, to show $\frac{\partial h_u(\mu)}{\partial \mu} < 0$, it suffices to show

$$-\frac{\partial f_{\mathsf{upper}}(x_\mu,y_\mu)}{\partial y_\mu} < \frac{\frac{\partial x_\mu}{\partial \mu}}{z \cdot \frac{\partial x_\mu}{\partial \mu} - \frac{\partial y_\mu}{\partial \mu}}. \tag{3.32}$$

We will simplify (3.32) by finding an upper bound for z. Using (3.30) and (3.31), we have

$$z = \left(\frac{1-\hat{\beta}_* y_\mu}{1-\hat{\beta}_* x_\mu}\right)\left(\frac{\hat{\beta}_*(2x_\mu+y_\mu-1)}{3+\hat{\beta}_*(2x_\mu+y_\mu-2)}\right) = \left(1+\frac{\hat{\beta}_*(x_\mu-y_\mu)}{1-\hat{\beta}_* x_\mu}\right)\left(\frac{\hat{\beta}_*(2x_\mu+y_\mu-1)}{3-\hat{\beta}_*(2-2x_\mu-y_\mu)}\right).$$

Since, by Lemma 3.3.20, $x_\mu > y_\mu$, $2x_u + y_y > 1$, $x_\mu > 0$ and (since $x_\mu + y_\mu < 1$) $2 > 2x_u + y_u$, z is an increasing function of $\hat{\beta}_*$. Since $d \ge 23$, we have $\hat{\beta}_* = 3/(d+1) \le 1/8$, so z is upper-bounded by its value with $\hat{\beta}_*$ replaced by $1/8$. This gives that

$$z \le \frac{(8-y_\mu)(2x_\mu+y_\mu-1)}{(8-x_\mu)(2x_\mu+y_\mu+22)}.$$

Moreover, using Mathematica, we show in Chapter 8.1.6 that

$$\frac{(8-y_\mu)(2x_\mu+y_\mu-1)}{(8-x_\mu)(2x_\mu+y_\mu+22)} < \frac{1}{24} \text{ for all } \mu > 1. \tag{3.33}$$

It follows that $z < 1/24$. Thus, we can re-write our goal from (3.32) — to prove the lemma, it suffices to show

$$-\frac{\partial f_{\mathsf{upper}}(x_\mu, y_\mu)}{\partial y_\mu} < \frac{\frac{\partial x_\mu}{\partial \mu}}{\frac{1}{24} \cdot \frac{\partial x_\mu}{\partial \mu} - \frac{\partial y_\mu}{\partial \mu}}. \tag{3.34}$$

The definitions of f_{upper} and R imply that $f_{\mathsf{upper}}(x, y) = 1/(1 + 2R(x, y))$. Therefore, using the fact that $\frac{a}{(1+2a)^2} \leq 1/8$ for all $a > 0$, we have

$$f_{\mathsf{upper}}(x, y)^2 R(x, y) = \frac{R(x, y)}{\big(1 + 2R(x, y)\big)^2} \leq \frac{1}{8}.$$

So, plugging this into the second equality in (3.30), recalling that $\frac{\partial f_{\mathsf{upper}}(x_\mu, y_\mu)}{\partial y_\mu} < 0$ and $\beta_* = 3/(d + 1)$, we get

$$-\frac{\partial f_{\mathsf{upper}}(x_\mu, y_\mu)}{\partial y_\mu} \leq \frac{d\hat{\beta}_*(3 + \hat{\beta}_*(2x_\mu + y_\mu - 2))}{8(1 - \hat{\beta}_*(1 - x_\mu - y_\mu))(1 - \hat{\beta}_* y_\mu)} \leq \frac{3(3 + \hat{\beta}_*(2x_\mu + y_\mu - 2))}{8(1 - \hat{\beta}_*(1 - x_\mu - y_\mu))(1 - \hat{\beta}_* y_\mu)}.$$

Let Y be the right-hand-side of the previous expression. Using $\hat{\beta}_* \leq 1/8$ and the inequalities from Lemma 3.3.20, we find that Y is increasing in $\hat{\beta}_*$ (see Mathematica code in Appendix 8.1.6). Thus, we can replace Y with its value with $\hat{\beta}_*$ replaced by $1/8$, which is $3(2x_\mu + y_\mu + 22)/((8 - y_\mu)(x_\mu + y_\mu + 7))$. Plugging this into (3.34), it suffices to show

$$\frac{3(2x_\mu + y_\mu + 22)}{(8 - y_\mu)(x_\mu + y_\mu + 7)} < \frac{\frac{\partial x_\mu}{\partial \mu}}{\frac{1}{24} \cdot \frac{\partial x_\mu}{\partial \mu} - \frac{\partial y_\mu}{\partial \mu}}. \tag{3.35}$$

We prove (3.35) in two cases.

Case 1: $\mu > 32$**:** Using the values $\frac{\partial x_\mu}{\partial \mu}, \frac{\partial y_\mu}{\partial \mu}$ that we calculated earlier, the right-hand side of (3.35) is $8/7$. The Mathematica code in Chapter 8.1.6 uses `Resolve` to show rigorously that there is no $\mu > 32$ satisfying (3.35).

Case 2: $\mu < 32$**:** Using the values that we calculated earlier, the right-hand side of (3.35) is

$$\frac{24\left(25\mu^2 + 60\mu + 3536\right)}{25\mu^2 + 60\mu + 73536}.$$

The Mathematica code in Appendix 8.1.6 uses `Resolve` to show rigorously that there is no $\mu > 1$ satisfying (3.35). □

We will use the following function in several of the remaining lemmas.

Definition 3.3.22. Let $\psi(d,z) = d/(d-3z+1) + \ln(d+1-3z)$.

Lemma 3.3.23. *Suppose $d \geq 23$. Then $h_u(d, 157/80) < 0$ and $h_u(d, 32) < 0$.*

Proof. Let $\zeta(d,x,y) := 2\psi(d,y) - \psi(d,x) - \psi(d,1-x-y)$, where ψ is the function defined in Definition 3.3.22. The derivative of $h_u(d,\mu)$ with respect to d is given as follows (see Chapter 8.1.7 for the Mathematica code).

$$\begin{aligned}\frac{\partial h_u(d,\mu)}{\partial d} &= \frac{\partial f_{\mathsf{upper}}(d,\beta_*(d),x_\mu,y_\mu)}{\partial d} \\ &= \frac{\left(1-\frac{3x_\mu}{d+1}\right)^{\frac{d}{2}}\left(1-\frac{3y_\mu}{d+1}\right)^{d}\left(1-\frac{3(1-x_\mu-y_\mu)}{d+1}\right)^{\frac{d}{2}}\zeta(d,x_\mu,y_\mu)}{\left(\left(1-\frac{3y_\mu}{d+1}\right)^{d} + 2\left(1-\frac{3x_\mu}{d+1}\right)^{\frac{d}{2}}\left(1-\frac{3(1-x_\mu-y_\mu)}{d+1}\right)^{\frac{d}{2}}\right)^2}. \end{aligned} \tag{3.36}$$

First, fix $\mu = 157/80$. We will prove three facts.

- **Fact 1:** For all $d \geq 23$, $\frac{\partial\zeta(d,x_\mu,y_\mu)}{\partial d} < 0$.
- **Fact 2:** $\lim_{d\to\infty}\zeta(d,x_\mu,y_\mu) = 0$.
- **Fact 3:** $\lim_{d\to\infty} h_u(d,\mu) < 0$.

Facts 1 and 2 guarantee that, for all $d \geq 23$, $\zeta(d,x_\mu,y_\mu) > 0$. Lemma 3.3.20 guarantees that all other factors in (3.36) are also positive. Thus, $\frac{\partial h_u(d,\mu)}{\partial d}$ is positive for all $d \geq 23$. Together with Fact 3, this proves the first part of the lemma, that $h_u(d,157/80) < 0$. The three facts are proved in the Mathematica code in Chapter 8.1.7.
Finally, fix $\mu = 32$. Lemma 3.3.20 again guarantees that all factors in (3.36) other than $\zeta(d,x_\mu,y_\mu)$ are positive. Thus, it suffices to prove the three facts for $\mu = 32$, and this is done in the Mathematica code in Chapter 8.1.7. □

Lemmas 3.3.21 and 3.3.23 have the following corollary.

Corollary 3.3.24. *For every $d \geq 23$ and $\mu \geq 157/80$, $h_u(d,\mu) < 0$.*

Proof. By Lemma 3.3.21, $h_u(d,\mu)$ is decreasing for $\mu \in [157/80, 32)$. Thus, for μ in this range, $h_u(d,\mu) \leq h_u(d,157/80)$ and by Lemma 3.3.23, $h_u(d,157/80) < 0$.
By Lemma 3.3.21, $h_u(d,\mu)$ is decreasing for $\mu > 32$. Thus, for $\mu > 32$, $h_(d,\mu) \leq h(d,32)$ and by Lemma 3.3.23, $h_u(d,32) < 0$. □

Lemma 3.3.25. *For every $d \geq 23$ and $\mu \geq 157/80$, $h_\ell(d,\mu) > 0$.*

Proof. We will show that

$$\frac{\partial h_\ell(d,\mu)}{\partial d} > 0 \text{ for all } d \geq 23 \text{ and } \mu \geq 157/80. \tag{3.37}$$

The Mathematica code in Chapter 8.1.8 verifies that $h_\ell(23,\mu) > 0$ for all $\mu \geq 157/80$. Together with (3.37), this proves the lemma.

Therefore, in the rest of the proof, we prove (3.37). Using Definitions 3.3.19 and 3.3.16, we have $h_\ell(d,\mu) = f_{\mathsf{lower}}(d,\beta_*(d),x_\mu,y_\mu) - y_\mu$. We use the following definitions in order to describe $\frac{\partial h_\ell(d,\mu)}{\partial d}$. Recall from Definition 3.3.22 that $\psi(d,z) = d/(d-3z+1) + \ln(d+1-3z)$. Let $A = (1-\frac{3x_\mu}{d+1})^d$ and $B = (1-\frac{3y_\mu}{d+1})^d$ and $C = (1-\frac{3(1-x_\mu-y_\mu)}{d+1})^d$. Then the derivative of $h_\ell(d,\mu)$ with respect to d is given as follows (see Chapter 8.1.8 for Mathematica assistance).

$$\frac{\partial h_\ell(d,\mu)}{\partial d} = \frac{\partial f_{\mathsf{lower}}(d,\beta_*(d),x_\mu,y_\mu)}{\partial d} = \frac{AC(\psi(d,x_\mu)-\psi(d,1-x_\mu-y_\mu)) + AB(\psi(d,x_\mu)-\psi(d,y_\mu))}{(A+B+C)^2}. \tag{3.38}$$

Lemma 3.3.20 guarantees that A, B and C are positive, so to prove (3.37), and hence the lemma, it suffices to show $\psi(d,x_\mu) > \psi(d,1-x_\mu-y_\mu)$ and $\psi(d,x_\mu) > \psi(d,y_\mu)$.

Note that

$$\frac{\partial\,\psi\left(d,\frac{1}{3}+t\right)}{\partial t} = \frac{9t}{(d-3t)^2}, \quad \text{and} \quad \frac{\partial\,\psi\left(d,\frac{1}{3}-t\right)}{\partial t} = \frac{9t}{(d+3t)^2}.$$

Thus, for fixed d, the function $\psi(d,z)$ is decreasing for $z \in [0,1/3]$. Since, by Lemma 3.3.20, $0 < y_\mu < 1-x_\mu-y_\mu < 1/3$, we have

$$\psi(d,y_\mu) \geq \psi(d,1-x_\mu-y_\mu). \tag{3.39}$$

The function $\psi(d,z)$ is increasing for $z \in [1/3,1]$. Since, Lemma 3.3.20 guarantees $1/3 < 2/3 - y_\mu < x_\mu < 1$, we have

$$\psi(d,x_\mu) \geq \psi(d,\tfrac{2}{3}-y_\mu). \tag{3.40}$$

Since the function $\psi(d,\frac{1}{3}+t)-\psi(d,\frac{1}{3}-t)$ is increasing for $t \in [0,1/3]$, and it is 0 at $t=0$, we have $\psi(d,\frac{1}{3}+t) \geq \psi(d,\frac{1}{3}-t)$ for $t \in [0,1/3]$. Lemma 3.3.20 guarantees $0 < y_\mu < 1/3$, so taking $t = 1/3 - y_\mu$, we get

$$\psi(d,\tfrac{2}{3}-y_\mu) \geq \psi(d,y_\mu). \tag{3.41}$$

Combining (3.40), (3.41) and (3.39) we obtain $\psi(d,x_\mu) > \psi(d,1-x_\mu-y_\mu)$ and $\psi(d,x_\mu) > \psi(d,y_\mu)$, which prove (3.37), and hence the lemma. □

Lemma 3.3.26. *If $d \geq 23$ then there is no solution to the system of equations*

$$\begin{cases} f_{\mathsf{upper}}(d, \beta_*(d), x, y) = x \\ f_{\mathsf{lower}}(d, \beta_*(d), x, y) = y \end{cases} \tag{3.42}$$

which satisfies $x \geq 157y/80 \geq 0$ and $2x + y \geq 1 \geq 2y + x$.

Proof. Consider any fixed $d \geq 23$ and, for the sake of contradiction, assume that such an (x, y) exists. Let $\mu = x/y$, so that (by Definition 3.3.16), (μ, y) is a solution to the equation

$$g_u(d, \mu, y) = g_\ell(d, \mu, y) = 0.$$

The conditions $x \geq 157y/80 \geq 0$ and $2x + y \geq 1 \geq 2y + x$ translate into $\mu \geq 157/80$ and $1/(2\mu + 1) \leq y \leq 1/(\mu + 2)$. Since $g_u(d, \mu, y) = 0$, by Lemma 3.3.17 and Corollary 3.3.24, we have $y < y_\mu$. Since $g_\ell(d, \mu, y) = 0$, by Lemmas 3.3.18 and 3.3.25, we have $y > y_\mu$. This yields a contradiction. □

Corollary 3.3.27. *For every integer $d \geq 23$, there exists a positive integer n_0 such that for all $n \geq n_0$,*

$$\frac{u_n(d, \beta_*(d))}{\ell_n(d, \beta_*(d))} \leq \frac{53}{27}.$$

Proof. Fix $d \geq 23$. For simplicity, we will write β_* instead of $\beta_*(d)$.
Recall that the sequences $\{u_n(d, \beta_*)\}$ and $\{\ell_n(d, \beta_*)\}$ converge to the limits $u_\infty(d, \beta_*)$ and $\ell_\infty(d, \beta_*)$, respectively (cf. (3.26)). Moreover, by Lemma 3.3.12, the pair $(x, y) = (u_\infty(d, \beta_*), \ell_\infty(d, \beta_*))$ is a solution to the system of equations (3.42) satisfying $0 < y \leq 1 - x - y \leq x < 1$. By Lemma 3.3.26, there is no solution (x, y) to (3.42) such that $x \geq 157y/80 \geq 0$ and $2x + y \geq 1 \geq 2y + x$. So, it must be the case that $u_\infty(d, \beta_*)/\ell_\infty(d, \beta_*) < 157/80$. Since $53/27 > 157/80$, there exists a positive integer n_0 such that for all $n \geq n_0$, $u_n(d, \beta_*)/\ell_n(d, \beta_*) \leq 53/27$. □

Corollary 3.3.27 accounts for integers $d \geq 23$. To account for integers $3 \leq d \leq 22$, we define the following two sequences.

$$\begin{cases} u'_0(d) = 1 \\ \ell'_0(d) = 0 \end{cases}$$

and for every non-negative integer n,

$$\begin{cases} u'_{n+1}(d) = \dfrac{\lceil 10000\, f_{\mathsf{upper}}(d, \beta_*(d), u'_n(d), \ell'_n(d)) \rceil}{10000} \\ \ell'_{n+1}(d) = \dfrac{\lfloor 10000\, f_{\mathsf{lower}}(d, \beta_*(d), u'_n(d), \ell'_n(d)) \rfloor}{10000} \end{cases}.$$

We have the following lemma, which is proved by brute force.

Lemma 3.3.28. *For every integer $d \in \{3, \dots, 22\}$ and every integer $n \in \{0, \dots, 60\}$, we have $u'_n(d) \geq u'_{n+1}(d)$, $\ell'_n(d) \leq \ell'_{n+1}(d)$, $2u'_n(d) + \ell'_n(d) \geq 1 \geq 2\ell'_n(d) + u_n(d)$ and $u'_{60}(d)/\ell'_{60}(d) \leq \frac{53}{27}$.*

Proof. In Appendix 8.1.9, we use Mathematica to compute all the values $u'_n(d)$ and $\ell'_n(d)$ for $n \in \{0, \dots, 60\}$ and $d \in \{3, \dots, 22\}$. We then check that all of the desired inequalities hold. □

We next show that the sequences $\{u'_n(d)\}$ and $\{\ell'_n(d)\}$ bound the sequences $\{u_n(d, \beta_*(d))\}$ and $\{\ell_n(d, \beta_*(d))\}$ for $n \leq 60$.

Lemma 3.3.29. *For every integer $d \in \{3, \dots, 22\}$ and every integer $n \in \{0, \dots, 60\}$, we have $u'_n(d) \geq u_n(d, \beta_*(d))$ and $\ell'_n(d) \leq \ell_n(d, \beta_*(d))$.*

Proof. Fix d to be an integer between 3 and 22. Since d is fixed, we simplify the notation by writing u_n for $u_n(d, \beta_*(d))$, u'_n for $u'_n(d)$, ℓ_n for $\ell_n(d, \beta_*(d))$, ℓ'_n for $\ell'_n(d)$, $f_{\mathsf{upper}}(x, y)$ for $f_{\mathsf{upper}}(d, \beta_*(d), x, y)$ and $f_{\mathsf{lower}}(x, y)$ for $f_{\mathsf{lower}}(d, \beta_*(d), x, y)$.
We prove the lemma by induction on n. For the base case $n = 0$, we have $u_n = u'_n = 1$ and $\ell_n = \ell'_n = 0$. For the inductive step, suppose $n > 0$. By Lemmas 3.3.11 and 3.3.28, we have

$$2\ell_{n-1} + u_{n-1} \leq 1 \leq 2u_{n-1} + \ell_{n-1}, \quad \text{and} \quad 2\ell'_{n-1} + u'_{n-1} \leq 1 \leq 2u'_{n-1} + \ell'_{n-1}.$$

By the induction hypothesis, we have

$$\ell'_{n-1} \leq \ell_{n-1} \leq u_{n-1} \leq u'_{n-1}.$$

Using Lemma 3.3.9, we therefore obtain that

$$\begin{aligned} u'_n &\geq f_{\mathsf{upper}}(u'_{n-1}, \ell'_{n-1}) \geq f_{\mathsf{upper}}(u_{n-1}, \ell_{n-1}) = u_n, \text{ and} \\ \ell'_n &\leq f_{\mathsf{lower}}(u'_{n-1}, \ell'_{n-1}) \leq f_{\mathsf{lower}}(u_{n-1}, \ell_{n-1}) = \ell_n. \end{aligned}$$

This completes the proof. □

Corollary 3.3.30. *For every integer $d \in \{3, \dots, 22\}$ and every integer $n \geq 60$, $\dfrac{u_n(d, \beta_*(d))}{\ell_n(d, \beta_*(d))} \leq \dfrac{53}{27}$.*

Proof. Fix an arbitrary integer d between 3 and 22. We have the following chain of inequalities (see below for explanation):

$$\frac{u_n(d,\beta_*(d))}{\ell_n(d,\beta_*(d))} \le \frac{u_{60}(d,\beta_*(d))}{\ell_{60}(d,\beta_*(d))} \le \frac{u'_{60}(d)}{\ell'_{60}(d)} \le \frac{53}{27}.$$

The first inequality holds by Lemma 3.3.11, since the sequence $\{u_n(d,\beta_*(d))\}$ is increasing and the sequence $\{\ell_n(d,\beta_*(d))\}$ is decreasing. The second inequality holds by Lemma 3.3.29. Finally, the third inequality holds by Lemma 3.3.28. □

We can now prove Lemma 3.3.1, which we restate here for convenience.

Lemma 3.3.2. *Let $q = 3$ and $c \in [3]$ be an arbitrary colour. For $d \ge 2$, consider the d-ary tree $\mathbb{T}_{d,n}$ with height n and let $\tau : \Lambda_{\mathbb{T}_{d,n}} \to [3]$ be an arbitrary configuration on the leaves. When $d = 2$, for all $\beta \in (0,1)$, for all sufficiently large n it holds that*

$$\frac{459}{2000} \le \Pr_{\mathbb{T}_{2,n}}[\sigma(v_{2,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{2,n}}) = \tau] \le \frac{1107}{2500}.$$

When $d \ge 3$, for all $\beta \in [1 - \frac{3}{d+1}, 1)$, there exist sequences $\{L_n\}$ and $\{U_n\}$ (depending on d and β) such that for all sufficiently large n

$$L_n \le \Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] \le U_n \text{ and } U_n/L_n \le 53/27.$$

Proof. The statement for $d = 2$ follows directly from Lemma 3.3.13.
Suppose $d \ge 3$. Let $U_n = u_n(d,\beta_*(d))$, $L_n = u_n(d,\beta_*(d))$. By Corollaries 3.3.27 and 3.3.30, there exists an integer n_0 such that for all $n \ge n_0$,

$$\frac{U_n}{L_n} = \frac{u_n(d,\beta_*(d))}{\ell_n(d,\beta_*(d))} \le \frac{53}{27}.$$

Furthermore, by Lemmas 3.3.10 and 3.3.15, for any $n \ge 0$, any configuration $\tau\colon \Lambda_{\mathbb{T}_{d,n}} \to [3]$ and any colour $c \in [3]$, we have

$$L_n = \ell_n(d,\beta_*(d)) \le \Pr_{\mathbb{T}_{d,n}}[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau] \le u_n(d,\beta_*(d)) = U_n.$$

This completes the proof. □

3.3.5 Verifying Condition 3.2.4

In this section, we assume that $q = 3$ and give the proof of the following lemma that verifies Condition 3.2.4 for all $\alpha \in (1, 53/27]$.

Lemma 3.3.31. *Let $q = 3$ and $d \geq 2$. Then, the pair (q, d) satisfies Condition 3.2.4 for all $\alpha \in (1, 53/27]$.*

Recall that for a pair (q, d), Condition 3.2.4, for a fixed value of $\alpha > 1$ and colours $c_1, c_2 \in [q]$, amounts to checking

$$h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) < \alpha^{1/d} \text{ for all } \mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha), \tag{3.43}$$

where the set $\mathrm{Ex}_{c_2}(\alpha)$ is given by (cf. (3.11))

$$\mathrm{Ex}_c(\alpha) = \big\{(p_1, \ldots, p_q) \in \{1, \alpha\}^q \mid p_c = 1 \wedge \exists c' \in [q] \text{ such that } p_{c'} = \alpha\big\}.$$

Note that, for $q = 3$, $\mathrm{Ex}_c(\alpha)$ has exactly 3 vectors. As we shall see shortly, we can capture the value of $h_{c_1,c_2,\beta}$ when $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_c(\alpha)$ using a function $\varphi(d, d_0, d_1, \alpha, \beta)$ which depends on α, β and d, but also on two non-negative integers d_0 and d_1 with $d_0 + d_1 \leq d$; roughly, d_0 is the number of $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}$ which are equal to the first vector in $\mathrm{Ex}_c(\alpha)$, d_1 to the second vector and the remaining $d - d_0 - d_1$ to the third vector. Let us first define the function $\varphi(d, d_0, d_1, \alpha, \beta)$.

Definition 3.3.32. Let $\beta \in [0, 1]$. Fix any $\alpha > 1$ and any integer $d \geq 2$. Let $x(\alpha, \beta) = 1 - \frac{(1-\beta)(\alpha-1)}{\beta+2\alpha}$ and $y(\alpha, \beta) = 1 - \frac{(1-\beta)(\alpha-1)}{\beta+\alpha+1}$. For any nonnegative integers d_0 and d_1 such that $d_0 + d_1 \leq d$, let

$$\varphi(d, d_0, d_1, \alpha, \beta) = 1 + \frac{(1-\beta)(1 - x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d_1})}{\beta + x(\alpha,\beta)^{d_0}(y(\alpha,\beta)^{d_1} + y(\alpha,\beta)^{d-d_0-d_1})}.$$

We then have the following lemma.

Lemma 3.3.33. *Suppose $q = 3$, $d \geq 2$ and $\beta \in [0, 1]$. Fix $\alpha > 1$ and distinct colours $c_1, c_2 \in [q]$, and let $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha)$. Then there are nonnegative integers d_0 and d_1 with $d_0 + d_1 \leq d$ such that*

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) = \varphi(d, d_0, d_1, \alpha, \beta).$$

Proof. Let c_3 be the remaining colour in $[q]$ (other than c_1 and c_2), so that $[q] = \{c_1, c_2, c_3\}$. Since $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha)$, for every $k \in [d]$ we have that $p^{(k)}_{c_2} = 1$ and one of three situations occurs:

$$\text{(i) } p^{(k)}_{c_1} = p^{(k)}_{c_3} = \alpha, \text{ (ii) } p^{(k)}_{c_1} = \alpha, p^{(k)}_{c_3} = 1, \text{ (iii) } p^{(k)}_{c_1} = 1, p^{(k)}_{c_3} = \alpha.$$

If situation (i) occurs, then, using the notation from Definition 3.3.32, we have

$$1 - \frac{(1-\beta)(p_{c_1}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)}} = x(\alpha, \beta). \tag{3.44}$$

If situation (ii) occurs, then this quantity is $y(\alpha, \beta)$. If situation (iii) occurs, then it is 1. Now let d_0 be the number of $k \in [d]$ for which situation (i) occurs and let d_1 be the number of $k \in [d]$ for which situation (ii) occurs. Then plugging (3.44) and the other similar observations above into the definition (3.5) of $g_{c_1,c_2,\beta}$ and $g_{c_3,c_2,\beta}$, we have

$$g_{c_1,c_2,\beta}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)}) = x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d_1}, \text{ and}$$
$$g_{c_3,c_2,\beta}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)}) = x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d-d_0-d_1}.$$

Plugging this into the definition of $h_{c_1,c_2,\beta}$, we have

$$h_{c_1,c_2,\beta}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)}) = 1 + \frac{(1-\beta)\Big(1 - g_{c_1,c_2,\beta}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)})\Big)}{\beta + \sum_{c \neq c_2} g_{c,c_2,\beta}(\hat{\mathbf{p}}^{(1)}, \dots, \hat{\mathbf{p}}^{(d)})} = \varphi(d, d_0, d_1, \alpha, \beta). \ \square$$

The following definition applies Definition 3.3.32 to the critical value of β for the special case where $d_0 + d_1 = d$; we will see that this special case is all that we need to consider to verify Condition 3.2.4 for $\alpha \in (1, 2)$.

Definition 3.3.34. Let $q = 3$. Fix any $\alpha > 1$ and any integer $d \geq 2$. Let $\beta_*(d) = 1 - q/(d+1)$. Let $X = x(\alpha, \beta_*(d))$ and $Y = y(\alpha, \beta_*(d))$. Let

$$\varphi_*(d, d_0, \alpha) = \varphi(d, d_0, d - d_0, \alpha, \beta_*(d)) = 1 + \frac{3}{d+1} \cdot \frac{1 - X^{d_0} Y^{d-d_0}}{\beta_*(d) + X^{d_0}(Y^{d-d_0} + 1)}.$$

The values X and Y from definition 3.3.34 are clearly functions of d and α, but we suppress this in the notation to avoid clutter. (These values will arise in proofs, but, unlike $\varphi_*(d, d_0, \alpha)$, they will not arise in statements of lemmas.)
The following lemma applies for $\alpha \in (1, 2)$.

Lemma 3.3.35. *Suppose $q = 3$, $d \geq 2$ and $1 - q/(d+1) \leq \beta \leq 1$. Fix $\alpha \in (1, 2)$ and distinct colours $c_1, c_2 \in [q]$. Suppose that $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \dots, \mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha)$. Then, there is a non-negative integer $d_0 \leq d$ such that*

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \leq \varphi_*(d, d_0, \alpha).$$

Proof. Since d is fixed, we simplify the notation by writing β_* for $\beta_*(d)$.
Note that both $x(\alpha,\beta)$ and $y(\alpha,\beta)$ are increasing functions of β, so φ is a decreasing function of β. Thus for all $\beta \in [\beta_*, 1]$, we have $\varphi(d, d_0, d_1, \alpha, \beta_*) \geq \varphi(d, d_0, d_1, \alpha, \beta)$. Combining this with Lemma 3.3.33, we find that there are nonnegative integers d_0 and d_1 with $d_0 + d_1 \leq d$ such that

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) = \varphi(d, d_0, d_1, \alpha, \beta) \leq \varphi(d, d_0, d_1, \alpha, \beta_*). \tag{3.45}$$

Let $d_2 = d - d_0 - d_1 \geq 0$. If $d_1 < d_2$, then

$$\varphi(d, d_0, d_2, \alpha, \beta_*) = 1 + \frac{(1-\beta_*)(1 - x(\alpha,\beta_*)^{d_0} y(\alpha,\beta_*)^{d_2})}{\beta_* + x(\alpha,\beta_*)^{d_0}(y(\alpha,\beta_*)^{d_1} + y(\alpha,\beta_*)^{d_2})} \geq \varphi(d, d_0, d_1, \alpha, \beta_*).$$

So we can further assume that $d_1 \geq d_2$. Since $1 < \alpha < 2$, and $\beta_* \leq 1$, we have

$$x(\alpha,\beta_*)^2 - y(\alpha,\beta_*) = \frac{(1-\beta_*)(\alpha-1)}{(\beta_*+2\alpha)(\beta_*+\alpha+1)} \left(\frac{\beta_*+\alpha+1}{\beta_*+2\alpha}(1-\beta_*)(\alpha-1) - (\beta_*+2) \right) \leq 0,$$

which implies

$$x(\alpha,\beta_*)^{d_0} y(\alpha,\beta_*)^{d_1} \geq x(\alpha,\beta_*)^{d_0+2} y(\alpha,\beta_*)^{d_1-1},$$

and

$$x(\alpha,\beta_*)^{d_0} y(\alpha,\beta_*)^{d_2} \geq x(\alpha,\beta_*)^{d_0+2} y(\alpha,\beta_*)^{d_2-1}.$$

These (together with the definition of φ) imply that if $d_1 \geq d_2 \geq 1$ then

$$\varphi(d, d_0, d_1, \alpha, \beta_*) \leq \varphi(d, d_0+2, d_1-1, \alpha, \beta_*).$$

Repeating this until $d_2 = 0$, we obtain

$$\varphi(d, d_0, d_1, \alpha, \beta_*) \leq \varphi(d, d_0+2d_2, d_1-d_2, \alpha, \beta_*) = \varphi_*(d, d_0+2d_2, \alpha).$$

Combining this with (3.45), we obtain

$$h_{c_1,c_2,\beta}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(d)}) \leq \varphi_*(d, d_0+2d_2, \alpha). \qquad \square$$

Lemma 3.3.36. *For every fixed $d \in \{2, \dots, 22\}$, $d_0 \in \{0, \dots, d\}$ and $\alpha \in (1, 53/27]$, we have $\varphi_*(d, d_0, \alpha) < \alpha^{1/d}$.*

Proof. This is rigorously verified using the `Resolve` function of Mathematica in Section 8.1.10. $\square$

Lemma 3.3.37. *For every fixed integer $d \geq 23$, $d_0 \in \{0, \dots, d\}$ and $\alpha \in (1, 53/27]$, we have $\varphi_*(d, d_0, \alpha) < \alpha^{1/d}$.*

Proof. Fix d and d_0 such that $d \geq 23$ and $d_0 \in \{0, \dots, d\}$. Since d is fixed, we simplify the notation by writing β_* for $\beta_*(d)$. Note that $\beta_* \in (0, 1)$. Given d, let X and Y be the functions of α defined in Definition 3.3.34, and observe that these are positive for all $\alpha \geq 1$. Let

$$\tilde{\varphi}(d, d_0, \alpha) = 1 + \frac{1 - X^{d_0} Y^{d-d_0}}{dX^{\frac{2d_0}{2+\beta_*}} Y^{\frac{d-d_0}{2+\beta_*}}} - \alpha^{1/d} = 1 + \frac{1}{d}\Big(X^{-\frac{2d_0}{2+\beta_*}} Y^{-\frac{d-d_0}{2+\beta_*}} - X^{\frac{d_0\beta_*}{2+\beta_*}} Y^{\frac{(d-d_0)(1+\beta_*)}{2+\beta_*}}\Big) - \alpha^{1/d}.$$

By the inequality of weighted arithmetic and geometric means[2] we have

$$\beta_* + X^{d_0}(Y^{d_1} + 1) \geq (2 + \beta_*) X^{\frac{2d_0}{2+\beta_*}} Y^{\frac{d_1}{2+\beta_*}},$$

which yields $\tilde{\varphi}(d, d_0, \alpha) \geq \varphi_*(d, d_0, \alpha) - \alpha^{1/d}$. Thus, our goal is to prove $\tilde{\varphi}(d, d_0, \alpha) < 0$ when $\alpha \in (1, 53/27]$. When $\alpha = 1$, X and Y are 1 so $\tilde{\varphi}(d, d_0, 1) = 0$. To prove the lemma, it suffices to show that $\frac{\partial \tilde{\varphi}}{\partial \alpha} < 0$ for $\alpha \in (1, 53/27]$. The rest of the proof is devoted to this technical fact. It is broken up into four steps.

Step 1. Let $\xi_1 = 2d_0(\frac{2+\alpha\beta_*}{2+\beta_*}) + (d - d_0)\alpha$, $M = -X^{d_0} Y^{d-d_0} \Big(d_0\beta_* \frac{2+\alpha\beta_*}{2+\beta_*} + (d - d_0)(\beta_* + 1)\alpha\Big)$, $S = X^{\frac{2d_0}{2+\beta_*}} Y^{\frac{d-d_0}{2+\beta_*}} (\alpha + \beta_* + 1)(\alpha\beta_* + 2)/(1 - \beta_*)$, and $\xi_2 = M + \alpha^{1/d} S$. The goal of Step 1 is to show that $\frac{\partial \tilde{\varphi}}{\partial \alpha} < 0$ follows from $\xi_1 \leq \xi_2$.

Technical details of Step 1. To calculate the derivative of the function $\tilde{\varphi}$, let

$$g_X := (2\alpha + \beta_*)(\alpha\beta_* + \alpha + 1), \qquad g_Y := (\alpha + \beta_* + 1)(\alpha\beta_* + 2).$$

Using Mathematica, we verify in Section 8.1.11 the formula

$$\frac{\partial \tilde{\varphi}}{\partial \alpha} = \left(\frac{(1 - \beta) X^{-\frac{2d_0}{2+\beta_*}} Y^{-\frac{d-d_0}{2+\beta_*}}}{d\alpha g_Y}\right) \left(2d_0\alpha \frac{g_Y}{g_X} + (d - d_0)\alpha + X^{d_0} Y^{d-d_0} \Big(d_0\beta_*\alpha \frac{g_Y}{g_X} + (d - d_0)(1 + \beta_*)\alpha\Big) - \alpha^{1/d} S\right). \tag{3.46}$$

For all $\alpha > 1$, we have that

$$\alpha(2 + \beta_*)(\alpha + \beta_* + 1) - (2\alpha + \beta_*)(\alpha\beta_* + \alpha + 1) = -\beta_*(\alpha - 1)^2 < 0,$$

[2] The inequality says that for non-negative $x_1, x_2, x_3, w_1, w_2, w_3$ with $w = w_1 + w_2 + w_3 > 0$, $w_1 x_1 + w_2 x_2 + w_3 x_3 \geq w x_1^{w_1/w} x_2^{w_2/w} x_3^{w_3/w}$. Take $x_1 = 1$, $x_2 = X^{d_0} Y^{d_1}$, $x_3 = X^{d_0}$, $w_1 = \beta_*$, $w_2 = 1$ and $w_3 = 1$.

so we obtain the bound

$$\frac{g_Y}{g_X} = \frac{(\alpha+\beta_*+1)(\alpha\beta_*+2)}{(2\alpha+\beta_*)(\alpha\beta_*+\alpha+1)} < \frac{(\alpha+\beta_*+1)(\alpha\beta_*+2)}{\alpha(2+\beta_*)(\alpha+\beta_*+1)} = \frac{2+\alpha\beta_*}{\alpha(2+\beta_*)}. \tag{3.47}$$

Now note that the first parenthesised expression in (3.46) is positive since $X, Y > 0$ for all $\alpha > 1$. Thus, to show $\frac{\partial\tilde{\varphi}}{\partial\alpha} < 0$, it suffices to show that the second parenthesised expression in (3.46) is less than 0. To do this, we can apply the strict upper bound on g_X/g_Y from (3.47), and show that the resulting expression, which is $\xi_1 - \xi_2$, is at most 0. Thus, we have completed Step 1.

Step 2. Let $W = 1 + \frac{\alpha-1}{d} - \frac{(d-1)(\alpha-1)^2}{2d^2}$ and $\xi_3 = M + WS$. The goal of Step 2 is to show $W \leq \alpha^{1/d}$, which implies $\xi_3 \leq \xi_2$. Given Step 1, this means that $\frac{\partial\tilde{\varphi}}{\partial\alpha} < 0$ will follow from showing that $\xi_1 \leq \xi_3$.

Technical details of Step 2. Recall that $d \geq 23$. Let $m(\alpha, d) = \alpha^{1/d} - W$. Note that

$$m(1,d) = \frac{\partial m}{\partial\alpha}\Big|_{\alpha=1} = \frac{\partial^2 m}{\partial\alpha^2}\Big|_{\alpha=1} = 0, \quad \text{and} \quad \frac{\partial^3 m}{\partial\alpha^3} = \frac{\left(2-\frac{1}{d}\right)\left(1-\frac{1}{d}\right)}{d\alpha^{3-\frac{1}{d}}} > 0 \text{ for all } \alpha > 1.$$

Thus, $m(\alpha, d) \geq 0$ for $\alpha > 1$.

Step 3. Let $\kappa = d_0/d$. After re-parameterising ξ_1 and ξ_3 (so that they depend on d, κ and α), we show that $\xi_1 \leq \xi_3$ follows from $\frac{\partial^2\xi_3}{\partial\alpha^2} > 0$. Given the other steps, this means that $\frac{\partial\tilde{\varphi}}{\partial\alpha} < 0$ follows from $\frac{\partial^2\xi_3}{\partial\alpha^2} > 0$. For future reference, the re-parametrisation is as follows. Let

$$\begin{aligned} s_0 &= -\left(\frac{(2d-1)d(1-\kappa)\alpha}{d+1} + \frac{(d-2)\kappa((d-2)\alpha+2(d+1))}{3(d+1)}\right), \text{ and} \\ t_0 &= \frac{((d+1)\alpha+2d-1)((d-2)\alpha+2(d+1))}{3(d+1)}\left(1+\frac{\alpha-1}{d}-\frac{(d-1)(\alpha-1)^2}{2d^2}\right). \end{aligned}$$

Then

$$\begin{aligned} \xi_1 &= \frac{\alpha-1}{3}(-d\kappa+3d-4\kappa) + d(\kappa+1), \text{ and} \\ \xi_3 &= s_0 X^{d\kappa} Y^{d(1-\kappa)} + t_0 X^{\frac{2\kappa(d+1)}{3}} Y^{\frac{(1-\kappa)(d+1)}{3}}. \end{aligned} \tag{3.48}$$

Technical details of Step 3.

The Mathematica code in Section 8.1.11 verifies that, at $\alpha = 1$, $\xi_3 = d(\kappa+1)$ and $\frac{\partial\xi_3}{\partial\alpha} = \frac{1}{3}(-d\kappa+3d-4\kappa)$. By Taylor's theorem and the Lagrange form of the remainder, there exists a number $\tilde{\alpha} \in (1, \alpha]$ such that

$$\xi_3 = d(\kappa+1) + \frac{1}{3}(-d\kappa+3d-4\kappa)(\alpha-1) + \frac{1}{2}\left(\frac{\partial^2\xi_3}{\partial\alpha^2}\Big|_{\alpha=\tilde{\alpha}}\right)(\alpha-1)^2. \tag{3.49}$$

Thus, if $\frac{\partial^2 \xi_3}{\partial \alpha^2} > 0$ for all $\alpha \in (1, 53/27]$, we can conclude that

$$\xi_3 \geq d(\kappa+1) + \frac{1}{3}(-d\kappa + 3d - 4\kappa)(\alpha - 1) = \xi_1. \tag{3.50}$$

Step 4. Using the parametrisation (3.48) of Step 3, we show that $\frac{\partial^2 \xi_3}{\partial \alpha^2} > 0$, for all $d \geq 23$, $\kappa \in [0,1]$ and $\alpha \in (1, 53/27]$, thus completing the proof.

Technical details of Step 4. We start with the observation that, for any k_1 and k_2 (not depending on α) and any function r of α, it holds that

$$\frac{\partial (r X^{k_1} Y^{k_2})}{\partial \alpha} = f(r, k_1, k_2) X^{k_1} Y^{k_2}, \text{ where } f(r, k_1, k_2) = \frac{\partial r}{\partial \alpha} + \frac{k_1 r}{X}\frac{\partial X}{\partial \alpha} + \frac{k_2 r}{Y}\frac{\partial Y}{\partial \alpha}.$$

Applying this observation twice to each of the two summands in the expression (3.48) for ξ_3, we see that there are rational functions s_2 and t_2 of α, d and κ such that

$$\frac{\partial^2 \xi_3}{\partial \alpha^2} = s_2 X^{d\kappa} Y^{d(1-\kappa)} + t_2 X^{\frac{2\kappa(d+1)}{3}} Y^{\frac{(1-\kappa)(d+1)}{3}}.$$

In Section 8.1.11 we use Mathematica to calculate t_2 explicitly and to verify that for every $d \geq 23$ and $\alpha \in (1, 53/27]$, we have $\frac{\partial^2 t_2}{\partial \kappa^2} \geq 0$, $\frac{\partial t_2}{\partial \kappa}\Big|_{\kappa=0} \geq 0$ and $t_2|_{\kappa=0} \geq 0$. We conclude that $t_2 \geq 0$ for all $\kappa \in [0,1]$ (for the given ranges of d and α). Since X and Y are less than or equal to 1 and $2\kappa(d+1)/3 \leq d\kappa$ and $(1-\kappa)(d+1)/3 \leq d(1-\kappa)$, the fact that $t_2 \geq 0$ implies

$$\frac{\partial^2 \xi_3}{\partial \alpha^2} = s_2 X^{d\kappa} Y^{d(1-\kappa)} + t_2 X^{\frac{2\kappa(d+1)}{3}} Y^{\frac{(1-\kappa)(d+1)}{3}} \geq (s_2 + t_2) X^{d\kappa} Y^{d(1-\kappa)}. \tag{3.51}$$

The final Mathematica code segment verifies that $s_2 + t_2 > 0$ for all $d \geq 23$, $\alpha \in (1, 53/27]$ and $\kappa \in [0,1]$. This, together with (3.51), completes the proof since X and Y are positive. □

We finish by giving the proof of Lemma 3.3.31.

Lemma 3.3.31. *Let $q = 3$ and $d \geq 2$. Then, the pair (q, d) satisfies Condition 3.2.4 for all $\alpha \in (1, 53/27]$.*

Proof. To prove Condition 3.2.4, it suffices to check (3.43) for $\alpha \in (1, 53/27)$. By Lemma 3.3.35, we only need to check that $\varphi_*(d, d_0, \alpha) < \alpha^{1/d}$ for $\alpha \in (1, 53/27)$ and integers $0 \leq d_0 \leq d$. This has been verified in Lemma 3.3.36 for all $d \leq 22$ and in Lemma 3.3.37 for all $d \geq 23$. □

3.4 Uniqueness threshold for $q = 4$

In this chapter, we will fix $q = 4$ and give the proof of Theorem 3.1.2. Note that when $q = 4$, it is difficult to characterise $h_{c_1,c_2,\beta}$ since there will be 7 cases instead of 3 cases in Lemma 3.3.33. So we cannot give a general result so far. However, we can still generalise our approach and introduce some simplified cases when $q = 4$. In Chapter 3.4.1, we will extend the method in Chapter 3.3.1, 3.3.2, 3.3.3 and 3.3.4 to arbitrary q, which can help us bound the ratio of marginal probabilities when $q = 4$ (Lemma 3.4.10 and Corollary 3.4.12). Then in Chapter 3.4.2 we will use Mathematica to verify Condition 3.2.4 in the bounded regime and complete the proof of Theorem 3.1.2.

3.4.1 A generalisation of f_{upper} and f_{lower}

We first define generalised f_{upper} and f_{lower}. Let $\hat{\beta} = 1 - \beta$ and $\hat{q} = q - 1$, then we define

$$f_{\mathsf{upper}}(q, d, \beta, x, y) = \frac{(1 - \hat{\beta}y)^d}{(1 - \hat{\beta}y)^d + \hat{q}\big((1 - \hat{\beta}x)^{\hat{q}-1}(1 - \hat{\beta}(1 - y - (\hat{q} - 1)x))\big)^{d/\hat{q}}}$$

and

$$f_{\mathsf{lower}}(q, d, \beta, x, y) = \frac{(1 - \hat{\beta}x)^d}{(1 - \hat{\beta}x)^d + (\hat{q} - 1)(1 - \hat{\beta}y)^d + (1 - \hat{\beta}(1 - x - (\hat{q} - 1)y))^d}.$$

We will use the following inequality.

Proposition 3.4.1 (Karamata's inequality)**.** *Let f be a convex function on an interval $[a, b]$. Suppose $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ are $2n$ real numbers in $[a, b]$ and $x_1 \geq x_2 \geq \cdots \geq x_n$ and $y_1 \geq y_2 \geq \cdots \geq y_n$. Then $f(x_1) + f(x_2) + \cdots + f(x_n) \geq f(y_1) + f(y_2) + \cdots + f(y_n)$ if $(x_1, x_2, \ldots, x_n)$ majorises $(y_1, y_2, \ldots, y_n)$.*
We say $(x_1, x_2, \ldots, x_n)$ majorises $(y_1, y_2, \ldots, y_n)$ if the following holds:

1. *for every $1 \leq m \leq n - 1$, $x_1 + x_2 + \cdots + x_m \geq y_1 + y_2 + \cdots + y_m$, and*
2. *$x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n$.*

Corollary 3.4.2. *Let f be a convex function on $(-\infty, 1]$, and $n+3$ numbers $a, b, c, x_1, \ldots, x_n$ satisfy that*

1. *$0 \leq a \leq b \leq 1$,*

2. *for every* $1 \leq i \leq n$, $a \leq x_i \leq b$,

3. $x_1 + x_2 + \cdots + x_n = c$ *where* $0 \leq c \leq 1$.

Then we have

$$f(x_1) + f(x_2) + \ldots f(x_n) \leq (n-1)f(a) + f(c-(n-1)a)\,,$$

and

$$f(x_1) + f(x_2) + \ldots f(x_n) \leq (n-1)f(b) + f(c-(n-1)b)\,.$$

Proof. Without loss of generality, we assume $x_1 \geq x_2 \geq \cdots \geq x_n$. Note that $c-(n-1)a \geq c - \sum_{i=2}^{n} x_i = x_1 \geq a$ and $c-(n-1)b \leq c - \sum_{i=1}^{n-1} x_i = x_n \leq b$. By Lemma 3.4.1, it suffices to show that both $(c-(n-1)a, a, \ldots, a)$ and $(b, \ldots, b, c-(n-1)b)$ majorise $(x_1, x_2, \ldots, x_n)$.

- For every $1 \leq m < n$, $c-(n-1)a + (m-1)a = c-(n-m)a \geq c - \sum_{i=m+1}^{n} x_i = x_1 + \cdots + x_m$. So $(c-(n-1)a, a, \ldots, a)$ majorises $(x_1, x_2, \ldots, x_n)$.
- For every $1 \leq m < n$, $mb \geq x_1 + \cdots + x_m$. So $(b, \ldots, b, c-(n-1)b)$ majorises $(x_1, x_2, \ldots, x_n)$.

Combining with the fact that $c-(n-1)b \leq c-(n-1)a \leq c \leq 1$ and $f(x)$ is a convex function on $(-\infty, 1]$, we complete the proof of this lemma. □

The following lemma gives recursively-generated bounds on the probability that the root of $\mathbb{T}_{d,n}$ is a given colour.

Lemma 3.4.3. *Suppose* $q \geq 3$, $d \geq 2$ *and* $\beta \in [0,1]$. *For any* $n \geq 0$, *let* $T = \mathbb{T}_{d,n}$ *with root* $v = v_T$ *and leaves* $\Lambda = \Lambda_T$. *Let* $v_1, \ldots, v_d$ *be the children of* v *in* T, T_i *be the subtree of* T *rooted at* v_i *and* Λ_i *denote the set of leaves of the subtree* T_i. *Consider any configuration* $\tau \colon \Lambda \to [q]$ *and any real numbers* $L, U \in [0,1]$ *such that, for all* $i \in [d]$ *and all* $j \in [q]$ *we have*

$$L \leq \Pr_{T_i}[\sigma(v_i) = j \mid \sigma(\Lambda_i) = \tau(\Lambda_i)] \leq U.$$

Then, for all colours $c \in [q]$, *we also have*

$$f_{\mathsf{lower}}(q,d,\beta,U,L) \leq \Pr_T[\sigma(v) = c \mid \sigma(\Lambda) = \tau] \leq f_{\mathsf{upper}}(q,d,\beta,U,L).$$

Proof. By symmetry between the colours, we may assume that $c = 1$. Let $p = \Pr_T[\sigma(v) = 1 \mid \sigma(\Lambda) = \tau]$. For any colour $c' \in [q]$ and any child $i \in [d]$, let $p_{i,c'} = \Pr_{T_i}[\sigma(v_i) = c' \mid \sigma(\Lambda_i) = \tau(\Lambda_i)]$.

Recall that for convenience let $\hat{\beta} := 1 - \beta \in [0, 1]$ and $\hat{q} = q - 1$. By Proposition 3.1.3, with $q \geq 3$ and $c = q$, we have that

$$p = \frac{1}{1+R}, \text{ where } R := \frac{\sum_{c'=1}^{\hat{q}} \prod_{i=1}^{d}(1 - \hat{\beta} p_{i,c'})}{\prod_{i=1}^{d}(1 - \hat{\beta} p_{i,q})}. \tag{3.52}$$

We first show that $p \geq f_{\mathsf{lower}}(q, d, \beta, U, L)$. Since $p_{i,q} \leq U$ for every $i \in [d]$, we obtain that

$$R \leq \frac{\sum_{c'=1}^{\hat{q}} \prod_{i=1}^{d}(1 - \hat{\beta} p_{i,c'})}{(1 - \hat{\beta} U)^d}. \tag{3.53}$$

For $c' \in [q]$, let $\bar{p}_{c'}$ denote the mean $\frac{1}{d}\sum_{i \in [d]} p_{i,c'}$. The function $f(x) = \ln(1 - \hat{\beta} x)$ is concave on the interval $(-\infty, 1]$ so by Jensen's inequality

$$\frac{1}{d}\sum_{i=1}^{d} \ln(1 - \hat{\beta} p_{i,c'}) \leq \ln(1 - \hat{\beta} \bar{p}_{c'}).$$

Thus, for $c' \in [\hat{q}]$ we have $\prod_{i=1}^{d}(1 - \hat{\beta} p_{i,c'}) \leq (1 - \hat{\beta} \bar{p}_{c'})^d$, which implies

$$\sum_{c'=1}^{\hat{q}} \prod_{i=1}^{d}(1 - \hat{\beta} p_{i,c'}) \leq (1 - \hat{\beta}\bar{p}_1)^d + \cdots + (1 - \hat{\beta}\bar{p}_{\hat{q}})^d. \tag{3.54}$$

Let $g(x) = (1 - \hat{\beta} x)^d$. Then $g(x)$ is a convex decreasing function on $(-\infty, 1]$. Since, for every $i \in [d]$, $p_{i,1} + \cdots + p_{i,q} = 1$, we have $\bar{p}_1 + \cdots + \bar{p}_{\hat{q}} = 1 - \bar{p}_q$, so Corollary 3.4.2 implies

$$\begin{aligned}\sum_{c'=1}^{\hat{q}} (1 - \hat{\beta}\bar{p}_{c'})^d &\leq (\hat{q} - 1)(1 - \hat{\beta} L)^d + (1 - \hat{\beta}(1 - \bar{p}_q - (\hat{q} - 1)L))^d \\ &\leq (\hat{q} - 1)(1 - \hat{\beta} L)^d + (1 - \hat{\beta}(1 - U - (\hat{q} - 1)L))^d .\end{aligned}$$

Plugging this into (3.54) and then into (3.53) we obtain that

$$R \leq \frac{(\hat{q} - 1)(1 - \hat{\beta} L)^d + (1 - \hat{\beta}(1 - U - (\hat{q} - 1)L))^d}{(1 - \hat{\beta} L)^d},$$

which yields the lower bound $p \geq f_{\mathsf{lower}}(q, d, \beta, U, L)$.

Next, we show that $p \leq f_{\mathsf{upper}}(q, d, \beta, U, L)$. To give an upper bound on p, it suffices to lower bound R. Since $p_{i,q} \geq L$ for every $i \in [d]$, we obtain the lower bound

$$R \geq \frac{\sum_{c'=1}^{\hat{q}} \prod_{i=1}^{d}(1 - \hat{\beta} p_{i,c'})}{(1 - \hat{\beta} L)^d}. \tag{3.55}$$

Using the inequality of arithmetic and geometric means we have

$$\sum_{c'=1}^{\hat{q}} \prod_{i=1}^{d} (1 - \hat{\beta} p_{i,c'}) \geq \hat{q} \prod_{i=1}^{d} \big((1 - \hat{\beta} p_{i,1}) \cdots (1 - \hat{\beta} p_{i,\hat{q}}) \big)^{1/\hat{q}}. \tag{3.56}$$

Now recall that $f(x) = \ln(1 - \hat{\beta}x)$, and fix any $i \in [d]$. Since $-f(x)$ is convex on $(-\infty, 1]$, Corollary 3.4.2 implies that $f(p_{i,1}) + \cdots f(p_{i,\hat{q}}) \geq (\hat{q} - 1) f(U) + f(1 - p_{i,q} - (\hat{q} - 1)U)$. Also, $f(x)$ is a decreasing function on $(-\infty, 1]$, so using the inequality $1 - p_{i,q} \leq 1 - L$ we obtain

$$f(p_{i,1}) + \cdots f(p_{i,\hat{q}}) \geq (\hat{q} - 1) f(U) + f(1 - L - (\hat{q} - 1)U) \,.$$

Thus, by exponentiating the both sides of the inequality, it gives that

$$(1 - \hat{\beta} p_{i,1}) \cdots (1 - \hat{\beta} p_{i,\hat{q}}) \geq (1 - \hat{\beta} U)^{\hat{q}-1} (1 - \hat{\beta}(1 - L - (\hat{q} - 1)U)).$$

Plugging this into (3.56) for each $i \in [d]$ and then into (3.55), we obtain that

$$R \geq \frac{\hat{q} \big((1 - \hat{\beta} U)^{\hat{q}-1} (1 - \hat{\beta}(1 - L - (\hat{q} - 1)U)) \big)^{d/\hat{q}}}{(1 - \hat{\beta} L)^d}.$$

Therefore, using (3.52), we obtain the upper bound $p \leq f_{\mathsf{upper}}(q, d, \beta, U, L)$. □

Similar to the case $q = 3$ in Chapter 3.3.2, we can also establish the following generalised properties for the generalised functions f_{upper} and f_{lower}.

Lemma 3.4.4. *For any fixed $d \geq 2$ and any fixed x and y satisfying $0 \leq y \leq x \leq 1$ and $x + \hat{q}y \leq 1 \leq \hat{q}x + y$,*

- *$f_{\mathsf{upper}}(q, d, \beta, x, y)$ is a decreasing function of β on the interval $(0, 1)$ and*
- *$f_{\mathsf{lower}}(q, d, \beta, x, y)$ is an increasing function of β on the interval $(0, 1)$.*

Proof. Recall that $\hat{\beta} = 1 - \beta$ and $\hat{q} = q - 1$. Let

$$W = \hat{q}^2 \hat{\beta} x(x - y) - \hat{q} \big(\hat{\beta}(x^2 - 3xy + x + (y - 1)y) + y \big) + (y - 1)(\hat{\beta} y - 1).$$

The derivative of f_{upper} with respect to β is given by

$$\frac{\partial f_{\mathsf{upper}}}{\partial \beta} = -{f_{\mathsf{upper}}}^2 \frac{d \big((1 - \hat{\beta} x)^{\hat{q}-1} (1 - \hat{\beta}(1 - (\hat{q} - 1)x - y)) \big)^{d/\hat{q}} W}{(1 - \hat{\beta} y)^{d+1} (1 - \hat{\beta} x)(1 - \hat{\beta}(1 - (\hat{q} - 1)x - y))}.$$

Using the conditions on x and y in the statement of the lemma, we find that

$$\begin{aligned} W &= 1 - qy + \hat{\beta}\big((qy-1)x + (\hat{q}x + y - 1)(\hat{q}-1)(x-y)\big) \\ &\geq 1 - qy + \hat{\beta}(qy-1)x \\ &= (1-qy)(1-\hat{\beta}x) \\ &\geq 0. \end{aligned}$$

We conclude that $\frac{\partial f_{\mathsf{upper}}}{\partial \beta} \leq 0$ so $f_{\mathsf{upper}}(q,d,\beta,x,y)$ is a decreasing function of β on the interval $[0,1]$. Similarly, the derivative of f_{lower} is given by

$$\frac{\partial f_{\mathsf{lower}}}{\partial \beta} = {f_{\mathsf{lower}}}^2 \frac{\big((2x + (\hat{q}-1)y - 1)(1-\hat{\beta}(1-x-(\hat{q}-1)y))^{d-1} + (\hat{q}-1)(x-y)(1-\hat{\beta}y)^{d-1}\big)d}{(1-\hat{\beta}x)^{d+1}}.$$

Using the following two inequalities

$$1 - \hat{\beta}y \geq 1 - \hat{\beta}(1-x-(\hat{q}-1)y) \geq 1 - \hat{\beta} \geq 0$$

and

$$(\hat{q}-1)(x-y) \geq 0,$$

we find that

$$\begin{aligned} &(2x + (\hat{q}-1)y - 1)(1-\hat{\beta}(1-x-(\hat{q}-1)y))^{d-1} + (\hat{q}-1)(x-y)(1-\hat{\beta}y)^{d-1} \\ \geq\ &\big((2x + (\hat{q}-1)y - 1) + (\hat{q}-1)(x-y)\big)(1-\hat{\beta}(1-x-(\hat{q}-1)y))^{d-1} \\ =\ &(qx-1)(1-\hat{\beta}(1-x-(\hat{q}-1)y))^{d-1} \geq 0, \end{aligned}$$

which yields that $\frac{\partial f_{\mathsf{lower}}}{\partial \beta} \geq 0$. So $f_{\mathsf{lower}}(q,d,\beta,x,y)$ is an increasing function of β on the interval $(0,1)$. □

Lemma 3.4.5. *For any fixed $d \geq 2$, $q \geq 3$ and $0 < \beta \leq 1$,*

1. *$f_{\mathsf{upper}}(q,d,\beta,x,y)$ is an increasing function of x when $\hat{q}x + y \geq 1$ and $x, y \in [0,1]$,*
2. *$f_{\mathsf{upper}}(q,d,\beta,x,y)$ is a decreasing function of y when $x, y \in [0,1]$.*

Proof. 1. Let $R = (1-\hat{\beta}x)^{\hat{q}-1}(1-\hat{\beta}(1-(\hat{q}-1)x-y)$. Then we obtain that

$$\frac{\partial R}{\partial x} = -\hat{\beta}^2\hat{q}(1-\hat{\beta}x)^{\hat{q}-1}(qx+y-1) \leq 0$$

since $\hat{q}x + y \geq 1$. So $f_{\mathsf{upper}} = (1-\hat{\beta}y)^d / \big((1-\hat{\beta}y)^d + \hat{q}R^{d/\hat{q}}\big)$ is an increasing function of x.

2. Let $R = 1/f_{\mathsf{upper}} - 1$. It is easy to see that

$$R = q\frac{\big((1-\hat{\beta}x)(1-\hat{\beta}(1-x-(\hat{q}-1)y))\big)^{d/\hat{q}}}{(1-\hat{\beta}y)^{2d/\hat{q}}}$$

is an increasing function of y if $x, y \in [0,1]$. Thus f_{upper} is a decreasing function of y.

□

Lemma 3.4.6. *For any fixed $d \geq 2$, $q \geq 3$ and $0 < \beta \leq 1$,*

1. *$f_{\mathsf{lower}}(q,d,\beta,x,y)$ is a decreasing function of x when $x, y \in [0,1]$,*
2. *$f_{\mathsf{lower}}(q,d,\beta,x,y)$ is an increasing function of y when $x + \hat{q}y \leq 1$ and $x, y \in [0,1]$.*

Proof. 1. Let $R = 1/f_{\mathsf{lower}} - 1$. It is easy to see that

$$R = \frac{(\hat{q}-1)(1-\hat{\beta}y)^d + (1-\hat{\beta}(1-x-(\hat{q}-1)y))^d}{(1-\hat{\beta}x)^d}$$

is an increasing function of x if $x, y \in [0,1]$. Thus f_{lower} is a decreasing function of x.

2. Let $R = (1-\hat{\beta}x)^d + (\hat{q}-1)(1-\hat{\beta}y)^d + (1-\hat{\beta}(1-x-(\hat{q}-1)y))^d$. Then we obtain that

$$\frac{\partial R}{\partial y} = d\hat{\beta}(\hat{q}-1)\big((1-\hat{\beta}(1-x-(\hat{q}-1)y))^{d-1} - (1-\hat{\beta}y)^{d-1}\big) \leq 0$$

since $1-x-(\hat{q}-1)y \geq y$. So $f_{\mathsf{lower}} = (1-\hat{\beta}x)^d/R$ is an increasing function of y.

□

Lemma 3.4.7. *For any fixed $d \geq 2$, $q \geq 3$, $0 < \beta \leq 1$ and $0 \leq y \leq x \leq 1$ such that $x + \hat{q}y \leq 1 \leq \hat{q}x + y$, we have*

$$f_{\mathsf{upper}}(q,d,\beta,x,y) + \hat{q}f_{\mathsf{lower}}(q,d,\beta,x,y) \leq 1 \leq \hat{q}f_{\mathsf{upper}}(q,d,\beta,x,y) + f_{\mathsf{lower}}(q,d,\beta,x,y).$$

Proof. For convenience, let

$$\mathsf{d_u} = (1-\hat{\beta}y)^d + \hat{q}\big((1-\hat{\beta}x)^{\hat{q}-1}(1-\hat{\beta}(1-(\hat{q}-1)x-y))\big)^{d/\hat{q}}$$

and

$$\mathsf{d_l} = (1-\hat{\beta}x)^d + (\hat{q}-1)(1-\hat{\beta}y)^d + (1-\hat{\beta}(1-x-(\hat{q}-1)y))^d.$$

Using Jensen's inequality and the condition $x + \hat{q}y \leq 1 \leq \hat{q}x + y$ we have the following inequalities:

1. $\mathsf{d}_{\mathsf{l}} \le (1-\hat{\beta}x)^d + \hat{q}(1-\hat{\beta}y)^d$;
2. $\mathsf{d}_{\mathsf{l}} \ge (1-\hat{\beta}x)^d + (\hat{q}-1)(1-\hat{\beta}y)^d + (1-\hat{\beta}((\hat{q}-2)(x-y)+x))^d \ge (1-\hat{\beta}y)^d + \hat{q}(1-\hat{\beta}x)^d$;
3. $\mathsf{d}_{\mathsf{u}} \ge (1-\hat{\beta}y)^d + \hat{q}(1-\hat{\beta}x)^d$;
4. $\mathsf{d}_{\mathsf{u}} - ((1-\hat{\beta}x)^d + \hat{q}(1-\hat{\beta}y)^d) = (1-\hat{\beta}x)^d \Big(\hat{q} \Big(\frac{1-\hat{\beta}(1-(\hat{q}-1)x-y)}{1-\hat{\beta}x} \Big)^{\frac{d}{\hat{q}}} - (\hat{q}-1)\frac{(1-\hat{\beta}y)^d}{(1-\hat{\beta}x)^d} - 1 \Big) \le 0$.

Here the last inequality is implied by

$$\begin{aligned}
(\hat{q}-1)\frac{(1-\hat{\beta}y)^d}{(1-\hat{\beta}x)^d} + 1 &\ge \hat{q}\Big(\frac{1-\hat{\beta}y}{1-\hat{\beta}x}\Big)^{(\hat{q}-1)d/\hat{q}} = \hat{q}\Big(1+\frac{\hat{\beta}(x-y)}{1-\hat{\beta}x}\Big)^{(\hat{q}-1)d/\hat{q}} \\
&\ge \hat{q}\Big(1+(\hat{q}-1)\frac{\hat{\beta}(x-y)}{1-\hat{\beta}x}\Big)^{d/\hat{q}} \\
&= \hat{q}\Big(\frac{1-\hat{\beta}(1-(\hat{q}-1)x-y+x+\hat{q}y-1)}{1-\hat{\beta}x}\Big)^{d/\hat{q}} \\
&\ge \hat{q}\Big(\frac{1-\hat{\beta}(1-(\hat{q}-1)x-y)}{1-\hat{\beta}x}\Big)^{d/\hat{q}}.
\end{aligned}$$

Thus, we conclude that

$$\hat{q}f_{\mathsf{lower}} + f_{\mathsf{upper}} \le \frac{\hat{q}(1-\hat{\beta}x)^d + (1-\hat{\beta}y)^d}{\min\{\mathsf{d}_{\mathsf{l}}, \mathsf{d}_{\mathsf{u}}\}} \le 1\,,$$

and

$$\hat{q}f_{\mathsf{upper}} + f_{\mathsf{lower}} \ge \frac{\hat{q}(1-\hat{\beta}y)^d + (1-\hat{\beta}x)^d}{\max\{\mathsf{d}_{\mathsf{l}}, \mathsf{d}_{\mathsf{u}}\}} \ge 1\,.$$

□

Lemma 3.4.8. *For any fixed $d \ge 2$, $q \ge 3$, $0 < \beta \le 1$ and $0 \le y_1 \le y_2 \le x_2 \le x_1 \le 1$ such that $\hat{q}y_1 + x_1 \le 1 \le \hat{q}x_1 + y_1$ and $\hat{q}y_2 + x_2 \le 1 \le \hat{q}x_2 + y_2$, we have*

$$f_{\mathsf{upper}}(q,d,\beta,x_2,y_2) \le f_{\mathsf{upper}}(q,d,\beta,x_1,y_1) \quad \textit{and} \quad f_{\mathsf{lower}}(q,d,\beta,x_2,y_2) \ge f_{\mathsf{lower}}(q,d,\beta,x_1,y_1).$$

Proof. Using Lemma 3.4.5, 3.4.6 and the condition in this lemma, we find that

$$f_{\mathsf{upper}}(q,d,\beta,x_2,y_2) \le f_{\mathsf{upper}}(q,d,\beta,x_1,y_2) \le f_{\mathsf{upper}}(q,d,\beta,x_1,y_1)\,,$$

and

$$f_{\mathsf{lower}}(q,d,\beta,x_2,y_2) \ge f_{\mathsf{lower}}(q,d,\beta,x_2,y_1) \ge f_{\mathsf{lower}}(q,d,\beta,x_1,y_1)\,.$$

□

Then for any $q \geq 3$, $\beta > 0$ and $d \geq 2$, we define generalised sequences u_n and ℓ_n:

$$\begin{cases} u_0(q,d,\beta) = 1, \\ \ell_0(q,d,\beta) = 0, \end{cases}$$

and for every non-negative integer n,

$$\begin{cases} u_{n+1}(q,d,\beta) &= f_{\mathsf{upper}}(q,d,\beta,u_n(q,d,\beta),\ell_n(q,d,\beta)), \\ \ell_{n+1}(q,d,\beta) &= f_{\mathsf{lower}}(q,d,\beta,u_n(q,d,\beta),\ell_n(q,d,\beta)). \end{cases} \tag{3.57}$$

Then Lemma 3.3.10, 3.3.11 and 3.3.15 also hold for the generalised version of u_n and ℓ. The proofs are the same except replacing 3 by q.

3.4.2 The case in which d is small

In this Chapter, we will prove the uniqueness threshold for $q = 4$ and $d \in \{4,5,6,7,8,9\}$. First, by Mathematica and brute-force, we have the following lemma:

Lemma 3.4.9. *The pairs $(q,d) = (4,4)$, $(q,d) = (4,5)$ and $(q,d) = (4,6)$ satisfy Condition 3.2.4 for all $\alpha > 1$.*

Proof. By symmetry among the colours, it suffices to verify the condition for colours $c_1 = 1$ and $c_2 = q$. We use Mathematica to enumerate all possible $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha)$ and then verify the following inequality

$$\forall\, \alpha > 1,\ h_{1,q,\beta_*}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) < \alpha^{1/d}\,.$$

For $(q,d) = (4,4)$, $(q,d) = (4,5)$ and $(q,d) = (4,6)$, all the resulting inequalities are satisfied. The code can be found in Chapter 8.1.13. □

For some larger d, we can do the following speedup computation. Similar to Lemma 3.3.28, we first bound the range of the ratio γ by calculating two sequences numerically using Mathematica. Let

$$\begin{cases} u'_0(d) = 1 \\ \ell'_0(d) = 0 \end{cases}$$

and for every non-negative integer n,

$$\begin{cases} u'_{n+1}(d) = \dfrac{\lceil 10000\, f_{\mathsf{upper}}(4,d,\beta_*(d),u'_n(d),\ell'_n(d))\rceil}{10000} \\ \ell'_{n+1}(d) = \dfrac{\lfloor 10000\, f_{\mathsf{lower}}(4,d,\beta_*(d),u'_n(d),\ell'_n(d))\rfloor}{10000} \end{cases}.$$

We have the following lemma, which is proved by brute force.

Lemma 3.4.10. *For every integer $d \in \{7,8,9\}$ and every integer $n \in \{0,\dots,60\}$, we have $u'_n(d) \geq u'_{n+1}(d)$, $\ell'_n(d) \leq \ell'_{n+1}(d)$, $3u'_n(d) + \ell'_n(d) \geq 1 \geq 3\ell'_n(d) + u_n(d)$ and $u'_{60}(d)/\ell'_{60}(d) \leq 41/10$.*

Proof. In Appendix 8.1.14, we use Mathematica to compute all the values $u'_n(d)$ and $\ell'_n(d)$ for $n \in \{0,\dots,60\}$ and $d \in \{7,8,9\}$. We then check that all of the desired inequalities hold. □

We next show that the sequences $\{u'_n(d)\}$ and $\{\ell'_n(d)\}$ bound the sequences $\{u_n(d,\beta_*(d))\}$ and $\{\ell_n(d,\beta_*(d))\}$ for $n \leq 60$.

Lemma 3.4.11. *For every integer $d \in \{7,8,9\}$ and every integer $n \in \{0,\dots,60\}$, we have $u'_n(d) \geq u_n(d,\beta_*(d))$ and $\ell'_n(d) \leq \ell_n(d,\beta_*(d))$.*

Proof. Fix d to be an integer in $\{7,8,9\}$. Since d is fixed, we simplify the notation by writing u_n for $u_n(d,\beta_*(d))$, u'_n for $u'_n(d)$, ℓ_n for $\ell_n(d,\beta_*(d))$, ℓ'_n for $\ell'_n(d)$, $f_{\mathsf{upper}}(x,y)$ for $f_{\mathsf{upper}}(q,d,\beta_*(d),x,y)$ and $f_{\mathsf{lower}}(x,y)$ for $f_{\mathsf{lower}}(q,d,\beta_*(d),x,y)$.
We prove the lemma by induction on n. For the base case $n=0$, we have $u_n = u'_n = 1$ and $\ell_n = \ell'_n = 0$. For the inductive step, suppose $n > 0$. By Lemma 3.3.11 and Lemma 3.3.28, we have

$$3\ell_{n-1} + u_{n-1} \leq 1 \leq 3u_{n-1} + \ell_{n-1}, \quad \text{and} \quad 3\ell'_{n-1} + u'_{n-1} \leq 1 \leq 3u'_{n-1} + \ell'_{n-1}.$$

By the induction hypothesis, we have

$$\ell'_{n-1} \leq \ell_{n-1} \leq u_{n-1} \leq u'_{n-1}.$$

Using Lemma 3.3.9, we therefore obtain that

$$\begin{aligned} u'_n &\geq f_{\mathsf{upper}}(u'_{n-1},\ell'_{n-1}) \geq f_{\mathsf{upper}}(u_{n-1},\ell_{n-1}) = u_n, \text{ and} \\ \ell'_n &\leq f_{\mathsf{lower}}(u'_{n-1},\ell'_{n-1}) \leq f_{\mathsf{lower}}(u_{n-1},\ell_{n-1}) = \ell_n. \end{aligned}$$

This completes the proof. □

Combining Lemma 3.4.10, Lemma 3.4.11 with Lemma 3.3.11, we can bound the ratio:

Corollary 3.4.12. *For every integer $d \in \{7,8,9\}$ and $n \geq 60$, we have*

$$\gamma(4,d,\beta_*(d),n) \leq \frac{u'(60)}{\ell'(60)} \leq \frac{41}{10}.$$

Now, similar to Lemma 3.3.33, we try to simplify $h_{c_1,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)})$. Note that $\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)} \in \mathrm{Ex}_{c_2}(\alpha)$. So we have $p_c^{(k)}/p_{c_2}^{(k)} = 1$ or α for all $c \in [q]$ and there exists $c_0 \in [q]$ such that $p_{c_0}^{(k)}/p_{c_2}^{(k)} = \alpha$.

Since $q = 4$, without loss of generality, let $\{c_1, c_2, c_3, c_4\}$ denote 4 colours. Now there are 3 cases:

1. Suppose that $p_{c_1}^{(k)} = p_{c_3}^{(k)} = p_{c_4}^{(k)} = \alpha p_{c_2}^{(k)}$. Then for all $i \in \{1,3,4\}$ we define

$$x(\alpha,\beta) = 1 - \frac{(1-\beta)(p_{c_i}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)} + p_{c_4}^{(k)}} = 1 - \frac{(1-\beta)(\alpha-1)}{\beta+3\alpha}. \tag{3.58}$$

2. Suppose that there exists exactly 1 colour $c \in \{c_1, c_3, c_4\}$ such that $p_c^{(k)} = p_{c_2}^{(k)}$. Then for all $c_i \neq c, c_2$ we define

$$y(\alpha,\beta) = 1 - \frac{(1-\beta)(p_{c_i}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)} + p_{c_4}^{(k)}} = 1 - \frac{(1-\beta)(\alpha-1)}{\beta+2\alpha+1} \tag{3.59}$$

   and it is easy to see that $1 - \frac{(1-\beta)(p_c^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)} + p_{c_4}^{(k)}} = 1$.

3. Suppose that there exists exactly 1 colour $c \in \{c_1, c_3, c_4\}$ such that $p_c^{(k)} = \alpha p_{c_2}^{(k)}$. Then we define

$$z(\alpha,\beta) = 1 - \frac{(1-\beta)(p_c^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)} + p_{c_4}^{(k)}} = 1 - \frac{(1-\beta)(\alpha-1)}{\beta+\alpha+2} \tag{3.60}$$

   and it is easy to see that for all $c_i \neq c, c_2$, we have $1 - \frac{(1-\beta)(p_{c_i}^{(k)} - p_{c_2}^{(k)})}{\beta p_{c_2}^{(k)} + p_{c_1}^{(k)} + p_{c_3}^{(k)} + p_{c_4}^{(k)}} = 1$.

Further, there are 7 detailed cases of $p_{c_1}^{(k)}, p_{c_3}^{(k)}, p_{c_4}^{(k)}$. For each case, we count the number of $k \in [d]$ for which that case occurs, respectively:

$p_{c_1}^{(k)}/p_{c_2}^{(k)}$	$p_{c_3}^{(k)}/p_{c_2}^{(k)}$	$p_{c_4}^{(k)}/p_{c_2}^{(k)}$	number of occurrence
α	α	α	d_0
α	α	1	d_1
α	1	α	d_2
1	α	α	d_3
α	1	1	d_4
1	α	1	d_5
1	1	α	d_6

where $d = d_0 + d_1 + d_2 + d_3 + d_4 + d_5 + d_6$.
Then using the definition of $g_{c_i,c_j,\beta}$ and plugging (3.58), (3.59) and (3.60), we have

$$\begin{aligned} g_{c_1,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) &= x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d_1+d_2} z(\alpha,\beta)^{d_5}\,, \\ g_{c_3,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) &= x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d_1+d_3} z(\alpha,\beta)^{d_6}\,, \\ g_{c_4,c_2,\beta}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) &= x(\alpha,\beta)^{d_0} y(\alpha,\beta)^{d_2+d_3} z(\alpha,\beta)^{d_7}\,. \end{aligned} \tag{3.61}$$

To analyse $h_{c_1,c_2,\beta}$, let $\mathbf{d} = (d_0,\dots,d_6)$ and

$$\phi(\alpha,\beta,\mathbf{d}) = h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)},\dots,\mathbf{p}^{(d)}) = 1 + \frac{(1-\beta)(1-g_{c_1,c_2,\beta})}{\beta + g_{c_1,c_2,\beta} + g_{c_3,c_2,\beta} + g_{c_4,c_2,\beta}}\,.$$

Now we can prove the following lemma:

Lemma 3.4.13. *The pairs $(q,d) = (4,7)$, $(q,d) = (4,8)$ and $(q,d) = (4,9)$ satisfy Condition 3.2.4 for all $1 < \alpha \le 41/10$.*

Proof. It suffice to show that for all $\alpha > 1$, $\beta \ge 1 - 4(d+1)$ and all possible $\mathbf{d}$ such that $d_0 + \cdots + d_6 = d$, it holds that

$$\phi(\alpha,\beta,\mathbf{d}) < \alpha^{1/d}\,.$$

Note that $x(\alpha,\beta)$, $y(\alpha,\beta)$, $z(\alpha,\beta)$ are all increasing functions of β. So $\phi(\alpha,\beta,\mathbf{d})$ is a decreasing function of β, and thus it is sufficient to show that

$$\phi(\alpha,\beta_*(d),\mathbf{d}) < \alpha^{1/d}\,.$$

Again, we use Mathematica to enumerate all possible tuples $\mathbf{d}$ and then for each $\mathbf{d}$ verify the inequality above. For $(q,d) = (4,7)$, $(q,d) = (4,8)$ and $(q,d) = (4,9)$, all the resulting inequalities are satisfied. The code can be found in Chapter 8.1.15. □

Combining Lemma 3.4.9 with Lemma 3.4.13, we conclude that when $q = 4, d \in \{4,5,6,7,8,9\}$, (q,d) satisfies Condition 3.2.4 for all $\alpha > 1$ if $d \le 6$ and all $1 < \alpha \le 41/10$ if $d \ge 7$. Then the proof of Theorem 3.1.2 follows directly from Lemma 3.4.12 and Proposition 3.2.5.

Chapter 4

Counting in Uniqueness I : Counting Four-colourings on Cubic Graphs

This chapter is based on the joint work with Pinyan Lu, Chihao Zhang and Minshen Zhu [56]. Our goal is to design an FPTAS for counting the number of proper 4-colourings on graphs whose maximum degree is at most 3.

4.1 Introduction

The problem of establishing the connection between the condition of uniqueness regime and the condition of approximability of counting problems has played a significant role in the study of counting complexity. In this chapter, we will discuss a special case of the antiferromagnetic Potts model – the proper q-colouring problem. It is easy to see that the proper q-colouring model corresponds to the Potts model with $\beta = 0$ (see Chapter 1.2.1). The problem of counting proper q-colourings has been extensively studied in computer science and statistical physics. It is known to be #P-hard for $q \geq 3$ even on graphs with bounded maximum degree $\Delta \geq 3$ [12]. So designing efficient approximation algorithms becomes an important study to this problem. Another well-known result is that the uniqueness condition for the q-colouring model on infinite d-ary trees is exactly $q \geq d + 2$ [50]. This fact supports the conjecture that $q = \Delta + 1$ is the optimal bound for approximate counting the number of proper q-colourings in general graphs of maximum degree Δ. However this goal is too difficult to achieve.

In this chapter we will present a new result. We design an FPTAS for approximate counting proper 4-colourings in the case of $\Delta = 3$. It is the first approximation algorithm

to achieve the optimal bound of $q = \Delta + 1$ and this algorithm is based on correlation decay approach (the correlation decay approach is introduced in Chapter 2.2.1).

4.1.1 The result

Theorem 4.1.1. *There exists an* FPTAS *to compute the number of proper four colourings on graphs with maximum degree three.*

Although we start with a standard graph colouring instance, where each vertex can choose from the same set of 4 different colours, we need to modify the colour lists of different vertices during our algorithm. Therefore our algorithm should work on list-colouring instances in general, where different vertices may have different colour lists. More formally, a list-colouring instance is specified by a graph-list pair (G, L), where $G = (V, E)$ is an undirected graph and $L : V \to 2^{[q]}$ associates each vertex v with a colour list $L(v) \subseteq [q]$. A proper colouring of (G, L) is an assignment $c : V \to [q]$ such that (1) $c(v) \in L(v)$ for every $v \in V$ and (2) no two ends of an edge share the same colour, i.e., $c(u) \neq c(v)$ for every $e = (u, v) \in E$.
The Gibbs distribution of an instance (G, L) is the uniform distribution over all proper colourings of (G, L). For every vertex $v \in V$ and colour $i \in [q]$, we use $\Pr_{G,L}[c(v) = i]$ to denote the marginal probability that the vertex v is coloured i in the Gibbs distribution. If there exists an efficient algorithm to estimate the marginal probability $\Pr_{G,L}[c(v) = i]$ with sufficient accuracy, then we can construct an FPTAS to count the number of proper colourings.

Lemma 4.1.2. *Suppose there exists an algorithm to compute a* $(1 \pm \varepsilon)$ *approximation of* $\Pr_{G,L}[c(v) = i]$ *for every list-colouring instance* (G, L) *with* $G = (V, E)$*,* $q = 4$*,* $\Delta = 3$*,* $|L(v)| \geq d_v + 1$ *for every* $v \in V$*, and every* $i \in [q]$ *in time* $\mathrm{poly}(|V|, \frac{1}{\varepsilon})$*. Then there exists an* FPTAS *to compute the number of proper* 4*-colourings on graphs with maximum degree three.*

The proof Lemma 4.1.2 is routine. It is almost the same as Lemma 2.1.1. (Or see e.g. [39].) Therefore, the remaining task is to approximate $\Pr_{G,L}[c(v) = i]$ for instances satisfying the conditions stated in Lemma 4.1.2 and the following theorem is equivalent to our main theorem (Theorem 4.1.1).

Lemma 4.1.3. *Let $q = 4$. There exists an algorithm such that for every list-colouring instance (G, L) with $G = (V, E)$ and maximum degree at most 3, every vertex $v \in V$, every colouring $i \in L(v)$ and every $0 < \varepsilon < 1$, it computes a number $\hat{p}$ in time $\mathrm{poly}\left(|V|, \frac{1}{\varepsilon}\right)$ satisfying*

$$(1-\varepsilon)\hat{p} \leq \Pr_{G,L}[c(v) = i] \leq (1+\varepsilon)\hat{p}.$$

Recall that our algorithm is based on correlation decay. So we will first describe our algorithm in Chapter 4.2. Then we verify the correlation property in Chapter 4.3 and 4.4, and finally complete the proof of Lemma 4.1.3 and Theorem 4.1.1 in Chapter 4.5.

4.1.2 Prior works

A number of literature has been devoted to the design of approximation algorithms of counting proper q-colourings [11, 62, 43, 42, 47, 12, 77, 23, 24, 44, 22]. The main algorithmic tool used in these works is the method of Markov chain Monte Carlo (MCMC), which is based on the simulation of a Markov chain on all proper q-colourings of a graph G whose stationary distribution is the uniform distribution. Although the Markov chains themselves are usually quite simple, it is challenging to prove the rapid mixing property of the chains and the interplay between the number q of colours and the maximum degree Δ of the graph G turns out to be a key measure for such property to hold.

Glauber dynamics is a natural Markov chain to sample colourings and it converges to the uniform distribution as long as $q \geq \Delta + 2$. Jerrum [47] and Salas and Sokal [68] independently showed that Glauber dynamics mixes rapidly if $q > 2\Delta$. The bound of 2Δ was considered as a barrier for the analysis of Glauber dynamics and was even conjectured as a threshold for the rapid mixing property to hold for a period of time. Later, the conjecture was refuted by Bubley et al. [12] by showing that Glauber dynamics indeed rapidly mixes when $\Delta = 3$ and $q = 5$. It is worth to note that this result attains the ergodicity threshold for Glauber dynamics ($q \geq \Delta + 2$) and thus it is the best one can achieve via this method. For general Δ, the state-of-the-art requires that $q > \frac{11}{6}\Delta$, due to Vigoda's work [77]. Recently, the constant was reduced to $\frac{11}{6} - \varepsilon_0$ for a very small constant ε_0 [16].

All the above algorithms based on MCMC provide randomized algorithms. Later, a deterministic polynomial-time approximation algorithm (FPTAS) was obtained in [39] when $q \geq 2.8432\Delta + c$ for some sufficiently large constant c on triangle-free graphs. The

bound was improved to $q \geq 2.581\Delta + 1$ on general graphs [57]. These new deterministic algorithms are based on the correlation decay techniques. The best known bound for deterministic algorithms is $q \geq 2\Delta$ for general graphs while $q \geq 1.7633\Delta + c$ for triangle-free graphs and some constant c [55]. Their algorithms are based on a so-called "polynomial interpolation" method in the zero-free region for the partition function of colourings on the complex plane.

Our algorithm is also based on correlation decay method. As the first algorithm achieving the optimal bound, we view it as a substantial step towards the optimal counting algorithms for general graphs. It overcomes an intrinsic barrier of MCMC (Glauber dynamics) based algorithms. For the case of $q = \Delta + 1$, the Glauber dynamics Markov chain is not ergodic and thus its stationary distribution is not unique.

4.2 The approximation algorithm

4.2.1 Tree recursion and our observations

Our approach to design the FPTAS is based on the correlation decay property and the amortised analysis, which were introduced in Chapter 2.2.1 and Chapter 2.2.2. Thus, the first step is to establish a tree recursion, by which we can compute marginal probabilities on trees. Since proper q-colouring is a special case of antiferromagnetic Potts model, we can claim the following recursion, which is similar to Proposition 3.1.3 and first appeared in [39].

Let (G, L) (where $G = (V, E)$) be an instance of list-colouring and $v \in V$ be a vertex. Let $N(v) = \{v_1, \ldots, v_d\}$ denote the set of neighbours of v in G, where d is the degree of v and let G_v be the graph obtained from G by removing vertex v and all edges incident on v. For every $k \in [d]$ and $i \in [q]$, let

$$L_{k,i}(u) = \begin{cases} L(u) \setminus \{i\}, & \text{if } u = v_\ell \text{ for some } \ell < k, \\ L(u) & \text{otherwise} \end{cases} \tag{4.1}$$

be colour lists. Then we have the following lemma:

Lemma 4.2.1. *Assuming above notations we have*

$$\Pr_{G,L}[c(v) = i] = \frac{\prod_{k=1}^{d} \left(1 - \Pr_{G_v, L_{k,i}}[c(v_k) = i]\right)}{\sum_{j \in L(v)} \prod_{k=1}^{d} \left(1 - \Pr_{G_v, L_{k,j}}[c(v_k) = j]\right)} \tag{4.2}$$

Proof. Let $Z_{G,L}$ denote the number of proper colourings on (G, L) and let $w_{G,L}(c(u) = i)$ denote the number of proper colourings on (G, L) that assigns vertex u with colour i for every vertex $u \in V$ and every colour $i \in [q]$. Then

$$\begin{aligned}
\Pr_{G,L}[c(v) = i] &= \frac{w_{G,L}(c(v) = i)}{\sum_{j \in L(v)} w_{G,L}(c(v) = j)} = \frac{w_{G_v,L}\left(\bigwedge_{k\in[d]} c(v_k) \neq i\right)}{\sum_{j\in L(v)} w_{G_v,L}\left(\bigwedge_{k\in[d]} c(v_k) \neq j\right)} \\
&= \frac{\Pr_{G_v,L}\left[\bigwedge_{k\in[d]} c(v_k) \neq i\right]}{\sum_{j\in L(v)} \Pr_{G_v,L}\left[\bigwedge_{k\in[d]} c(v_k) \neq j\right]} \\
&= \frac{\prod_{k=1}^{d} \Pr_{G_v,L}\left[c(v_k) \neq i \mid \bigwedge_{\ell<k} c(v_\ell) \neq i\right]}{\sum_{j\in L(v)} \prod_{k=1}^{d} \Pr_{G_v,L}\left[c(v_k) \neq j \mid \bigwedge_{\ell<k} c(v_\ell) \neq j\right]} \\
&= \frac{\prod_{k=1}^{d} \left(1 - \Pr_{G_v,L_{k,i}}[c(v_k) = i]\right)}{\sum_{j\in L(v)} \prod_{k=1}^{d} \left(1 - \Pr_{G_v,L_{k,j}}[c(v_k) = j]\right)}\,.
\end{aligned}$$

□

Remark. Let $G = (V, E)$. Note that for all list-colouring instances appearing in this computation tree, we have $q = 4$, $\Delta = 3$, and for every $v \in V$, $|L(v)| \geq d_v + 1$. In the definition of new colour lists (4.1), the size of the colour list associated with vertex u is reduced by one only if u is v's neighbour. However, the degrees of v's neighbours are also reduced by one in the new graph G_v. Therefore the condition $|L(u)| \geq d_u + 1$ still holds for every $u \in V \setminus \{v\}$. For every probability $\Pr_{G',L'}[c(u) = j]$ in the computation tree except the root, the degree d_u has the property $d_u \leq \Delta - 1 = 2$ since we reach this instance by removing a neighbour of u and thus the degree is reduced by at least one. A more subtle and crucial new observation is that for every probability $\Pr_{G',L'}[c(u) = j]$ on the computation tree except the root, it holds that $|L(u)| \geq d_u + 2$ since the degree of u is reduced by one while the size of the colour list for u remains the same in the definition of (4.1).

In all prior works of analysing (4.2), the recursion functions were viewed as functions which have dq free and independent variables. For each single variable, an identical recursion function was used to expand to a set of dq new free and independent variables. However, when we apply the recursion function (4.2), the marginal probabilities $\Pr_{G_v,L_{k,j}}[c(v_k) = j]$ for different subinstances $(G_v, L_{k,j})$ are not independent. A key idea is to consider the relations among these variables to reduce redundancy and improve the efficiency of the recursion. Here are two important observations:

- For different colours i and j, the recursions for $\Pr_{G,L}[c(v) = i]$ and $\Pr_{G,L}[c(v) = j]$

involve exactly the same set of dq variables.

- If $k = 1$, $L_{k,j}$ are identical for all colour j, and $\Pr_{G_v,L_{1,j}}[c(v_1) = j]$ are simply the marginal probabilities of vertex v_1 for different colours in the same subinstance.

Using these two observations, we can further apply the same recursion function to $\Pr_{G_v,L_{1,j}}[c(v_1) = j]$ while we use the recursion (4.2) to compute $\Pr_{G,v}[c(v) = i]$. More specifically, let d_1 be the degree of v_1 in the graph G_v and $u_1, u_2, \ldots, u_{d_1}$ be the neighbours of v_1 in the graph G_v. Denote by G_{v,v_1} the graph obtained from G_v by removing v_1 and all edges incident on v_1. For every $k \in [d_1]$ and $i \in [q]$, we use $L'_{k,i}$ to denote the colour list such that

$$L'_{k,i}(u) = \begin{cases} L(u) \setminus \{i\}, & \text{if } u = u_\ell \text{ for some } \ell < k, \\ L(u), & \text{otherwise.} \end{cases}$$

Applying recursion (4.2), we obtain that for every $j \in L(v_1)$, it holds that

$$\Pr_{G_v,L_{1,j}}[c(v_1) = j] = \frac{\prod_{k=1}^{d_1} \left(1 - \Pr_{G_{v,v_1},L'_{k,j}}[c(u_k) = j]\right)}{\sum_{l \in L(v_1)} \prod_{k=1}^{d_1} \left(1 - \Pr_{G_{v,v_1},L'_{k,l}}[c(u_k) = l]\right)}. \tag{4.3}$$

Then we substitute (4.3) into recursion (4.2) and get a new recursion for computing $\Pr_{G,L}[c(v) = i]$. We view this new recursion as one step in the computation tree and analyse its correlation decay property.

Another crucial idea in our proof is to get better bounds for variables $\Pr_{G_v,L_{k,j}}[c(v_k) = j]$ and then we only need to prove the correlation decay property in these bounded ranges. This idea was used in prior works for counting colourings and many other problems. However, in our setting of $q = 4$ and $\Delta = 3$, these values could be as large as 1 and as small as 0. These are trivial bounds for a probability in general. Here, we have the following observation to refine bounds.

Note that the situation that a marginal probability $\Pr_{G,L}[c(v) = i] = 1$ only happens at the root of the recursion tree and for all other variable the value is between 0 and $1/2$. It is possible that a marginal probability is 0 or $1/2$. Thus in general the bound cannot be improved. However, we can easily find in which cases a marginal probability $\Pr_{G,L}[c(v) = i]$ is 0 or $1/2$. So our algorithm can output the exact value in these cases. Otherwise, we can obtain better bounds (see 4.3.7).

Finally, as in most of the correlation decay approach, we use a potential function to amortise the decay rate. It remains the most important and magic ingredient of the

proof. There is no general method to design potential function. Based on some numerical computation, we propose a new potential function in our work. Comparing to the previous potential functions for colouring problem, the main new feature of our function is non-monotonicity, which captures the property of the problem. We conjecture that a potential function of a similar shape can be used for general graph colouring problem for similar set of recursions.

4.2.2 The algorithm for approximately estimate marginal probabilities

In this chapter, we describe our algorithm to estimate marginals.
The main idea of our algorithm to estimate $\Pr_{G,L}[c(v) = i]$ is to recursively apply recursions (4.2) and (4.3) up to some depth D. For the convenience of analysis, we distinguish between following cases, depending on the degree of v and its neighbours.

- Our algorithm terminates in one of the following three boundary cases.
  1. The colour i is not in the colour list $L(v)$, i.e., $i \notin L(v)$, for which we return 0;
  2. The recursion depth is 0, for which we return $\frac{1}{|L(v)|}$;
  3. The degree of v in G is 0, i.e. v is an isolated vertex, for which we return $\frac{1}{|L(v)|}$.
- If the degree of v in G is 1, the algorithm branches into three cases according to the size of $L(v)$.
  1. If $|L(v)| = 2$, we directly apply recursion (4.2);
  2. If $|L(v)| = 4$, note that the sum of the marginal probabilities of colours $j \in L(v)$ on v_1 in G_v is 1, the denominator of the recursion (4.2) becomes a constant 3;
  3. If $|L(v)| = 3$, for the same reason we can denote the denominator of the recursion (4.2) by $2 + y$, where y is the marginal probability of colour $j \in [4] \setminus L(v)$ (the absent colour) on v_1 in G_v.
- If the degree of v in G is 2 or 3, we faithfully apply recursion (4.2) and (4.3) to estimate the marginal probability. In order to simplify the analysis, we use the following convention in the case of $\deg_G(v) = 2$: Let the neighbours of v be v_1, v_2. We always assume $\deg_G(v_1) \geq \deg_G(v_2)$ and if $\deg_G(v_1) = \deg_G(v_2) = 1$, then $i \notin L(v_1)$ implies $i \notin L(v_2)$.

The whole algorithm is described below. We use procedure $P(G, L, v, i, D)$ to estimate $\Pr_{G,L}[c(v) = i]$ up to depth D.

Algorithm 1: Estimating $\Pr_{G,L}[c(v) = i]$

Input : Graph G, colour lists L, vertex v, colour i, recursion depth D;
Output : A real value $P \in [0, 1]$ estimating $\Pr_{G,L}[c(v) = i]$ up to depth D.

Function $P(G, L, v, i, D)$
begin
  if $i \notin L(v)$ **then**
    return 0;
  end
  if $D \leq 0$ **or** $\deg_G(v) = 0$ **then**
    return $\frac{1}{|L(v)|}$;
  end
  if $\deg_G(v) = 1$ **then**
    return $P_1(G, L, v, i, D)$;
  end
  if $\deg_G(v) = 2$ **then**
    return $P_2(G, L, v, i, D)$;
  end
  if $\deg_G(v) = 3$ **then**
    return $P_3(G, L, v, i, D)$;
  end
end

The functions $P_1(G, L, v, i, D)$, $P_2(G, L, v, i, D)$ and $P_3(G, L, v, i, D)$ deal with the cases of $\deg_G(v) = 1$, $\deg_G(v) = 2$ and $\deg_G(v) = 3$ respectively.

Case $\deg_G(v) = 3$**:** v has three neighbours v_1, v_2, and v_3.

Algorithm 2: Estimating $\Pr_{G,L}[c(v) = i]$ when $\deg_G(v) = 3$

Function $P_3(G, L, v, i, D)$
begin
  for $j \in L(v)$ **do**
    $x_j \leftarrow P(G_v, L_{1,j}, j, D)$;
    $y_j \leftarrow P(G_v, L_{2,j}, j, D)$;
    $z_j \leftarrow P(G_v, L_{3,j}, j, D)$;
  end
  return $\frac{(1-x_i)(1-y_i)(1-z_i)}{\sum_{j \in L(v)}(1-x_j)(1-y_j)(1-z_j)}$;
end

Case $\deg_G(v) = 2$**:** v has two neighbours v_1 and v_2. Without loss of generality, suppose that $\deg_G(v_1) \geq \deg_G(v_2)$ and v_1 has neighbours $\{u_1, \ldots, u_{d_1}\}$ in the graph G_v. We can also assume that if $\deg_G(v_1) = \deg_G(v_2) = 1$. Then $i \notin L(v_1)$ implies $i \notin L(v_2)$.

Algorithm 3: Estimating $\Pr_{G,L}[c(v) = i]$ when $\deg_G(v) = 2$

Function $P_2(G, L, v, i, D)$
begin
  for $j \in L(v)$ **do**
    $f_j \leftarrow 0$;
    $y_j \leftarrow P(G_v, L_{2,j}, v_2, j, D-1)$;
    if $j \in L(v_1)$ **then**
      for $k \in [d_1]$, $w \in L(v_1)$ **do**
        $x_{k,w} \leftarrow P(G_{v,v_1}, L'_{k,w}, u_k, w, D-1)$;
      end
      $f_j \leftarrow \frac{\prod_{k=1}^{d_1}(1-x_{k,j})}{\sum_{w \in L(v_1)} \prod_{k=1}^{d_1}(1-x_{k,w})}$;
    end
  end
  return $\frac{(1-f_i)(1-y_i)}{\sum_{j \in L(v)}(1-f_j)(1-y_j)}$;
end

Case $\deg_G(v) = 1$**:** v has only one neighbour v_1.

Algorithm 4: Estimating $\Pr_{G,L}[c(v) = i]$ when $\deg_G(v) = 1$

Function $P_1(G, L, v, i, D)$ **begin**
  if $|L(v)| = 2$ **then**
    $x \leftarrow P(G_v, L_{1,i}, v_1, i, D-1)$;
    $y \leftarrow P(G_v, L_{1,j}, v_1, j, D-1)$ where j is the other colour in $L(v)$;
    return $\frac{1-x}{2-x-y}$;
  end
  if $|L(v)| = 4$ **then**
    $x \leftarrow P(G_v, L_{1,i}, v_1, i, D-1)$;
    return $\frac{1-x}{3}$;
  end
  `/* In the following case,` $|L(v)| = 3$. `*/`
  $x \leftarrow P(G_v, L_{1,i}, v_1, i, D-1)$;
  $y \leftarrow P(G_v, L_{1,j}, v_1, j, D-1)$ where j is the colour in the singleton set $[4] \setminus L(v)$;
  return $\frac{1-x}{2+y}$;
end

Proposition 4.2.2. *Let $q = 4$. Given a list-colouring instance $(G = (V, E), L)$ with maximum degree 3, a vertex $v \in V$ satisfying $\deg_G(v) \leq 2$ and $|L(v)| \geq \deg_G(v) + 2$, a*

nonnegative integer D, we have

$$\sum_{i=1}^{4} P(G, L, v, i, D) = 1$$

Proof. We will prove the proposition by induction on D. When $D = 0$ we have $\sum_{i=1}^{4} P(G, L, v, i, 0) = \sum_{i \in L(v)} \frac{1}{|L(v)|} = 1$. Suppose the proposition holds for $D - 1$. To obtain the proof for D, we will do a case analysis depending on the degree of v.

1. $\deg_G(v) = 0$.

   Clearly $\sum_{i=1}^{4} P(G, L, v, i, D) = \sum_{i \in L(v)} \frac{1}{|L(v)|} = 1$.

2. $\deg_G(v) = 1$.

   Let $x_i = P(G_v, L_{1,i}, v_1, i, D - 1)$. By definition we have $L_{1,i} = L$ for all $i \in [4]$. Therefore $\sum_{i=1}^{4} x_i = \sum_{i=1}^{4} P(G_v, L, v_1, i, D - 1) = 1$ by induction hypothesis.

   If $|L(v)| = 4$, $\sum_{i=1}^{4} P(G, L, v, i, D) = \sum_{i=1}^{4} \frac{1-x_i}{3} = \frac{4-1}{3} = 1$.

   If $|L(v)| = 3$, assume $j \notin L(v)$. Then $\sum_{i=1}^{4} P(G, L, v, i, D) = \sum_{i \in L(v)} \frac{1-x_i}{2+x_j} = \frac{3-(1-x_j)}{2+x_j} = 1$.

3. $\deg_G(v) = 2$.

   In this case $|L(v)| = 4$. So

$$\sum_{i=1}^{4} P(G, L, v, i, D) = \sum_{i=1}^{4} \frac{(1 - f_i)(1 - y_i)}{\sum_{j \in L(v)} (1 - f_j)(1 - y_j)} = \frac{\sum_{i=1}^{4} (1 - f_i)(1 - y_i)}{\sum_{j=1}^{4} (1 - f_j)(1 - y_j)} = 1 .$$

□

Using the same proof, we can also have $\sum_{j=1}^{4} f_j = 1$, where f_j is defined in Algorithm 3. As the end of this section, we claim that by choosing appropriate D our algorithm can compute a number $\hat{p}$ in time $\mathrm{poly}\left(|V|, \frac{1}{\varepsilon}\right)$ satisfying $(1-\varepsilon)\hat{p} \leq \Pr_{G,L}[c(v) = i] \leq (1+\varepsilon)\hat{p}$, which can prove Lemma 4.1.3. The proof of our claim and Lemma 4.1.3 will be given in Chapter 4.5.

4.3 Bounding the marginal probability

In this chapter, we introduce upper and lower bounds for values computed in the algorithm. These bounds will play a crucial role in our proof.

Definition 4.3.1. We say a tuple $(G = (V, E), L, v \in V)$ (a list-colouring instance together with a vertex in the graph) is *reachable* if the following conditions are satisfied: $\deg_G(u) \leq 3$ and $|L(u)| \geq \deg_G(u) + 1$ for every $u \in V$, $\deg_G(v) \leq 2$ and $|L(v)| \geq \deg_G(v) + 2$.

It follows the discussion in Chapter 4.2.1 that for all probabilities $\Pr_{G,L}[c(v) = i]$ that appeared in the computation tree except the root, (G, L, v) is reachable.

Proposition 4.3.2. *Let (G, L, v) be a reachable tuple, $i \in [4]$ be a colour, and D be a nonnegative integer. Then it holds that*

$$0 \leq P(G, L, v, i, D) \leq \frac{1}{2}.$$

Proof. We now prove the proposition by induction on D. For the base case $D = 0$, $P(G, L, v, i, D)$ will return $\frac{1}{|L(v)|}$ if $D = 0$, so the proposition holds since $|L(v)| \geq 2$. Assume the induction hypothesis that the proposition holds for $D - 1$. We will show that the proposition is also true for D. We now do a case analysis depending on the degree of v.

1. $\deg_G(v) = 0$.

   In this case $P(G, L, v, i, D)$ will return $\frac{1}{|L(v)|}$ where $|L(v)| \geq \deg_G(v) + 2 = 2$, hence we have $0 \leq P(G, L, v, i, D) \leq \frac{1}{2}$.

2. $\deg_G(v) = 1$.

   Let $x = P(G_v, L_{1,i}, v_1, i, D - 1)$ and $y = P(G_v, L_{1,j}, j, D - 1)$, as defined in Algorithm 4. Then $0 \leq x, y \leq \frac{1}{2}$ by induction hypothesis.

   According to algorithm, $P(G, L, v, i, D)$ will return $\frac{1-x}{3}$ or $\frac{1-x}{2+y}$, and in both cases this return value is bounded by $\frac{1}{2}$ given $x, y \geq 0$.

3. $\deg_G(v) = 2$.

   Let f_j, $x_{k,w}$ and y_j be the variables defined in Algorithm 3. By induction hypothesis we have $0 \leq x_{k,w}, y_j \leq \frac{1}{2}$. As for f_j, we need to further do a case analysis depending on d_1.

   If $d_1 \in \{0, 1\}$ then $f_j \leq \frac{1}{2}$ immediately follows, as we have already seen in previous

two cases. If $d_1 = 2$, we also have

$$\begin{aligned} f_j &= \frac{\prod_{k=1}^{2}(1-x_{k,j})}{\sum_{w\in L(v_1)}\prod_{k=1}^{2}(1-x_{k,w})} \\ &\leq \frac{1-x_{1,j}}{(1-x_{1,j})+\frac{1}{2}\sum_{w\in L(v_1)\setminus\{j\}}(1-x_{1,w})} \\ &= \frac{1-x_{1,j}}{(1-x_{1,j})+\frac{1}{2}(2+x_{1,j})} \\ &\leq \frac{1}{2}. \end{aligned}$$

Here we used the fact that $\sum_{w\in L(v_1)} x_{1,w} = 1$, since $|L(v_1)| = 4$ when $d_1 = 2$. Similarly we have

$$P(G,L,v,i,D) = \frac{(1-f_i)(1-y_i)}{\sum_{j\in L(v)}(1-f_j)(1-y_j)} \leq \frac{1}{2}.$$

□

Proposition 4.3.3. *Let (G,L,v) be a reachable tuple, $i \in L(v)$ be a colour, and D be a nonnegative integer. Then it holds that*

1. *if $\deg_G(v) = 2$, then*
$$P(G,L,v,i,D) \geq \frac{1}{13};$$
2. *if $\deg_G(v) \leq 1$, then*
$$P(G,L,v,i,D) \geq \frac{1}{6}.$$

Proof. If $\deg_G(v) = 0$ or $D = 0$, we have $P(G,L,v,i,D) = \frac{1}{|L(v)|} \geq \frac{1}{4}$. In the following, we assume $D \geq 1$ and $\deg_G(v) \geq 1$. We now again do a case analysis depending on the degree of v.

1. $\deg_G(v) = 2$

   It must be the case that $|L(v)| = 4$. Therefore we have

$$\begin{aligned} P(G,L,v,i,D) &= \frac{(1-f_i)(1-y_i)}{(1-f_i)(1-y_i)+\sum_{j\in L(v)\setminus\{i\}}(1-f_j)(1-y_j)} \\ &\geq \frac{\left(1-\frac{1}{2}\right)^2}{\left(1-\frac{1}{2}\right)^2+\sum_{j\in L(v)\setminus\{i\}}(1-0)} = \frac{\frac{1}{4}}{\frac{1}{4}+3} = \frac{1}{13}. \end{aligned}$$

   The upper bound $\frac{1}{2}$ for f_j and y_j is guaranteed by Proposition 4.3.2.

2. $\deg_G(v) = 1$

   If $|L(v)| = 4$, $P(G, L, v, i, D) = \frac{1-x}{3} \geq \frac{1-\frac{1}{2}}{3} = \frac{1}{6}$.

   If $|L(v)| = 3$, $P(G, L, v, i, D) = \frac{1-x}{2+y} \geq \frac{1-\frac{1}{2}}{2+\frac{1}{2}} = \frac{1}{5} > \frac{1}{6}$.

□

Note that overall we have lower bounds $\frac{1}{13}$ for $P(G, L, v, i, D)$, regardless of the degree of v. Furthermore, for f_j's defined in Algorithm 3 we can draw a similar conclusion: if $j \in L(v_1)$ then $f_j \geq \frac{1}{13}$.

Proposition 4.3.4. *Let (G, L, v) be a reachable tuple and D be a nonnegative integer. Then for every colour $i \in [4]$ such that $i \in L(u)$ for some neighbour u of v, we have*

1. *if $\deg_G(v) = 2$, then*
$$P(G, L, v, i, D) \leq \frac{12}{25};$$
2. *if $\deg_G(v) = 1$, then*
$$P(G, L, v, i, D) \leq \frac{6}{13}.$$

Proof. If $i \notin L(v)$ then $P(G, L, v, i, D)$ returns 0 and we are done. If $D = 0$ and $i \in L(v)$, then $P(G, L, v, i, 0) = \frac{1}{|L(v)|} \leq \frac{1}{3}$ since v have at least one neighbour. In the following, we assume $i \in L(v)$ and $D \geq 1$.

1. $\deg_G(v) = 2$

   If $i \in L(v_1)$, by Proposition 4.3.3 we know that $f_i \geq \frac{1}{13}$.

$$\begin{aligned} P(G, L, v, i, D) &= \frac{(1-f_i)(1-y_i)}{(1-f_i)(1-y_i) + \sum_{j \in L(v)\setminus\{i\}} (1-f_j)(1-y_j)} \\ &\leq \frac{(1-f_i)(1-0)}{(1-f_i)(1-0) + \sum_{j \in L(v)\setminus\{i\}} (1-f_j)\left(1-\frac{1}{2}\right)} \\ &= \frac{2(1-f_i)}{2(1-f_i) + (3 - \sum_{j \in L(v)\setminus\{i\}} f_j)} \\ &= \frac{2(1-f_i)}{4-f_i} \leq \frac{24}{51} < \frac{12}{25}. \end{aligned}$$

Here we used the fact that $\sum_{j\in L(v)} f_j = 1$. On the other hand, if $i \in L(v_2)$ then $y_i \geq \frac{1}{13}$. So

$$\begin{aligned} P(G,L,v,i,D) &= \frac{(1-f_i)(1-y_i)}{(1-f_i)(1-y_i)+\sum_{j\in L(v)\setminus\{i\}}(1-f_j)(1-y_j)} \\ &\leq \frac{(1-f_i)\left(1-\frac{1}{13}\right)}{(1-f_i)\left(1-\frac{1}{13}\right)+\sum_{j\in L(v)\setminus\{i\}}(1-f_j)\left(1-\frac{1}{2}\right)} \\ &= \frac{\frac{12}{13}(1-f_i)}{\frac{12}{13}(1-f_i)+\frac{1}{2}\sum_{j\in L(v)\setminus\{i\}}(1-f_j)} = \frac{24(1-f_i)}{50-11f_i} \leq \frac{12}{25}. \end{aligned}$$

2. $\deg_G(v) = 1$

   Clearly $i \in L(u)$ where u is the only neighbour of v. So $x = P(G_v, L, u, i, D-1) \geq \frac{1}{13}$.

   If $|L(v)| = 4$ then $P(G,L,v,i,D) = \frac{1-x}{3} < \frac{1}{3} < \frac{6}{13}$.

   If $|L(v)| = 3$ then $P(G,L,v,i,D) = \frac{1-x}{2+y} \leq \frac{1-\frac{1}{13}}{2} = \frac{6}{13}$.

□

Proposition 4.3.5. *Let (G,L,v) be a reachable tuple, $i \in [4]$ be a colour. Assume that $\deg_G(v) = 2$. Then one of the following holds:*

1. *the vertex v and its two neighbours form a triangle in G;*
2. *$P(G,L,v,i,D) \leq \frac{13}{27}$ for all integer $D \geq 2$.*

Proof. Without loss of generality we assume $i = 1$. Denote by v_1 and v_2 the two neighbours of v in G. We only need to consider the case when $1 \notin L(v_1)$ and $1 \notin L(v_2)$, i.e. $f_1 = y_1 = 0$, since otherwise by Proposition 4.3.4 we immediately have $P(G,L,v,1,D) \leq \frac{12}{25} < \frac{13}{27}$. In this case v_1 and v_2 each only have up to one neighbour in G_v, which we will denote by u_1 and u_2 respectively. We now continue to discuss in two cases.

1. $\deg_{G_v}(v_2) = 0$

   According to the algorithm $y_j = \frac{1}{|L(v)|}$ for $j \in L(v_2)$ and $y_j = 0$ for $j \notin L(v_2)$.

If $|L(v_2)| = 3$ then $y_1 = 0$ and $y_2 = y_3 = y_4 = \frac{1}{3}$. We have

$$\begin{aligned} P(G, L, v, i, D) &= \frac{1}{1 + \sum_{j \in L(v) \setminus \{1\}} (1 - f_j)(1 - y_j)} \\ &= \frac{1}{1 + \frac{2}{3}\left(3 - \sum_{j \in L(v) \setminus \{1\}} f_j\right)} \\ &= \frac{1}{1 + \frac{4}{3}} = \frac{3}{7} < \frac{13}{27}. \end{aligned}$$

If $|L(v_2)| = 2$ we can assume $2 \notin L(v_2)$, thus $y_1 = y_2 = 0$ and $y_3 = y_4 = \frac{1}{2}$. We have

$$\begin{aligned} P(G, L, v, i, D) &= \frac{1}{1 + \sum_{j \in L(v) \setminus \{i\}} (1 - f_j)(1 - y_j)} \\ &\leq \frac{1}{1 + (1 - f_2)(1 - 0) + (1 - f_3)(1 - \frac{1}{2}) + (1 - f_4)(1 - \frac{1}{2})} \\ &= \frac{1}{1 + \frac{1}{2}(1 - f_2) + \frac{1}{2}\sum_{j \in L(v) \setminus \{1\}} (1 - f_j)} \\ &= \frac{1}{2 + \frac{1}{2}(1 - f_2)} \leq \frac{4}{9} < \frac{13}{27}. \end{aligned}$$

2. $\deg_{G_v}(v_2) = 1$.

   Since v, v_1 and v_2 do not form a triangle, $L_{2,k}(u_2) = L(u_2)$ for all colour k, thus it follows from Proposition 4.3.4 that for every $j \in L_{2,j}(u_2) = L(u_2)$, we have

$$y_j = P(G_v, L_{2,j}, v_2, j, D - 1) \leq \frac{6}{13}.$$

   So if $L(v_2) \subseteq L(u_2)$ we have

$$\begin{aligned} P(G, L, v, 1, D) &= \frac{1}{1 + \sum_{j \in L(v) \setminus \{1\}} (1 - f_j)(1 - y_j)} \\ &\leq \frac{1}{1 + \left(1 - \frac{6}{13}\right) \sum_{j \in L(v) \setminus \{1\}} (1 - f_j)} \\ &= \frac{1}{1 + \frac{7}{13} \cdot 2} = \frac{13}{27}. \end{aligned}$$

   On the other hand, consider $L(v_2) \not\subseteq L(u_2)$. Notice that $1 \notin L(v_2)$ so there should be some other colour, say colour 2, satisfying $2 \in L(v_2) \setminus L(u_2)$. This also forces $\deg_{G_{v,v_2}}(u_2) \leq 1$. Let

$$t_{kj} \triangleq P(G_{v,v_2}, L_{2,k}, u_2, j, D - 2),$$

where $G_{v,v_2} \triangleq (G_v)_{v_2}$, i.e., the graph obtained from G by removing v and v_2 and all edges incident to them. Since $2 \notin L(u_2)$ we have $t_{k2} = 0$ for all k. We need to further distinguish between two cases.

(a) $1 \in L(u_2)$.

Recall that $L_{2,k}(u_2) = L(u_2)$ so for all $k \in L(v)$, $1 \in L_{2,k}(u_2)$. Combining $\deg_{G_{v,v_2}}(u_2) \leq 1$, by Proposition 4.3.3 we obtain that for all $k \in L(v)$, $t_{k1} \geq \frac{1}{6}$. Specifically we have $t_{21} \geq \frac{1}{6}$. Now 1 is the colour in singleton set $[4] \setminus L(v_2)$, so according to Algorithm 4

$$y_2 = \frac{1 - t_{22}}{2 + t_{21}} \leq \frac{6}{13}.$$

As a consequence, we again have $y_j \leq \frac{6}{13}$ for all $j \in L(v_2)$ and the Proposition follows from this bound.

(b) $1 \notin L(u_2)$. In this case u_2 is isolated in G_{v,v_2} with colour list $\{3, 4\}$. So it is clear that $t_{k1} = t_{k2} = 0$ and $t_{k3} = t_{k4} = \frac{1}{2}$ for every k. Further we have $y_2 = \frac{1}{2}$ and $y_3 = y_4 = \frac{1}{4}$. Now

$$\begin{aligned} P(G, L, v, 1, D) &= \frac{1}{1 + \sum_{j \in L(v) \setminus \{1\}} (1 - f_j)(1 - y_j)} \\ &= \frac{1}{1 + (1 - \frac{1}{2})(1 - f_2) + (1 - \frac{1}{4})(1 - f_3) + (1 - \frac{1}{4})(1 - f_4)} \\ &= \frac{1}{1 + \frac{3}{4} \sum_{j \in L(v) \setminus \{1\}} (1 - f_j) - \frac{1}{4}(1 - f_2)} \\ &= \frac{4}{9 + f_2} \quad \leq \frac{4}{9} \quad < \frac{13}{27}. \end{aligned}$$

□

Combining above propositions with the $\deg_G(v) \leq 1$ case, we have the following theorem for bounds on marginal probabilities computed:

Theorem 4.3.6. *Let (G, L, v) be a reachable tuple, $i \in [4]$ be a colour. Then one of the following propositions holds:*

1. *$P(G, L, v, i, D) = \frac{1}{2}$ for all integer $D \geq 2$;*
2. *$P(G, L, v, i, D) \leq \frac{13}{27}$ for all integer $D \geq 2$. Specifically $P(G, L, v, i, D) \leq \frac{6}{13}$ when $\deg_G(v) \leq 1$.*

Furthermore, when $P(G, L, v, i, D) = \frac{1}{2}$ *for some integer* $D \geq 2$ *the local structure of* G *around* v *falls into one of the following three cases (see Figures 4.1, 4.2 and 4.3):*

1. $\deg_G(v) = 0$. *Then* $j, w \notin L(v)$ *for two distinct colours* j, w *other than* i *(Figure 4.1).*

2. $\deg_G(v) = 1$. *Denote by* u *neighbour of* v. *Then* $i \notin L(u)$ *and* $j \notin L(u) \cup L(v)$ *for some colour* $j \neq i$ *(Figure 4.2).*

3. $\deg_G(v) = 2$. *Denote by* u_1 *and* u_2 *two neighbours of* v. *Then* v, u_1, u_2 *form a triangle, and* $i \notin L(u_1) \cup L(u_2)$ *(Figure 4.3).*

Proof. Assume w.l.o.g. $i = 1$, and we will assume $1 \in L(v)$, otherwise the statement is trivial. From Proposition 4.3.5 we know if $P(G, L, v, i, D) = \frac{1}{2}$ then v and its two neighbours v_1, v_2 must form a triangle, and $1 \notin L(v_1) \cup L(v_2)$, as depicted in Figure 4.3. If this is not the case then $P(G, L, v, i, D) \leq \frac{13}{27}$. Now we focus on those $\deg_G(v) \leq 1$ cases.

1. $\deg_G(v) = 0$.

   We know $|L(v)| \geq 2$. If $|L(v)| \geq 3$ then apparently $P(G, L, v, 1, D) \leq \frac{1}{3} < \frac{13}{27}$. If $|L(v)| = 2$ then $P(G, L, v, i, D) = \frac{1}{2}$ and this is just the case depicted in Figure 4.1

2. $\deg_G(v) = 1$.

   By Algorithm 4, if $|L(v)| = 4$ then $P(G, L, v, 1, D)$ will return $\frac{1-x}{3} < \frac{13}{27}$. So we focus on $|L(v)| = 3$. Assume 2 is the colour in singleton set $[4] \setminus L(v)$.

   According to the algorithm, $P(G, L, v, 1, D)$ now returns $\frac{1-x}{2+y}$ where

   $$x = P(G_v, L_{1,1}, u, 1, D-1)$$
   $$y = P(G_v, L_{1,2}, u, 2, D-1)$$

   Notice that $\frac{1-x}{2+y}$ could reach $\frac{1}{2}$ if and only if $x = y = 0$. Otherwise at least one of x and y is bounded by $\frac{1}{6}$ from below, thus $\frac{1-x}{2+y}$ is bounded by $\max\left\{\frac{1-0}{2+1/6}, \frac{1-1/6}{2+0}\right\} = \frac{6}{13}$.

   Moreover, $x = y = 0$ indicates that $1 \notin L_{1,1}(u)$ and $2 \notin L_{1,2}(u)$. Recall $L_{1,1} = L_{1,2} = L$, it immediately follows that $\deg_{G_v}(u) = 0$ and $1 \notin L(u)$ and $2 \notin L(u)$. Together with $2 \notin L(v)$ we have $2 \notin L(u) \cup L(v)$ which completes the proof (This case is depicted in Figure 4.2).

□

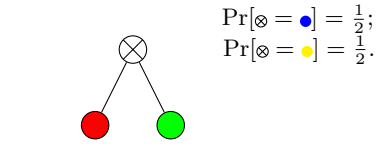


Figure 4.1: Boundary case one

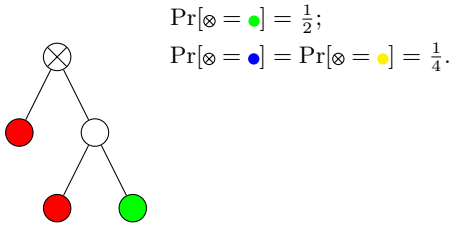


Figure 4.2: Boundary case two

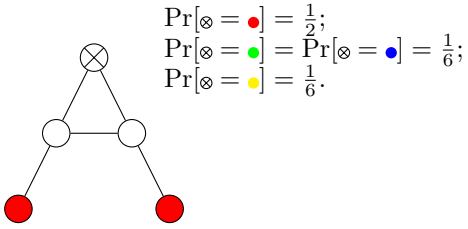


Figure 4.3: Boundary case three

Consider a depth D that is large enough(larger than the size of G). Then clearly $P(G, L, v, i, D)$ should return $\Pr_{G,L}[c(v) = i]$. Therefore we can actually draw the same conclusions for true value $\Pr_{G,L}[c(v) = i]$. Thus we can present the following theorem, which is an immediate corollary of Proposition 4.3.3 and Theorem 4.3.6.

Theorem 4.3.7. *Let (G, L, v) be a reachable tuple and $i \in [4]$ be a colour. Then either $\Pr_{G,L}[c(v) = i]$ is 0 or $1/2$, or $\Pr_{G,L}[c(v) = i] \in \left[\frac{1}{13}, \frac{13}{27}\right]$. Moreover, let $D \geq 2$ be an integer. We also have that*

1. *$P(G, L, v, i, D) = 0$ if and only if $\Pr_{G,L}[c(v) = i] = 0$;*
2. *$P(G, L, v, i, D) = \frac{1}{2}$ if and only if $\Pr_{G,L}[c(v) = i] = \frac{1}{2}$;*
3. *$P(G, L, v, i, D) \in \left[\frac{1}{13}, \frac{13}{27}\right]$ if and only if $\Pr_{G,L}[c(v) = i] \in \left[\frac{1}{13}, \frac{13}{27}\right]$.*

4.4 Verifying the correlation decay property

In this chapter, we discuss the correlation decay property of our recursion. First we present the following theorem, which captures the property of correlation decay.

Theorem 4.4.1. *Suppose $D \geq 3$ and $q = 4$. Let $\lambda = \frac{9996}{10000}$ be a constant. Then for any list-colouring instance $(G = (V, E), L)$ satisfying $|L(v)| \geq \deg_G(v) + 1$ for every $v \in V$, we have*

$$|P(G, L, v, i, D) - \Pr_{G,L}[c(v) = i]| \leq C \cdot \lambda^{D-3},$$

where $C > 0$ is some constant.

We can view the one step recursion $P(G, L, v, i, D)$ as a function F_i where each input of F_i is obtained by calling a depth-$(D-1)$ recursion on some list-colouring instance (G_k, L_k). Therefore F_i has 2 main variants, depending on whether P_1 or P_2 is called.
As we explained in Chapter 2.2.2, the so-called "amortised analysis" is usually a powerful method to prove the correlation decay property. We will apply this method to our proof. We set

$$\varphi(x) = 2\ln x - 2\ln\left(\frac{1}{2} - x\right)$$

whose derivative (potential function) is

$$\Phi(x) = \frac{1}{x(\frac{1}{2} - x)}$$

and let

$$M = \frac{3}{2} - \sqrt{2} = \sup_{0 \leq x \leq \frac{1}{2}} \frac{1}{(1-x)\Phi(x)}.$$

Pick a list-colouring instance $(G = (V, E), L)$ with maximum degree 3, a vertex v in G with neighbour(s) v_1 and v_2 if exist satisfying $|L(v)| \geq \deg_G(v) + 2$ and a colour i. To prove Theorem 4.4.1, the idea is to apply induction on D, which can be formalized by the following lemma.

Lemma 4.4.2. *Let $\lambda = \frac{9996}{10000}$ be a constant and F_i be the function associated with $P(G, L, v, i, D)$. Then one of the following statements holds:*

1. $F_i(\mathbf{x}) = F_i(\mathbf{x}^*) = 0$;
2. $F_i(\mathbf{x}) = F_i(\mathbf{x}^*) = \frac{1}{2}$;

3. $|\varphi(F_i(\mathbf{x})) - \varphi(F_i(\mathbf{x}^*))| \leq \lambda \cdot \max_{j:x_j \in (0,\frac{1}{2})} \left|\varphi(x_j) - \varphi(x_j^*)\right|$,

where $\mathbf{x}$ *are the return values of subroutines called by* $P(G, L, v, i, D)$ *and* $\mathbf{x}^*$ *are true values of those called instances.*

We shall point out here if the first two cases do not occur then $\varphi(F_i(\mathbf{x}))$ and $\varphi(F_i(\mathbf{x}^*))$ are always well-defined. This is a simple corollary of Lemma 4.3.7. Instead of proving this lemma, we will introduce Lemma 4.4.3 which can directly imply Lemma 4.4.2.
To ease the notation we first define the following. For any function $f(x)$ and any d-dimensional vector $\mathbf{x}$ where $d \in \mathbb{N}$, let $f(\mathbf{x})$ denote the vector $(f(x_1), f(x_2), \cdots, f(x_d))$.

Lemma 4.4.3. *Suppose* d *is the arity of* F_i*. Define the contraction rate*

$$\alpha(\mathbf{x}) = \sum_{j=1}^{d} \frac{\Phi(F_i(\mathbf{x}))}{\Phi(x_j)} \left|\frac{\partial F_i(\mathbf{x})}{\partial x_j}\right|.$$

Then for all $\mathbf{x} \in Dom(F_i) \subseteq [0, \frac{1}{2}]^d$*, we have*

$$\alpha(\mathbf{x}) \leq \lambda$$

where $\lambda = \frac{9996}{10000}$.

Before delving into the proof, we first show that how to prove Lemma 4.4.2 by Lemma 4.4.3.

Proof of Lemma 4.4.2. Let $\mathcal{I}$ be the index set of variables of F_i. Let $\mathbf{x}_0 = \left\{x_i \mid i \in \mathcal{I}, x_i \in \left\{0, \frac{1}{2}\right\}\right\}$ and $\mathbf{x}_1 = \left\{x_i \mid i \in \mathcal{I}, x_i \in (0, \frac{1}{2})\right\}$. Let $\mathcal{I}_0$ and $\mathcal{I}_1$ be the corresponding index set of $\mathbf{x}_0$ and $\mathbf{x}_1$. Define $\mathbf{x}_0^*$ and $\mathbf{x}_1^*$ similarly.
Let $\mathbf{u}_1 = \varphi(\mathbf{x}_1)$, $\mathbf{u}_1^* = \varphi(\mathbf{x}_1^*)$, and since φ is strictly increasing we have $\mathbf{x}_1 = \varphi^{-1}(\mathbf{u}_1)$ and $\mathbf{x}_1^* = \varphi^{-1}(\mathbf{u}_1^*)$. Notice $\mathbf{u}_1^*$ is well-defined because we know $x_i \in (0, \frac{1}{2})$ if and only if $x_i^* \in (0, \frac{1}{2})$ by Lemma 4.3.7. In other words, $\mathbf{x}_0$ and $\mathbf{x}_1$ shares the same index set with $\mathbf{x}_0^*$ and $\mathbf{x}_1^*$, respectively.
Introduce

$$g(t) = \varphi(F_i(\mathbf{x}_0, \varphi^{-1}(t\mathbf{u}_1 + (1-t)\mathbf{u}_1^*))$$

so that $\varphi(F_i(\mathbf{x})) - \varphi(F_i(\mathbf{x}^*)) = \varphi(F_i(\mathbf{x}_0, \mathbf{x}_1)) - \varphi(F_i(\mathbf{x}_0^*, \mathbf{x}_1^*)) = g(1) - g(0)$. By *Mean Value Theorem* there exists $\tilde{t} \in (0, 1)$ such that

$$\frac{g(1) - g(0)}{1 - 0} = g'(\tilde{t}).$$

For convenience we denote $\tilde{\mathbf{u}}_1 = \tilde{t}\mathbf{u}_1 + (1-\tilde{t})\mathbf{u}_1^*$ and $\tilde{\mathbf{x}}_1 = \varphi^{-1}(\tilde{\mathbf{u}}_1)$. Clearly each component of $\tilde{\mathbf{x}}_1$ lies between 0 and $\frac{1}{2}$ since φ is a monotone function. Simple derivative calculation yields

$$\begin{aligned}
|\varphi(F_i(\mathbf{x})) - \varphi(F_i(\mathbf{x}^*))| &= \left| \sum_{j \in \mathcal{I}_1} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(\tilde{x}_j)} \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \cdot (u_j - u_j^*) \right| \\
&\leq \sum_{j \in \mathcal{I}_1} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(\tilde{x}_j)} \left| \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \right| \cdot \left| u_j - u_j^* \right| \\
&\leq \left(\sum_{j \in \mathcal{I}_1} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(\tilde{x}_j)} \left| \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \right| \right) \cdot \max_{j \in \mathcal{I}_1} \left| u_j - u_j^* \right|.
\end{aligned}$$

Finally note that if $x_j \in \left\{0, \frac{1}{2}\right\}$ then $\frac{1}{\Phi(x_j)} = 0$, thus,

$$\begin{aligned}
&\sum_{j \in \mathcal{I}_1} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(\tilde{x}_j)} \left| \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \right| \\
=& \sum_{j \in \mathcal{I}_0} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(x_j)} \left| \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \right| + \sum_{j \in \mathcal{I}_1} \frac{\Phi(F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1))}{\Phi(\tilde{x}_j)} \left| \frac{\partial F_i(\mathbf{x}_0, \tilde{\mathbf{x}}_1)}{\partial x_j} \right| \\
\leq& \sup_{\mathbf{x} \in [0, \frac{1}{2}]^d} \alpha(\mathbf{x}) \qquad \leq \lambda.
\end{aligned}$$

This completes the proof. □

We make some remarks. Here F_i is just a general concept representing the function of our algorithm. We use different recursions to compute the marginal probability as the degrees of v and v's neighbours change. As a consequence, the specific form, including arity of F_i has several variants, and depends on actual situations. Moreover, in our analysis we will frequently refine the domain of F_i because in some cases both true value and computed value never exceed a certain bound. Nevertheless, we can always obtain the expression of this contraction rate $\alpha(\mathbf{x})$, and it turns out that we can bound this rate for all variants of F_i.

The rest of this section is dedicated to proving Lemma 4.4.3. Our proof is based on the case analysis of the degree of v. Because of the symmetry between colours, we will only need to prove for $i = 1$. The proofs for other colours are identical.

4.4.1 The case $\deg_G(v) = 1$

Denote by v_1 the only neighbour of v. In this case F_1 has three variants.

$$F_1 = \begin{cases} \frac{1-x}{3} & |L(v)| = 4 \\ \frac{1-x}{2+y} & 1 \in L(v), j \notin L(v) \\ 0 & 1 \notin L(v) \end{cases}$$

where $x = P(G_v, L_{1,1}, v_1, 1, D-1)$ and $y = P(G_v, L_{1,j}, v_1, j, D-1)$. We shall prove Lemma 4.4.3 for the first two variants since the last one is trivial.

1. $|L(v)| = 4$.

   The contraction rate writes as

   $$\alpha(\mathbf{x}) = \frac{\Phi(F_1(x))}{\Phi(x)} \left| \frac{\partial F_1(x)}{\partial x} \right|.$$

   Moreover we have the following upper bound

   $$\frac{\Phi(F_1(x))}{\Phi(x)} \left| \frac{\partial F_1(x)}{\partial x} \right| = \frac{1}{3} \cdot \frac{x\left(\frac{1}{2} - x\right)}{\frac{1-x}{3}\left(\frac{1}{2} - \frac{1-x}{3}\right)} = \frac{3x(1-2x)}{(1-x)(1+2x)} \leq \frac{3x}{1+2x} \leq \frac{3}{4} < 1.$$

2. $1 \in L(v), j \notin L(v)$.

   In this case F_1 is a binary function. The contraction rate writes as

   $$\alpha(x, y) = \frac{\Phi(F_1(x,y))}{\Phi(x)} \left| \frac{\partial F_1(x,y)}{\partial x} \right| + \frac{\Phi(F_1(x,y))}{\Phi(y)} \left| \frac{\partial F_1(x,y)}{\partial y} \right|.$$

   We further consider the following three cases.

   (a) $1 \notin L(v_1)$ and $j \notin L(v_1)$.

   In this case x and y are accurately computed, hence no error occurs in our computation.

   (b) $1 \notin L(v_1)$ and $j \in L(v_1)$.

   Denote by d_1 degree of v_1 in graph G_v. Then $y = \frac{\prod_{k=1}^{d_1}(1-z_{jk})}{\sum_{l \in L(v_1)} \prod_{k=1}^{d_1}(1-z_{lk})} \geq \frac{\left(1-\frac{1}{2}\right)^2}{\left(1-\frac{1}{2}\right)^2 + (1-0)\times 3} = \frac{1}{13}$. This lower bound also holds for y^*.

If $1 \notin L(v_1)$, then $x = x^* = 0$. Let $F_0 = F_1(0, \cdot)$ be the function obtained by fixing $x = 0$ in F_1. The contraction rate of F_0 is

$$\begin{aligned}\alpha(y) &= \frac{\Phi(F_0(y))}{\Phi(y)} \left|\frac{\partial F_0(y)}{\partial y}\right| \\ &= \frac{y\left(\frac{1}{2}-y\right)}{\frac{1}{2+y}\left(\frac{1}{2}-\frac{1}{2+y}\right)} \cdot \frac{1}{(2+y)^2} \\ &= (1-2y) \leq \frac{11}{13}.\end{aligned}$$

(c) $1 \in L(v_1)$.

Similarly we could have $x, x^* \geq \frac{1}{13}$. Then

$$\begin{aligned}\alpha(x,y) = &\frac{\Phi(F_1(x,y))}{\Phi(x)} \left|\frac{\partial F_1(x,y)}{\partial x}\right| + \frac{\Phi(F_1(x,y))}{\Phi(y)} \left|\frac{\partial F_1(x,y)}{\partial y}\right| \\ =&\frac{1}{\frac{1-x}{2+y}\left(\frac{1}{2}-\frac{1-x}{2+y}\right)} \left(\frac{x\left(\frac{1}{2}-x\right)}{2+y} + \frac{(1-x)y\left(\frac{1}{2}-y\right)}{(2+y)^2}\right) \\ =&\frac{x\left(\frac{1}{2}-x\right)(2+y)+y\left(\frac{1}{2}-y\right)(1-x)}{(1-x)\left(x+\frac{y}{2}\right)}.\end{aligned}$$

We show that $\frac{x\left(\frac{1}{2}-x\right)(2+y)+y\left(\frac{1}{2}-y\right)(1-x)}{(1-x)\left(x+\frac{y}{2}\right)} \leq \lambda$ for $\lambda = \frac{9996}{10000}$, which is equivalent to

$$x\left(\frac{1}{2}-x\right)(2+y) + y\left(\frac{1}{2}-y\right)(1-x) \leq \lambda \cdot (1-x)\left(x+\frac{y}{2}\right). \tag{4.4}$$

Inequality (4.4) can be simplified to

$$(1-x)y^2 + \left(x^2 - \frac{\lambda}{2}x - \frac{1-\lambda}{2}\right)y + x^2 - (1-\lambda)x \geq 0. \tag{4.5}$$

Using the fact that $\frac{1}{13} \leq x \leq \frac{1}{2}$, we know the LHS of (4.5) is minimized at $y = \frac{2x^2-\lambda x-1+\lambda}{4x-4}$. Plugging this into (4.5) and it can be simplified to

$$1-2\lambda+\lambda^2+(16-14\lambda-2\lambda^2)x+(-52+36\lambda+\lambda^2)x^2+(32-20\lambda)x^3+4x^4 \leq 0,$$

which holds for $\frac{1}{13} \leq x \leq \frac{1}{2}$.

To summarise the analysis in Chapter 4.4.1, we have

$$\alpha(\mathbf{x}) \leq \lambda = \frac{9996}{10000}.$$

4.4.2 The case $\deg_G(v) = 2$

Denote by v_1, v_2 two neighbours of v. Let $d_i = \deg_{G_v}(v_i)$ and we have $d_1 \geq d_2$. In this case

$$F_i = F_i(\mathbf{x}, \mathbf{y}) = \frac{(1-f_i)(1-y_i)}{\sum_{j\in L(v)}(1-f_j)(1-y_j)}$$

where

$$f_i = \begin{cases} \frac{\prod_{k=1}^{d_1}(1-x_{k,i})}{\sum_{j\in L(v_1)}\prod_{k=1}^{d_1}(1-x_{k,j})} & i \in L(v_1) \\ 0 & i \notin L(v_1). \end{cases}$$

4.4.2.1 A. The case $d_1 = 2$

We first note that for $i, j \in L(v_1)$, f_i/f_j is bounded by constants.

Proposition 4.4.4. *If $d_1 = 1$ or 2 and for every $1 \leq k \leq d_1$, $j \in L(v_1)$, we have $0 \leq x_{k,j} \leq \frac{1}{2}$, then for every $i, j \in L(v_1)$, it holds that $\frac{1}{4} \leq f_i/f_j \leq 4$ and $f_i \geq \frac{1}{13}$.*

Proof. For every $i, j \in L(v_1)$, we have

$$\frac{f_i}{f_j} = \frac{\prod_{k=1}^{d_1}(1-x_{k,i})}{\prod_{k=1}^{d_1}(1-x_{k,j})}.$$

Then the bound for the ratio follows from $d_1 = 1, 2$ and $0 \leq x_{k,j} \leq \frac{1}{2}$ for every $1 \leq k \leq d_1$, $j \in L(v_1)$.
To see the lower bound for f_i, we note that $|L(v)| \leq 4$ and thus $1 = \sum_{j\in L(v)} f_j \leq f_i + 4\sum_{j\in L(v)\setminus\{i\}} f_i \leq 13 f_i$. □

To prove Lemma 4.4.3 it suffices to bound the contraction rate

$$\alpha(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{2}\sum_{j\in L(v_1)} \frac{\Phi(F_1)}{\Phi(x_{ji})}\left|\frac{\partial F_1(\mathbf{x})}{\partial x_{ji}}\right| + \sum_{j=1}^{4}\frac{\Phi(F_1)}{\Phi(y_j)}\left|\frac{\partial F_1(\mathbf{y})}{\partial y_j}\right|.$$

Simple calculation yields

$$\begin{aligned}
&\sum_{i=1}^{2} \sum_{j \in L(v_1)} \frac{\Phi(F_1)}{\Phi(x_{ji})} \left| \frac{\partial F_1}{\partial x_{ji}} \right| \\
&= \sum_{i=1}^{2} \left(\frac{\Phi(F_1)}{\Phi(x_{1i})} \cdot \frac{F_1 f_1}{1 - x_{1i}} \cdot \sum_{k=2}^{4} \frac{F_k}{1 - f_k} + \sum_{j \in L(v_1) \setminus \{1\}} \frac{\Phi(F_1)}{\Phi(x_{ji})} \cdot \frac{F_1 f_j}{1 - x_{ji}} \left| \frac{1}{1 - f_1} - \sum_{\substack{k=1 \\ k \neq j}}^{4} \frac{F_k}{1 - f_k} \right| \right) \\
&\leq \sum_{i=1}^{2} \left(M \cdot \Phi(F_1) F_1 \left(f_1 \sum_{k=2}^{4} \frac{F_k}{1 - f_k} + \sum_{j \in L(v_1) \setminus \{1\}} f_j \left| \frac{1}{1 - f_1} - \sum_{\substack{k=1 \\ k \neq j}}^{4} \frac{F_k}{1 - f_k} \right| \right) \right) \triangleq 2 \cdot P_1(\mathbf{f}, \mathbf{y}),
\end{aligned}$$

$$\sum_{j=1}^{4} \frac{\Phi(F_1)}{\Phi(y_j)} \left| \frac{\partial F_1}{\partial y_j} \right| = \frac{\Phi(F_1)}{\Phi(y_1)} \cdot \frac{F_1(1 - F_1)}{1 - y_1} + \sum_{j=2}^{4} \frac{\Phi(F_1)}{\Phi(y_j)} \cdot \frac{F_1 F_j}{1 - y_j} \triangleq P_2(\mathbf{f}, \mathbf{y}).$$

Now we only need to bound

$$\alpha(\mathbf{x}, \mathbf{y}) = 2P_1(\mathbf{f}, \mathbf{y}) + P_2(\mathbf{f}, \mathbf{y}).$$

Notice that after substituting M for $\frac{1}{(1-x)\Phi(x)}$ we can ignore $\mathbf{x}$ and treat P_1 and P_2 as functions of $\mathbf{f}$ and $\mathbf{y}$, with some constraints on $\mathbf{f}$ as we will see soon.

Discussion on the absolute value. Let $D_j \triangleq \frac{1}{1-f_1} - \sum_{\substack{k=1 \\ k \neq j}}^{4} \frac{F_k}{1-f_k}$ for $j = 2, 3, 4$. We show that at least two of these D_j's are nonnegative. Assume for the contraction that $D_2, D_3 < 0$. Then we obtain

$$\begin{aligned}
\frac{1}{1 - f_1} - \frac{F_1}{1 - f_1} - \frac{F_3}{1 - f_3} - \frac{F_4}{1 - f_4} &< 0 \\
\frac{1}{1 - f_1} - \frac{F_1}{1 - f_1} - \frac{F_2}{1 - f_2} - \frac{F_4}{1 - f_4} &< 0.
\end{aligned}$$

This is equivalent to

$$(1 - f_2)(1 - y_2) + (f_1 - f_3)(1 - y_3) + (f_1 - f_4)(1 - y_4) < 0 \tag{4.6}$$

$$(1 - f_3)(1 - y_3) + (f_1 - f_2)(1 - y_2) + (f_1 - f_4)(1 - y_4) < 0 \tag{4.7}$$

(4.6)+(4.7) gives

$$(1 + f_1 - 2f_3)(1 - y_3) + (1 + f_1 - 2f_2)(1 - y_2) + 2(f_1 - f_4)(1 - y_4) < 0$$

Since $1+f_1-2f_3, 1+f_1-2f_2 > 0$ and $0 < y_2, y_3, y_4 < \frac{1}{2}$,

$$3f_1 + 1 - f_2 - f_3 - 2f_4 < 0.$$

Since $d_1 = 2$ we have $|L(v)| = 4$ so Proposition 4.4.4 holds for all pairs of f_i, f_j, $1 \le i < j \le 4$. Applying $f_1 + f_2 + f_3 + f_4 = 1$, we obtain $4f_1 < f_4$, which is a contradiction.

Therefore, we have either all D_j for $j = 2, 3, 4$ are nonnegative or at most one of it is negative. Assume D_2 is negative, i.e.,

$$(1-f_2)(1-y_2) + (f_1-f_3)(1-y_3) + (f_1-f_4)(1-y_4) < 0.$$

Since $(1-f_2)(1-y_2) \ge 0$, we have either $f_1 < f_3$ or $f_1 < f_4$ or both. W.l.o.g. assume $f_1 < f_3$, now we distinguish between two cases:

- ($f_1 < f_4$) In this case, we can let $y_2 = \frac{1}{2}$ and $y_3 = y_4 = 0$, this gives

  $$1 - f_2 + 2(f_1 - f_3) + 2(f_1 - f_4) < 0.$$

  Using the identity $f_1 + f_2 + f_3 + f_4 = 1$, we obtain

  $$6f_1 + f_2 - 1 < 0.$$

- ($f_1 \ge f_4$) In this case, we can let $y_2 = \frac{1}{2}$, $y_3 = 0$ and $f_4 = f_1$, this gives

  $$1 - f_2 + 2(f_1 - f_3) < 0.$$

  Using $f_3 = 1 - f_1 - f_2 - f_4 \le \frac{7}{8} - f_1 - f_2$, we obtain

  $$4f_1 + f_2 - \frac{3}{4} < 0.$$

Now we can continue our analysis of $\alpha(\mathbf{x}, \mathbf{y})$.

Case 1: All D_j are nonnegative for $j = 2, 3, 4$. Introduce the following function of w and f

$$G_\xi(w, f) = \frac{1-f}{\Phi(1-\frac{w}{1-f})} + 4M\xi \cdot \frac{w}{1-f}$$

where $\xi \in [0, 1]$ is some constant parameter. The following two lemmas will be very useful in our analysis.

Lemma 4.4.5. *$G_0(w,f)$ is concave when $f \in [0,\frac{1}{2}]$ and $\frac{1-f}{2} \le w \le 1-f$, hence for all w_i, f_i satisfying $f_i \in [0,\frac{1}{2}]$ and $\frac{1-f_i}{2} \le w_i \le 1-f_i, i = 1,2,\cdots,n$, we have*

$$\frac{G_0(w_1,f_1)+G_0(w_2,f_2)+\cdots+G_0(w_n,f_n)}{n} \le G_0\left(\frac{w_1+w_2+\cdots+w_n}{n}, \frac{f_1+f_2+\cdots+f_n}{n}\right).$$

Proof. The Hessian of $G_0(w,f)$ is $\begin{bmatrix} -\frac{2}{1-f} & -\frac{2w}{(1-f)^2} \\ -\frac{2w}{(1-f)^2} & -\frac{2w^2}{(1-f)^3} \end{bmatrix}$, which is negative semi-definite when $f \in [0,\frac{1}{2}]$. □

Lemma 4.4.6. *For all $w_1,w_2,w_3 \in [0,\frac{1}{2}]$ and $f_1,f_2,f_3 \in [\frac{1}{13},\frac{1}{2}]$ such that $\frac{1-f_i}{2} \le w_i \le 1-f_i, i = 1,2,3$, we have*

$$\frac{1}{2}\left(G_\xi(w_1,f_1)+G_\xi(w_2,f_2)\right) \le \kappa \cdot G_\xi\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right)$$
$$\frac{1}{3}\left(G_\xi(w_1,f_1)+G_\xi(w_2,f_2)+G_\xi(w_3,f_3)\right) \le \kappa \cdot G_\xi\left(\frac{w_1+w_2+w_3}{3}, \frac{f_1+f_2+f_3}{3}\right)$$

holds for any $\xi \in [0,1]$, where $\kappa = \frac{1038}{1000}$.

Proof. First we shall point out that if the lemma holds for $\xi = 1$, then it should hold for any other $0 \le \xi < 1$.

Suppose the lemma holds for $\xi = 1$. That is

$$\frac{1}{2}\left(G_1(w_1,f_1)+G_1(w_2,f_2)\right) \le \kappa \cdot G_1\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right).$$

Rewrite

$$G_\xi(w,f) = (1-\xi)G_0(w,f) + \xi G_1(w,f).$$

Recall that G_0 is concave, thus

$$\begin{aligned}\frac{1}{2}\left(G_\xi(w_1,f_1)+G_\xi(w_2,f_2)\right) &\le (1-\xi)G_0\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right) + \xi\kappa \cdot G_1\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right) \\ &\le (1-\xi)\kappa \cdot G_0\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right) + \xi\kappa \cdot G_1\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right) \\ &= \kappa \cdot G_\xi\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right).\end{aligned}$$

The same argument works for the 3-variable case. So it remains to prove the $\xi = 1$ case. It can be rigorously proved by Mathematica (the codes are in Chapter 8.2.2) that for all $w_1,w_2,w_3 \in [0,\frac{1}{2}]$ and $f_1,f_2,f_3 \in [\frac{1}{13},\frac{1}{2}]$ such that $\frac{1-f_i}{2} \le w_i \le 1-f_i, i = 1,2,3$, we have

$$\frac{1}{2}(G_1(w_1,f_1)+G_1(w_2,f_2)) \le \kappa_1 \cdot G_1\left(\frac{w_1+w_2}{2}, \frac{f_1+f_2}{2}\right)$$

$$\frac{1}{3}(G_1(w_1,f_1)+2G_1(w_2,f_2)) \le \kappa_2 \cdot G_1\left(\frac{w_1+2w_2}{3}, \frac{f_1+2f_2}{3}\right).$$

Here $\kappa_1 = \frac{10195}{10000}, \kappa_2 = \frac{10181}{10000}$ and $\kappa_1\kappa_2 \le \kappa$. As a consequence,

$$\begin{aligned}
&\frac{1}{3}(G_1(w_1,f_1)+G_1(w_2,f_2)+G_1(w_3,f_3)\\
&= \frac{1}{2}\left(\frac{1}{3}(2G_1(w_1,f_1)+G_1(w_2,f_2)) + \frac{1}{3}(G_1(w_2,f_2)+2G_1(w_3,f_3))\right)\\
&\le \kappa_2 \cdot \frac{1}{2}\left(G_1\left(\frac{2w_1+w_2}{3}, \frac{2f_1+f_2}{3}\right) + G_1\left(\frac{w_2+2w_3}{3} + \frac{f_2+2f_3}{3}\right)\right)\\
&\le \kappa_1\kappa_2 \cdot G_1\left(\frac{1}{2}\left(\frac{2w_1+w_2}{3} + \frac{w_2+2w_3}{3}\right), \frac{1}{2}\left(\frac{2f_1+f_2}{3} + \frac{f_2+2f_3}{3}\right)\right)\\
&\le \kappa \cdot G_1\left(\frac{w_1+w_2+w_3}{3}, \frac{f_1+f_2+f_3}{3}\right).
\end{aligned}$$

□

Recall that $f_j = 0$ for $j \notin L(v_1)$, and

$$\begin{aligned}
&\sum_{j\in L(v_1)\setminus\{1\}} f_j\left(\frac{1}{1-f_1} - \sum_{\substack{k=1\\k\neq j}}^{4}\frac{F_k}{1-f_k}\right)\\
&= \sum_{j=2}^{4} f_j\left(\frac{1}{1-f_1} - \sum_{\substack{k=1\\k\neq j}}^{4}\frac{F_k}{1-f_k}\right) = 1 - \sum_{j=2}^{4} f_j \sum_{\substack{k=1\\k\neq j}}^{4}\frac{F_k}{1-f_k}\\
&= 1 - \sum_{j=1}^{4} f_j \sum_{\substack{k=1\\k\neq j}}^{4}\frac{F_k}{1-f_k} + f_1\sum_{k=2}^{4}\frac{F_k}{1-f_k} = 1 - \sum_{k=1}^{4}\frac{F_k}{1-f_k}\sum_{\substack{j=1\\j\neq k}}^{4} f_j + f_1\sum_{k=2}^{4}\frac{F_k}{1-f_k}\\
&= f_1\sum_{j=2}^{4}\frac{F_j}{1-f_j}.
\end{aligned}$$

So we have

$$\alpha = \Phi(F_1)F_1\left(\frac{1-F_1}{(1-y_1)\Phi(y_1)} + \sum_{j=2}^{4}\frac{F_j}{(1-y_j)\Phi(y_j)} + 4Mf_1\sum_{j=2}^{4}\frac{F_j}{1-f_j}\right).$$

Define symmetric forms of F_k as follows.

$$\hat{F}_k(f_1,f_2,y_1,y_2) = \frac{(1-f_k)(1-y_k)}{(1-f_1)(1-y_1)+3(1-f_2)(1-y_2)}, \quad k=1,2.$$

Then we can define the symmetric form of α

$$\hat{\alpha}(f_1, f_2, y_1, y_2) = \Phi(\hat{F}_1)\hat{F}_1 \left(\frac{1-\hat{F}_1}{(1-y_1)\Phi(y_1)} + \frac{3\hat{F}_2}{(1-y_2)\Phi(y_2)} + 12Mf_1 \cdot \frac{\hat{F}_2}{1-f_2} \right).$$

Lemma 4.4.7. *For all* $\mathbf{f}, \mathbf{y} \in [0, \frac{1}{2}]^4$ *such that* $\frac{1}{13} \leq f_1, f_2, f_3, f_4 \leq \frac{1}{2}$ *and* $f_1+f_2+f_3+f_4 = 1$*, there exists* $\hat{f}_2, \hat{y}_2 \in [0, \frac{1}{2}]$ *such that* $f_1 + 3\hat{f}_2 = 1$ *and*

$$\alpha(\mathbf{f}, \mathbf{y}) \leq \kappa \cdot \hat{\alpha}(f_1, \hat{f}_2, y_1, \hat{y}_2)$$

where $\kappa = \frac{1038}{1000}$.

Proof. Let $w_k = (1-f_k)(1-y_k), k = 2, 3, 4$, $A(\mathbf{f}, \mathbf{y}) = \sum_{j=1}^{4}(1-f_j)(1-y_j)$ be the denominator of F_k, and $\hat{A}(f_1, f_2, y_1, y_2) = (1-f_1)(1-y_1) + 3(1-f_2)(1-y_2)$ be the denominator of $\hat{F}_k$. Then

$$\begin{aligned}
\alpha &= \Phi(F_1)F_1 \left(\frac{1-F_1}{(1-y_1)\Phi(y_1)} + \frac{1}{A}\sum_{j=2}^{4} \frac{1-f_j}{\Phi(1-\frac{w_j}{1-f_j})} + 4Mf_1 \cdot \frac{w_j}{1-f_j} \right) \\
&= \Phi(F_1)F_1 \left(\frac{1-F_1}{(1-y_1)\Phi(y_1)} + \frac{1}{A}\sum_{j=2}^{4} G_{f_1}(w_j, f_j) \right).
\end{aligned}$$

Take $\hat{w}_2$ and $\hat{f}_2$ such that $3\hat{w}_2 = w_1 + w_2 + w_3$, $3\hat{f}_2 = f_1 + f_2 + f_3$, and take $\hat{y}_2 = 1 - \frac{\hat{w}_2}{1-\hat{f}_2}$. Therefore $f_1 + 3\hat{f}_2 = f_1 + f_2 + f_3 + f_4 = 1$ and

$$A(\mathbf{f}, \mathbf{y}) = \hat{A}(f_1, \hat{f}_2, y_1, \hat{y}_2)$$

$$F_1(\mathbf{f}, \mathbf{y}) = \hat{F}_1(f_1, \hat{f}_2, y_1, \hat{y}_2).$$

Furthermore, w_j and y_j satisfy the condition of Lemma 4.4.6 hence

$$\begin{aligned}
\alpha &\leq \Phi(F_1)F_1 \left(\frac{1-F_1}{(1-y_1)\Phi(y_1)} + 3\kappa \cdot \frac{G_{f_1}(\hat{w}_2, \hat{f}_2)}{A} \right) \\
&= \Phi(\hat{F}_1)\hat{F}_1 \left(\frac{1-\hat{F}_1}{(1-y_1)\Phi(y_1)} + 3\kappa \cdot \left(\frac{\hat{F}_2}{(1-\hat{y}_2)\Phi(\hat{y}_2)} + 4Mf_1 \cdot \frac{\hat{F}_2}{1-\hat{f}_2} \right) \right) \\
&\leq \kappa \cdot \Phi(\hat{F}_1)\hat{F}_1 \left(\frac{1-\hat{F}_1}{(1-y_1)\Phi(y_1)} + \frac{3\hat{F}_2}{(1-\hat{y}_2)\Phi(\hat{y}_2)} + 12Mf_1 \cdot \frac{\hat{F}_2}{1-\hat{f}_2} \right) \\
&= \kappa \cdot \hat{\alpha}(f_1, \hat{f}_2, y_1, \hat{y}_2).
\end{aligned}$$

□

Lemma 4.4.8. *For all $f_1, f_2, y_1, y_2 \in [0, \frac{1}{2}]$ such that $\frac{1}{13} \le f_1 \le \frac{1}{2}$ and $f_1 + 3f_2 = 1$, we have*

$$\hat{\alpha}(f_1, f_2, y_1, y_2) \le \frac{963}{1000}.$$

Proof. The lemma can be rigorously proved by Mathematica. The codes are in Chapter 8.2.3. □

Theorem 4.3.7 and Proposition 4.4.4 provide the condition for Lemma 4.4.7. Combining with Lemma 4.4.8, it gives

$$\alpha(\mathbf{f}, \mathbf{y}) \le \kappa \cdot \hat{\alpha}(f_1, \hat{f}_2, y_1, \hat{y}_2) \le \frac{1038}{1000} \cdot \frac{963}{1000} < \frac{9996}{10000}.$$

Case 2: D_j is negative for some j. Without loss of generality, we can assume $j = 2$, i.e., $\frac{1}{1-f_1} - \sum_{k=1,3,4} \frac{F_k}{1-f_k} < 0$. Therefore,

$$\sum_{j=2}^{4} f_j \left| \frac{1}{1-f_1} - \sum_{\substack{k=1 \\ k\neq j}}^{4} \frac{F_k}{1-f_k} \right| = f_1 \sum_{j=2}^{4} \frac{F_j}{1-f_j} - 2f_2 \left(\frac{1}{1-f_1} - \sum_{\substack{k=1 \\ k\neq 2}}^{4} \frac{F_k}{1-f_k} \right).$$

So we have

$$\alpha = \Phi(F_1)F_1 \left(\frac{1-F_1}{(1-y_1)\Phi(y_1)} + \sum_{j=2}^{4} \left(\frac{F_j}{(1-y_j)\Phi(y_j)} + \frac{4Mf_1F_j}{1-f_j} \right) - 4Mf_2 \left(\frac{1}{1-f_1} - \sum_{\substack{k=1 \\ k\neq 2}}^{4} \frac{F_k}{1-f_k} \right) \right)$$

which is a function of $\mathbf{f}, \mathbf{y} \in [0,1]^4$ where $f_1 + f_2 + f_3 + f_4 = 1$.
Similarly, by exploiting the symmetry of f_3 and f_4, we define the symmetric form of F_1.

$$\hat{F}_1(f_1, f_2, f_3, y_1, y_2, y_3) = \frac{(1-f_1)(1-y_1)}{\hat{A}}$$

where

$$\hat{A} = (1-f_1)(1-y_1) + (1-f_2)(1-y_2) + 2(1-f_3)(1-y_3).$$

Then we can define the symmetric form of α

$$\hat{\alpha} = \frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}} \left(\hat{A}(1-\hat{F}_1)P_1 + P_2 + P_3 \right)$$

where

$$P_1 = \frac{1}{(1-y_1)\Phi(y_1)} - \frac{4Mf_2}{1-f_1},$$
$$P_2 = \frac{1-f_2}{\Phi(y_2)} + 4Mf_1(1-y_2),$$
$$P_3 = \frac{2(1-f_3)}{\Phi(y_3)} + 8M(f_1+f_2)(1-y_3).$$

So $\hat{\alpha}$ is a function of $\mathbf{f}, \mathbf{y} \in [0, \frac{1}{2}]^3$.

Lemma 4.4.9. *For all $\mathbf{f}, \mathbf{y} \in [0, \frac{1}{2}]^4$ such that $\frac{1}{13} \le f_3, f_4 \le \frac{1}{2}$ and $f_1+f_2+f_3+f_4 = 1$, there exists $\hat{f}_3, \hat{y}_3 \in [0, \frac{1}{2}]$ such that $f_1+f_2+2\hat{f}_3 = 1$ and*

$$\alpha(\mathbf{f}, \mathbf{y}) \le \kappa \cdot \hat{\alpha}(f_1, f_2, \hat{f}_3, y_1, y_2, \hat{y}_3)$$

where $\kappa = \frac{1038}{1000}$.

Proof. Let $w_j = (1-f_j)(1-y_j)$ for $j = 3, 4$, and denote $A = A(w_1, w_2, w_3, w_4) = \sum_{j=1}^4 w_j$ be the denominator of F_k. Then

$$\begin{aligned}
\alpha &= \frac{\Phi(F_1)F_1}{A}\left(A(1-F_1)P_1 + P_2 + \sum_{j=3}^4 \frac{1-f_j}{\Phi(y_j)} + 4M(f_1+f_2)(1-y_j)\right) \\
&= \frac{\Phi(F_1)F_1}{A}\left(A(1-F_1)P_1 + P_2 + \sum_{j=3}^4 \frac{1-f_j}{\Phi(1-\frac{w_j}{1-f_j})} + 4M(f_1+f_2)\cdot\frac{w_j}{1-f_j}\right) \\
&= \frac{\Phi(F_1)F_1}{A}\left(A(1-F_1)P_1 + P_2 + \sum_{j=3}^4 G_{f_1+f_2}(w_j, f_j)\right).
\end{aligned}$$

Take $\hat{w}_3$ and $\hat{f}_3$ such that $2\hat{w}_3 = w_3 + w_4, 2\hat{f}_3 = f_3 + f_4$, and take $\hat{y}_3 = 1 - \frac{\hat{w}_3}{1-\hat{f}_3}$. Then we have $f_1+f_2+2\hat{f}_3 = f_1+f_2+f_3+f_4 = 1$. Let $\hat{A}(w_1, w_2, w_3) = w_1+w_2+2w_3$. It is clear that $A(w_1, w_2, w_3, w_4) = \hat{A}(w_1, w_2, \hat{w}_3)$. Since $f_1+f_2 \in [0,1]$ by Lemma 4.4.6 we have

$$\begin{aligned}
\alpha(\mathbf{f}, \mathbf{y}) &\le \frac{\Phi(F_1)F_1}{\hat{A}}\left(\hat{A}(1-\hat{F}_1)P_1 + P_2 + 2G_{f_1+f_2}(\hat{w}_3, \hat{f}_3)\right) \\
&= \frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}}\left(\hat{A}(1-\hat{F}_1)P_1 + P_2 + 2\kappa\cdot\left(\frac{1-\hat{f}_3}{\Phi(\hat{y}_3)} + 4M(f_1+f_2)(1-\hat{y}_3)\right)\right) \\
&\le \kappa\cdot\frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}}\left(\hat{A}(1-\hat{F}_1)P_1 + P_2 + \frac{2(1-\hat{f}_3)}{\Phi(\hat{y}_3)} + 8M(f_1+f_2)(1-\hat{y}_3)\right) \\
&= \kappa\cdot\hat{\alpha}(f_1, f_2, \hat{f}_3, y_1, y_2, \hat{y}_3).
\end{aligned}$$

□

Lemma 4.4.10. *For all* $f_1, f_2, f_3, y_1, y_2, y_3 \in [0, \frac{1}{2}]$ *satisfying*

$$\begin{aligned} f_1 + f_2 + 2f_3 &= 1, \\ 6f_1 + f_2 - 1 &< 0, \\ 4f_1 + f_2 - \frac{3}{4} &< 0, \end{aligned}$$

and

$$\frac{1}{13} \le f_1 \le \frac{1}{2}, \quad 0 \le f_2, f_3 \le \frac{1}{2},$$

we have

$$\hat{\alpha}(f_1, f_2, f_3, y_1, y_2, y_3) \le \frac{9163}{10000}.$$

Proof. Recall that

$$\hat{\alpha} = \frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}} \left(\hat{A}(1 - \hat{F}_1)P_1 + P_2 + P_3 \right)$$

where

$$\begin{aligned} P_1 &= \frac{1}{(1-y_1)\Phi(y_1)} - \frac{4Mf_2}{1-f_1}, \\ P_2 &= \frac{1-f_2}{\Phi(y_2)} + 4Mf_1(1-y_2), \\ P_3 &= \frac{2(1-f_3)}{\Phi(y_3)} + 8M(f_1+f_2)(1-y_3) \\ &= \frac{1+f_1+f_2}{\Phi(y_3)} + 8M(f_1+f_2)(1-y_3). \end{aligned}$$

Denote

$$A_1 = \hat{A}(1 - \hat{F}_1) = (1-f_2)(1-y_2) + 2(1-f_3)(1-y_3).$$

So

$$\hat{\alpha} = \frac{2\left(A_1P_1 + P_2 + P_3\right)}{A_1 - (1-y_1)(1-f_1)}.$$

We substitute P_1 for $P_1' = \frac{1}{(1-y_1)\Phi(y_1)} - \frac{4Mf_2}{1-1/13} \ge P_1$ and obtain an upper bound

$$\hat{\alpha} \le \frac{2(A_1P_1' + P_2 + P_3)}{A_1 - (1-y_1)(1-f_1)}.$$

Notice now both numerator and denominator are linear functions of f_1. Therefore it reaches the maximum value when f_1 is at its boundary. The next step is to let f_1 take its boundary values and simplify the formula.

1. $f_1 = \frac{1}{6}(1 - f_2)$.

$$\alpha_1 = \frac{2(A_1P_1' + P_2' + P_3')}{A_1 - (1 - y_1)(1 - \frac{1}{6}(1 - f_2))}$$

where

$$P_2' = \frac{1 - f_2}{\Phi(y_2)} + \frac{2}{3}M(1 - f_2)(1 - y_2),$$
$$P_3' = \frac{7 + 5f_2}{6\Phi(y_3)} + \frac{4}{3}M(5f_2 + 1)(1 - y_3).$$

It can be rigorously proved by Mathematica that $\alpha_1 \leq \frac{9138}{10000}$. The codes are in Chapter 8.2.4.

2. $f_1 = \frac{1}{4}\left(\frac{3}{4} - f_2\right)$.

$$\alpha_2 = \frac{2(A_1P_1' + P_2' + P_3')}{A_1 - (1 - y_1)(1 - \frac{1}{4}(\frac{3}{4} - f_2))}$$

where

$$P_2' = \frac{1 - f_2}{\Phi(y_2)} + M\left(\frac{3}{4} - f_2\right)(1 - y_2),$$
$$P_3' = \frac{19 + 12f_2}{16\Phi(y_3)} + 6M\left(f_2 + \frac{1}{4}\right)(1 - y_3).$$

It can be rigorously proved by Mathematica that $\alpha_1 \leq \frac{9163}{10000}$. The codes are in Chapter 8.2.5.

3. $f_1 = \frac{1}{13}$.

$$\alpha_3 = \frac{2(A_1P_1' + P_2' + P_3')}{A_1 - (1 - y_1)(1 - \frac{1}{13})}$$

where

$$P_2' = \frac{1 - f_2}{\Phi(y_2)} + \frac{4}{13}M(1 - y_2),$$
$$P_3' = \frac{14 + f_2}{13\Phi(y_3)} + 8M\left(f_2 + \frac{1}{13}\right)(1 - y_3).$$

It can be rigorously proved by Mathematica that $\alpha_3 \leq \frac{9102}{10000}$. The codes are in Chapter 8.2.6.

To conclude we have $\hat{\alpha} \leq \max\left\{\frac{9138}{10000}, \frac{9163}{10000}, \frac{9102}{10000}\right\} = \frac{9163}{10000}$. □

The discussion of absolute values provides the condition for Lemma 4.4.9, and combining Lemma 4.4.10 gives

$$\alpha(\mathbf{f}, \mathbf{y}) \leq \kappa \cdot \hat{\alpha}(f_1, f_2, \hat{f}_3, y_1, y_2, \hat{y}_3) \leq \frac{1038}{1000} \cdot \frac{9163}{10000} < \frac{9512}{10000}.$$

To summarise the analysis in Chapter 4.4.2.1, we have

$$\alpha(\mathbf{f}, \mathbf{y}) \leq \max\left\{\frac{9512}{10000}, \frac{9996}{10000}\right\} = \frac{9996}{10000}.$$

4.4.2.2 B. The case $d_1 = 1$

When $d_1 = 1$ we need to bound $\alpha(\mathbf{x}, \mathbf{y}) = P_1(\mathbf{f}, \mathbf{y}) + P_2(\mathbf{f}, \mathbf{y})$. Furthermore, if $1 \in L(v_1)$ then we still have $\frac{1}{13} \leq f_1 \leq \frac{1}{2}, 0 \leq f_2, f_3, f_4 \leq \frac{1}{2}$. In this case, the proof in Chapter 4.4.2.1 can all go through once we obtain the symmetric form of α by the following lemma. This is a modified version of Lemma 4.4.6 that can fit the situation of $d_1 = 1$.

Lemma 4.4.11. *For all $w_1, w_2, w_3 \in [0, \frac{1}{2}]$ and $f_1, f_2, f_3 \in [0, \frac{1}{2}]$ such that $\frac{1-f_i}{2} \leq w_i \leq 1 - f_i, i = 1, 2, 3$, we have*

$$\frac{1}{2}\left(G_\xi(w_1, f_1) + G_\xi(w_2, f_2)\right) \leq \kappa \cdot G_\xi\left(\frac{w_1 + w_2}{2}, \frac{f_1 + f_2}{2}\right)$$
$$\frac{1}{3}\left(G_\xi(w_1, f_1) + G_\xi(w_2, f_2) + G_\xi(w_3, f_3)\right) \leq \kappa \cdot G_\xi\left(\frac{w_1 + w_2 + w_3}{3}, \frac{f_1 + f_2 + f_3}{3}\right)$$

holds for any $\xi \in [0, \frac{1}{4}]$, where $\kappa = \frac{1019}{1000}$.

Proof. The proof is almost the same as Lemma 4.4.6, except that here we only need to prove for $\xi = \frac{1}{4}$. This is also achieved by proving that for all $w_1, w_2, w_3 \in [0, \frac{1}{2}]$ and $f_1, f_2, f_3 \in [0, \frac{1}{2}]$ such that $\frac{1-f_i}{2} \leq w_i \leq 1 - f_i, i = 1, 2, 3$, we have

$$\frac{1}{2}(G_{\frac{1}{4}}(w_1, f_1) + G_{\frac{1}{4}}(w_2, f_2)) \leq \kappa_1 \cdot G_{\frac{1}{4}}\left(\frac{w_1 + w_2}{2}, \frac{f_1 + f_2}{2}\right)$$

$$\frac{1}{3}(G_{\frac{1}{4}}(w_1, f_1) + 2G_{\frac{1}{4}}(w_2, f_2)) \leq \kappa_2 \cdot G_{\frac{1}{4}}\left(\frac{w_1 + 2w_2}{3}, \frac{f_1 + 2f_2}{3}\right)$$

where $\kappa_1 = \frac{1009}{1000}, \kappa_2 = \frac{1009}{1000}$ and $\kappa_1\kappa_2 \leq \kappa$. The Mathematica code to verify the lemma is in Chapter 8.2.7. □

Now it remains to handle the case when $1 \notin L(v_1)$. So in the rest of this section we will assume that $f_1 = 0$.

According to the convention in Algorithm 3, we have either $1 \notin L(v_2)$ or $d_2 = 0$. We will defer the discussion of this $d_2 = 0$ case to the end of this section. If $1 \notin L(v_2)$ we have $f_1 = y_1 = 0$ and this is true for both actual value and computed value. So we fix f_1, y_1 to be zero in our recursion and discuss the contraction rate of this partially fixed function

$$F_1 = \frac{1}{1 + \sum_{j=2}^{4}(1-f_j)(1-y_j)}.$$

The contraction rate should not involve the derivatives of f_1 and y_1, namely

$$\alpha(\mathbf{f}, \mathbf{y}) = P_1(\mathbf{f}, \mathbf{y}) + P_2(\mathbf{f}, \mathbf{y})$$

where

$$P_1(\mathbf{f}, \mathbf{y}) = \Phi(F_1)F_1 \cdot M \cdot \sum_{j \in L(v_1)\setminus\{1\}} f_j \left| 1 - \sum_{\substack{k=1 \\ k\neq j}}^{4} \frac{F_k}{1-f_k} \right|,$$

$$P_2(\mathbf{f}, \mathbf{y}) = \Phi(F_1)F_1 \cdot \sum_{j=2}^{4} \frac{F_j}{(1-y_j)\Phi(y_j)},$$

and

$$F_k = \frac{(1-f_k)(1-y_k)}{1 + \sum_{j=2}^{4}(1-f_j)(1-y_j)}$$

is also partially fixed accordingly.

Discussion on the absolute values. Let $D_j \triangleq 1 - \sum_{\substack{k=1 \\ k\neq j}}^{4} \frac{F_k}{1-f_k}$ for $j = 2, 3, 4$. Recall that

$$\sum_{j=2}^{4} D_j = \sum_{j \in L(v_1)\setminus\{1\}} f_j \left(1 - \sum_{\substack{k=1 \\ k\neq j}}^{4} \frac{F_k}{1-f_k} \right) = f_1 \sum_{j=2}^{4} \frac{F_j}{1-f_j} = 0,$$

so it cannot be the case that all D_j's have the same sign. We will always, without loss of generality, assume D_2 has the opposite sign against others. Then $|D_2| + |D_3| + |D_4|$ is either $2D_2$ or $-2D_2$.

Case 1: D_2 **is negative.** In this case

$$\alpha(\mathbf{f},\mathbf{y}) = \Phi(F_1)F_1\left(M\cdot(-2D_2) + \sum_{j=2}^{4}\frac{F_j}{(1-y_j)\Phi(y_j)}\right).$$

Denote $A \triangleq 1+\sum_{j=2}^{4}(1-f_j)(1-y_j)$ the denominator of F_1.
We first consider the case when $y_j = \frac{1}{2}$ for some $j\in\{2,3,4\}$. By Theorem 4.3.6 we know that all y_j's should be accurately computed given the recursion depth D is at least 3. So we can further discard all derivatives of y_j and obtain

$$\begin{aligned}\alpha(\mathbf{f},\mathbf{y}) &= 2Mf_2\cdot\Phi(F_1)F_1\left(\sum_{\substack{k=1\\k\neq 2}}^{4}\frac{F_k}{1-f_k}-1\right)\\ &= \frac{4Mf_2(3-y_3-y_4-A)}{A-2}.\end{aligned}$$

Notice that α is monotonically increasing on y_2, so we take $y_2=\frac{1}{2}$. After substituting $1-f_2$ for f_3+f_4 we get

$$\alpha(\mathbf{f},\mathbf{y}) \le 4Mf_2\cdot\frac{f_3(\frac{1}{2}-y_3)+f_4(\frac{1}{2}-y_4)}{(\frac{1}{2}-y_3)(1-f_3)+(\frac{1}{2}-y_4)(1-f_4)} \le 2M$$

where the last inequality is due to $f_2,f_3,f_4\le\frac{1}{2}$ and the monotonicity on f_3 and f_4.
On the other aspect, if $y_j\neq\frac{1}{2}$ for all $j\in\{2,3,4\}$, then by Theorem 4.3.6 we have $y_j\le\frac{6}{13}$ for all $j\in\{2,3,4\}$ since d_2 is at most 1. Let $w_j=(1-f_j)(1-y_j)$, by Lemma 4.4.11

$$\begin{aligned}\alpha(\mathbf{f},\mathbf{y}) &= \Phi(F_1)F_1\left(2Mf_2\left(\sum_{\substack{k=1\\k\neq 2}}^{4}\frac{F_k}{1-f_k}-1\right)+\sum_{j=2}^{4}\frac{F_j}{(1-y_j)\Phi(y_j)}\right)\\ &= \frac{\Phi(F_1)F_1}{A}\left(\frac{1-f_2}{\Phi(y_2)}+2Mf_2(1-A)+\sum_{j=3}^{4}\frac{1-f_j}{\Phi(y_j)}+2Mf_2(1-y_j)\right)\\ &= \frac{\Phi(F_1)F_1}{A}\left(\frac{1-f_2}{\Phi(y_2)}+2Mf_2(1-A)+\sum_{j=3}^{4}\frac{1-f_j}{\Phi(1-\frac{w_j}{1-f_j})}+2Mf_2\frac{w_j}{1-f_j}\right)\\ &= \frac{\Phi(F_1)F_1}{A}\left(\frac{1-f_2}{\Phi(y_2)}+2Mf_2(1-A)+\sum_{j=3}^{4}G_{\frac{f_2}{2}}(w_j,f_j)\right)\\ &\le \kappa\cdot\frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}}\left(\frac{1-f_2}{\Phi(y_2)}+2Mf_2(1-\hat{A})+2G_{\frac{f_2}{2}}(\hat{w}_3,\hat{f}_3)\right)\end{aligned}$$

where $\hat{w}_3 = \frac{w_3+w_4}{2}$, $\hat{f}_3 = \frac{f_3+f_4}{2}$, $\hat{F}_1 = \frac{1}{1+w_2+2\hat{w}_3}$ and $\hat{A} = 1 + w_2 + 2\hat{w}_3$. If we take $\hat{y}_3 = 1 - \frac{\hat{w}_3}{1-\hat{f}_3}$ then we can get the symmetric form of α:

$$\hat{\alpha}(\mathbf{f},\mathbf{y}) = \frac{\Phi(\hat{F}_1)\hat{F}_1}{\hat{A}}\left(\frac{1-f_2}{\Phi(y_2)} + 2Mf_2(1-\hat{A}) + \frac{2(1-\hat{f}_3)}{\Phi(\hat{y}_3)} + 4Mf_2(1-\hat{y}_3)\right).$$

Lemma 4.4.12. *For all $f_2, f_3, y_2, y_3 \in [0, \frac{1}{2}]$ satisfying*

$$\begin{aligned} f_2 + 2f_3 &= 1, \\ \frac{1}{13} \le f_2 &\le \frac{1}{2}, \\ 0 \le y_2, y_3 &\le \frac{6}{13}, \end{aligned}$$

we have

$$\hat{\alpha}(f_2, f_3, y_2, y_3) \le \frac{9231}{10000}.$$

Proof. The lemma can be verified by Mathematica. The codes are in Chapter 8.2.8. □

In conclusion we have

$$\alpha(\mathbf{f},\mathbf{y}) \le \max\left\{2M, \kappa\cdot\hat{\alpha}(f_2,\hat{f}_3,y_2,\hat{y}_3)\right\} \le \max\left\{2M, \frac{1018}{1000}\cdot\frac{9231}{10000}\right\} < \frac{94}{100}.$$

Case 2: D_2 is positive In this case

$$\alpha(\mathbf{f},\mathbf{y}) = \Phi(F_1)F_1\left(M\cdot 2D_2 + \sum_{j=2}^{4}\frac{F_j}{(1-y_j)\Phi(y_j)}\right).$$

As we did in Case 1, we first consider when $y_j = \frac{1}{2}$ for some $j \in \{2,3,4\}$. We similarly obtain

$$\begin{aligned} \alpha(\mathbf{f},\mathbf{y}) &= 2Mf_2\cdot\Phi(F_1)F_1\left(1 - \sum_{\substack{k=1\\k\neq 2}}^{4}\frac{F_k}{1-f_k}\right) \\ &= \frac{4Mf_2(A-3+y_3+y_4)}{A-2}. \end{aligned}$$

Notice that $\alpha(\mathbf{f},\mathbf{y})$ is monotonically increasing on y_3 and y_4, so we take $y_3 = y_4 = \frac{1}{2}$ which yields

$$\alpha(\mathbf{f},\mathbf{y}) \le 4Mf_2 \le 2M.$$

Now we once more assume $y_j \le \frac{6}{13}$ for all $j \in \{2,3,4\}$. Recall $\lambda = \frac{9996}{10000}$, we now prove that

$$\alpha(\mathbf{f},\mathbf{y}) = \frac{\sum_{j=2}^{4}(1-f_j)y_j(\frac{1}{2}-y_j) + 2Mf_2(A-3+y_3+y_4)}{\frac{A}{2}-1} < \lambda.$$

Since the denominator of $\alpha(\mathbf{f},\mathbf{y})$ is positive, $\alpha(\mathbf{f},\mathbf{y}) < \lambda$ is equivalent to

$$G \triangleq \sum_{j=2}^{4}(1-f_j)y_j(\frac{1}{2}-y_j) + 2Mf_2\left(A-3+y_3+y_4\right) - \lambda\left(\frac{1}{2}A-1\right) < 0.$$

Note that G is quadratic on y_3, we can write it as

$$\begin{aligned} G &= -(1-f_3)y_3^2 + \left(2Mf_2 + \frac{1}{2}(1-f_3) + \frac{1}{2}\lambda(1-f_3) - 2Mf_2(1-f_3)\right)y_3 + C \\ &= (1-f_3)\left(-y_3^2 + \left(\frac{2Mf_2}{1-f_3} + \frac{1+\lambda}{2} - 2Mf_2\right)y_3\right) + C, \end{aligned}$$

where C is a polynomial containing no y_3.
Therefore, G is increasing in $[-\infty, x_0]$ where $x_0 = \frac{Mf_2}{1-f_3} + \frac{1+\lambda}{4} - Mf_2 \ge \frac{1+\lambda}{4} \ge \frac{6}{13}$. Since y_3 and y_4 are symmetric, the same argument holds for y_4.
We only need to prove that $G' \triangleq G|_{y_3=y_4=\frac{6}{13}} < 0$. Applying $f_2+f_3+f_4=1$, a direct calculation yields

$$\begin{aligned} G' = &\frac{2}{13}Mf_2^2(-6+13y_2) + \frac{1}{338}(6-91\lambda+169y_2+169\lambda y_2-338y_2^2) \\ &+ \frac{1}{338}f_2\left(-13\lambda(-6+13y_2) + (-6+13y_2)(-1-52M+26y_2)\right). \end{aligned}$$

Since $y_2 \le \frac{6}{13}$, G' is increasing in $[-\infty, x_1]$ where $x_1 = \frac{1+52M+13\lambda-26y_2}{104M} > \frac{1}{2}$. Therefore, we only need to prove that

$$G'' \triangleq G'|_{f_2=\frac{1}{2}} = \frac{9}{338} + \frac{3M}{13} - \frac{2\lambda}{13} + \left(\frac{1}{4} - \frac{M}{2} + \frac{\lambda}{4}\right)y_2 - \frac{y_2^2}{2} < 0,$$

which holds for $y_2 \in \left[0, \frac{6}{13}\right]$.
In conclusion we have

$$\alpha(\mathbf{f},\mathbf{y}) \le \max\{2M, \lambda\} = \lambda.$$

The case $d_2 = 0$. At last we come to the discussion for $d_2 = 0$. In this case y_1 is not necessarily 0, but all y_j's are accurately computed. Redefine

$$F_1 = \frac{1 - y_1}{1 - y_1 + \sum_{j=2}^{4}(1 - f_j)(1 - y_j)},$$
$$A = 1 - y_1 + \sum_{j=2}^{4}(1 - f_j)(1 - y_j).$$

As we did before, we shall discard the derivatives of y_j's and assume D_2 has the opposite sign against others. Now

$$\begin{aligned}\alpha(\mathbf{f}, \mathbf{y}) &= 2Mf_2 \cdot \Phi(F_1)F_1 \left|1 - \sum_{\substack{k=1 \\ k\neq 2}}^{4} \frac{F_k}{1 - f_k}\right| \\ &= \frac{4Mf_2\,|A - (1 - y_1) - (1 - y_3) - (1 - y_4)|}{A - 2(1 - y_1)} \\ &= \frac{4Mf_2\left|\sum_{j=2}^{4}(1 - f_j)(1 - y_j) - 2 + y_3 + y_4\right|}{\sum_{j=2}^{4}(1 - f_j)(1 - y_j) - (1 - y_1)}\end{aligned}$$

is monotonically decreasing on y_1. So we can take $y_1 = 0$ and this is reduced to a situation we have discussed before.

So far we have exhausted all possible cases when $\deg_G(v) = 2$. To summarise the analysis in Chapter 4.4.2.2, we have

$$\alpha(\mathbf{f}, \mathbf{y}) \leq \max\left\{\frac{94}{100}, \lambda\right\} = \lambda.$$

4.4.3 Proof of the correlation decay property

Putting together the conclusions of Chapter 4.4.1 and Chapter 4.4.2, we complete the proof of Lemma 4.4.3.

Thus we can prove Theorem 4.4.1 now.

Proof of Theorem 4.4.1. Let $\lambda = \frac{9996}{10000}$ be constant.

We first claim that if a vertex v satisfies $\deg_G(v) \leq 2$ and $|L(v)| \geq \deg_G(v) + 2$, then one of the following statements holds:

- $P(G, L, v, i, D) = \Pr_{G,L}[c(v) = i]$;

- $|\varphi(P(G,L,v,i,D)) - \varphi(\Pr_{G,L}[c(v)=i])| \leq C_1 \cdot \lambda^{D-2}$, where $\varphi(x) = 2\ln x - 2\ln\left(\frac{1}{2}-x\right)$ and $C_1 > 0$ is a constant.

Given the claim, we have for some constant $C_2 > 0$, it holds that

$$|P(G,L,v,i,D) - \Pr_{G,L}[c(v)=i]| = \frac{1}{\Phi(\tilde{x})} \cdot |\varphi(P(G,L,v,i,D)) - \varphi(\Pr_{G,L}[c(v)=i])| \leq C_2 \cdot \lambda^D,$$

where $\Phi(x) \triangleq \varphi'(x) = \frac{1}{x\left(\frac{1}{2}-x\right)}$ and $\tilde{x}$ is some real between $\varphi(P(G,L,v,i,D))$ and $\varphi(\Pr_{G,L}[c(v)=i])$.

Now assume $(G=(V,E),L)$ satisfies $|L(v)| \geq \deg_G(v)+1$ for every $v \in V$. Let $v \in V$ be an arbitrary vertex and consider the computation tree of $P(G,L,v,i,D)$. According to the construction in Chapter 4.2.1, all the sub-instances $P(G',L',v',i',D')$ called by the procedure satisfy $|L(v)| \geq \deg_{G'}(v')+2$ and $\deg_{G'}(v') \leq 2$, i.e., the condition specified in the above claim. Further note that in all cases, the 1-norm of the gradients of our recursions

$$\begin{aligned}
F(x,y,z) &= \frac{1-x}{3-x-y} && \text{if } \deg_G(v)=1 \text{ and } |L(v)|=2;\\
F(x,y) &= \frac{1-x}{2+y} && \text{if } \deg_G(v)=1 \text{ and } |L(v)|=3;\\
F(x) &= \frac{1-x}{3} && \text{if } \deg_G(v)=1 \text{ and } |L(v)|=4;\\
F(\mathbf{f},\mathbf{y}) &= \frac{(1-f_i)(1-y_i)}{\sum_{j\in L(v)}(1-f_j)(1-y_j)} && \text{if } \deg_G(v)=2;\\
F(\mathbf{x},\mathbf{y},\mathbf{z}) &= \frac{(1-x_i)(1-y_i)(1-z_i)}{\sum_{j\in L(v)}(1-x_j)(1-y_j)(1-z_j)} && \text{if } \deg_G(v)=3,
\end{aligned}$$

are bounded above by some constants for parameters in the range $[0,\frac{1}{2}]$. Therefore it follows from the mean value theorem and the claim that

$$|P(G,L,v,i,D) - \Pr_{G,L}[c(v)=i]| \leq C \cdot \lambda^D.$$

for some constant $C > 0$.

It remains to prove the claim. We apply induction on D. The base case is that $D=2$. It follows from Theorem 4.3.6 and Lemma 4.3.7 that if $\Pr_{G,L}[c(v)=i]$ is 0 or $\frac{1}{2}$, then the algorithm return the correct value, i.e., $P(G,L,v,i,D) = \Pr_{G,L}[c(v)=i]$. Otherwise, the function $\varphi(\cdot)$ is bounded from above and thus the claim holds. For $D>2$, the claim follows from the induction hypothesis and Lemma 4.4.2. □

4.5 Proof of Theorem 4.1.1

In this Chapter, we will complete the proof of Lemma 4.1.3 and Theorem 4.1.1. We start by first analysing the running time of Algorithm 1.

Let $G = (V, E)$ be a graph with $|V| = n$, L be its colour lists, $v \in V$ be a vertex, $i \in \{1, 2, 3, 4\}$ be a colour and D be a nonnegative integer. Denote by $\tau(G, L, v, i, D)$ the running time of the procedure $P(G, L, v, i, D)$. Then we have:

Lemma 4.5.1. *If* $\deg_G(v) \leq 2$*, then* $\tau(G, L, v, i, D) = O\left(n^3 12^D\right)$.

Proof. We apply induction on n to show that for some constant $C \geq 0$, $\tau(G, L, v, i, D) \leq C \cdot n^3 12^D$. The base case is that $n = 1$, then the algorithm terminates in constant time. For general n, we need to analyse cases $\deg_G(v) = 1, 2$ respectively.

Case $\deg_G(v) = 1$**:** Algorithm 4 contains two subcases. We use an adjacency matrix to represent a graph. Thus we can construct in n^2 time the graph G_v which contains $n - 1$ vertices. We then have the following recursions for the two cases respectively (assuming notations in the description of Algorithm 4):

$$\tau(G, L, v, i, D) \leq \tau(G_v, L_{1,i}, v_1, i, D - 1) + n^2$$
$$\tau(G, L, v, i, D) \leq \tau(G_v, L_{1,i}, v_1, i, D - 1) + \tau(G_v, L_{1,j}, v_1, j, D - 1) + n^2$$

Then the lemma follows from the induction hypothesis.

Case $\deg_G(v) = 2$**:** Algorithm 3 has at most 12 branches, we have (assuming notations in the description of Algorithm 3):

$$\tau(G, L, v, i, D) \leq \sum_{k \in d_1} \sum_{w \in L(v_1)} \tau(G_{v,v_1}, L'_{k,w}, w, D - 1) + \sum_{j \in L(v)} \tau(G_v, L_{2,j}, j, D - 1) + n^2.$$

Then the lemma follows from the induction hypothesis. □

If $\deg_G(v) = 3$, then the algorithm $P(G, L, v, i, D)$ will call $P3(G, L, v, i, D)$ described in Algorithm 2. However, since the maximum degree of G is at most three hence in further recursion call to Algorithm 1, the degree of a vertex decreases by at least one. Therefore Algorithm 2 can be called at most once. Combining Lemma 4.5.1, we have

Lemma 4.5.2. $\tau(G, L, v, i, D) = O\left(n^3 12^D\right)$.

Now we prove Lemma 4.1.3.

Proof of Lemma 4.1.3. First, we need to bound the value $\Pr_{G,L}[c(v) = i]$ on the computation tree. If $\Pr_{G,L}[c(v) = i] = 0$ then it is clear to see $P(G, L, v, i, D) = 0$ thus we are done. Otherwise we have $\Pr_{G,L}[c(v) = i] \geq \frac{1}{13}$ if (G, L, v) is a *reachable* instance. In previous discussion we know that (G, L, v) is on the root of our computation tree if this instance is not *reachable*. In this case

$$\Pr_{G,L}[c(v) = i] = \frac{\prod_{k=1}^{d}(1 - \Pr_{G_v,L_{k,i}}[c(v_k) = i])}{\sum_{j\in L(v)}\prod_{k=1}^{d}(1 - \Pr_{G_v,L_{k,j}}[c(v_k) = j])},$$

where $d = \deg_G(v) \leq 3$ and $|L(v)| \leq 4$. It yields

$$\Pr_{G,L}[c(v) = i] \geq \frac{\left(1 - \frac{1}{2}\right)^3}{\left(1 - \frac{1}{2}\right)^3 + 1 + 1 + 1} = \frac{1}{25}.$$

Combining with the bound of reachable cases it implies $\Pr_{G,L}[c(v) = i] \geq \frac{1}{25}$ for all instances in computation tree.

By Theorem 4.4.1, there exists constants $\lambda = \frac{9996}{10000}$ and $C > 0$ such that for every list-colouring instance (G, L) satisfying conditions in the statement of the theorem, it holds that

$$|P(G, L, v, i, D) - \Pr_{G,L}[c(v) = i]| \leq C \cdot \lambda^{D-3}$$

for all $D \geq 3$.

For any $0 < \varepsilon < 1$, let t be the smallest integer such that $C \cdot \lambda^{t-3} \leq \frac{\varepsilon}{25}$ and let $\hat{p} = P(G, L, v, i, t)$. We can show that Algorithm 1 up to depth t is the algorithm outputs $\hat{p}$ such that

$$(1 - \varepsilon)\hat{p} \leq \Pr_{G,L}[c(v) = i] \leq (1 + \varepsilon)\hat{p}$$

in time $\mathrm{poly}(|V|, \frac{1}{\varepsilon})$.

Theorem 4.4.1 implies

$$\Pr_{G,L}[c(v) = i] - \frac{\varepsilon}{25} \leq \hat{p} \leq \Pr_{G,L}[c(v) = i] + \frac{\varepsilon}{25}$$

and thus by the bound of $\Pr_{G,L}[c(v) = i]$ above it holds that

$$(1 - \varepsilon)\Pr_{G,L}[c(v) = i] \leq \hat{p} \leq (1 + \varepsilon)\Pr_{G,L}[c(v) = i]\ .$$

So

$$(1 - \varepsilon)\hat{p} \leq \frac{1}{1 + \varepsilon}\hat{p} \leq \Pr_{G,L}[c(v) = i] \leq \frac{1}{1 - \varepsilon}\hat{p} \leq (1 + \varepsilon)\hat{p}\,.$$

Next we show that Algorithm 1 up to depth t is a polynomial time algorithm with respect to $|V|$ and $\frac{1}{\varepsilon}$. By Lemma 4.5.2, $\tau(G, L, v, x, t) = O\left(n^3 12^t\right)$. Since t is the smallest integer such that $C \cdot \lambda^{t-3} \leq \frac{\varepsilon}{25}$, we have

$$t - 4 \leq \log_\lambda \frac{\varepsilon}{25C} \leq t - 3\,,$$

which implies $\tau(G, L, v, x, t) = O\left(n^3 12^{\log_\lambda \frac{\varepsilon}{25C}}\right) = O\left(n^3 \left(\frac{25C}{\varepsilon}\right)^{-\log_\lambda 12}\right)$. λ and C are constants, so $\tau(G, L, v, x, t) = \mathrm{poly}\left(|V|, \frac{1}{\varepsilon}\right)$. □

Finally, combining Lemma 4.1.3 and Lemma 4.1.2 completes the proof of Theorem 4.1.1.

Chapter 5

Counting in Uniqueness II : Sampling Potts on Random Regular Graphs

This chapter is based on the joint work with Antonio Blanca , Andreas Galanis , Leslie Ann Goldberg , Daniel Štefankovič and Eric Vigoda [7].

5.1 Introduction

Connecting the uniqueness phase transition with the performance of algorithms is a difficult task that is largely under development. This connection is well-understood on the grid, where it is known that the mixing time of local Markov chains, such as the Glauber dynamics, switches from polynomial to exponential at the corresponding uniqueness threshold, see for example [59, 58, 73, 2, 60, 4]. Also, for certain two-state models, such as the ferromagnetic Ising model and the hard-core model, it has been proved that Glauber dynamics mixes rapidly when the underlying parameter is in uniqueness and that the dynamics mixes slowly otherwise, see [64, 65, 30]. So the same picture is conjectured to hold for the Potts model as well, but in general this remains open. In Chapter 4, we introduced the algorithmic result of proper q-colouring model in the uniqueness regime for $q = 4$. In this chapter, we will discuss general cases of the antiferromagnetic Potts model.

There has been significant progress in the last years in understanding the complexity of sampling from the Gibbs distribution in two-state systems, but for multi-state systems progress has been slower, especially on the algorithmic side. In this chapter, for all

integers $q \geq 3$ and $\Delta \geq 3$, we design approximate sampling algorithms for the q-state antiferromagnetic Potts model on random Δ-regular graphs (regular graphs with n vertices chosen uniformly at random), when the parameter β lies in the uniqueness regime of the regular tree. Our algorithm is not based on a Markov chain approach but proceed by iteratively adding the edges of the graph and performing a resampling step at each stage. As such, this algorithm can produce samples that are within error $1/n^{\delta}$ from the Potts distribution for some fixed constant $\delta > 0$ (which depends on β, q, Δ).

5.1.1 The result

Theorem 5.1.1. *Let $\Delta \geq 3$, $q \geq 3$ and $\beta \in (0,1)$ be in the uniqueness regime of the $(\Delta-1)$-ary tree with $\beta \neq (\Delta - q)/\Delta$. Then, there exists a constant $\delta > 0$ such that, for all sufficiently large n, the following holds with probability $1 - o(1)$ over the choice of a random Δ-regular graph $G = (V, E)$ with n vertices.*
There is a polynomial-time algorithm which, on input the graph G, outputs a random assignment $\sigma : V \to [q]$ whose distribution ν_σ is within total variation distance $O(1/n^{\delta})$ from the Potts distribution μ_G with parameter β, i.e.,

$$\|\nu_\sigma - \mu_G\|_{\mathrm{TV}} = O(1/n^{\delta}).$$

Note that the sampling algorithm in the antiferromagnetic case works throughout uniqueness apart from the point $(\Delta - q)/\Delta$, where uniqueness on the tree is expected to hold but the model is conjectured to be at criticality. In particular, for all $\beta \neq (\Delta - q)/\Delta$ which are in the uniqueness regime of $\mathbb{T}_\Delta$, it can be shown that the decay on the tree is exponentially small in its height. In contrast, even if uniqueness on the tree holds for $\beta = (\Delta - q)/\Delta$, it can be shown that the decay on the tree is polynomial in its height.

Remark 5.1.1. There are certain "bad" Δ-regular graphs where the algorithms will fail to produce samples with the desired accuracy; saying that the algorithms work on random Δ-regular graphs means that the number of these "bad" graphs with n vertices is a vanishing fraction of all Δ-regular graphs with n vertices for large n. Moreover, we can recognise the "good" graphs (where our algorithms will successfully produce samples with the desired accuracy) in polynomial time.

5.1.2 Prior works

Recall that, we define the uniqueness of the Potts model on d-ary trees in Chapter 1.2.2, which is restated below:

Definition 1.2.2. The q-state Potts model with parameter β exhibits *uniqueness* on the infinite d-ary tree $\mathbb{T}_d$ if, for all colours $c \in [q]$, it holds that

$$\limsup_{n\to\infty} \max_{\tau\colon \Lambda_{\mathbb{T}_{d,n}} \to [q]} \left| \Pr_{\mathbb{T}_{d,n}} \left[\sigma(v_{d,n}) = c \mid \sigma(\Lambda_{\mathbb{T}_{d,n}}) = \tau\right] - \frac{1}{q} \right| = 0. \tag{5.1}$$

It exhibits *non-uniqueness* otherwise.

Let $\Delta = d + 1$ be the maximum degree of d-ary trees. For the antiferromagnetic Potts model ($\beta < 1$), the uniqueness threshold on the tree is not yet known in full generality. It is a folklore conjecture that the model has uniqueness iff $q \geq \Delta$ and $\beta \in (0, 1)$, or $q < \Delta$ and $\beta \geq 1 - q/\Delta$. It is known that the model has non-uniqueness when $\beta < 1 - q/\Delta$ [32]. Establishing the uniqueness side of the conjecture is more difficult; we have proved this conjecture for small values of q and Δ in Chapter 3 but generally it is still open. For the q-colourings model ($\beta = 0$), Jonasson [49], building on work of Brightwell and Winkler [10], established that the model has uniqueness iff $q > \Delta$.

The connection between the uniqueness condition and the efficiency of approximate counting problems has been well studied for many spin systems, such as the Ising model and the hard-core model. However, for the Potts model, this study becomes a difficult task and is still open. So in this chapter, we focus on the antiferromagnetic Potts model on random graphs. Understanding spin systems on random graphs [64, 65, 79, 9, 63, 21, 33, 29, 30] gives insights about the complexity of counting and the efficiency of approximate sampling algorithms.

Our sampling algorithm is inspired by Efthymiou's algorithm [29] for sampling q-colourings on $G(n, d/n)$; the algorithm there also proceeds by iteratively adding the edges of the graph and exploits the uniqueness on the tree to show that the sampling error is small. However, for the antiferromagnetic Potts model, the resampling step turns out to be significantly more involved and we need substantial amount of work to ensure that it can be carried out efficiently.

Our sampling algorithm is also based on some properties of random graphs (which will be introduced in Chapter 5.2) and some properties of the Potts model on tree-like structures

(which will be introduced in Chapter 5.4). The whole sampling algorithm and the proof of Theorem 5.1.1 will be presented in Chapter 5.3 and a ReSample subroutine of sampling from the Ising model will be presented in Chapter 5.5.

5.2 Properties of random regular graphs

In this section, we state and prove structural properties of random Δ-regular graphs which ensure that our algorithms for the random-cluster and Potts models have the desired accuracy (cf. Remark 5.1.1).

We will work in the well-known configuration model. Precisely, for $n\Delta$ even, let $\mathcal{G} := \mathcal{G}_{n,\Delta}$ denote the uniform distribution on Δ-regular graphs which is obtained by taking a perfect matching of the set $[n] \times [\Delta]$ and collapsing for each $u \in [n]$ the elements $(u, 1), \dots, (u, \Delta)$ into a single vertex u; the elements of the set $[n] \times [\Delta]$ are called *points*. Technically, the distribution $\mathcal{G}$ is supported on multigraphs but it can be shown that the probability that $G \sim \mathcal{G}$ is simple is asymptotically a positive constant as $n \to \infty$; conditioned on that event, G is uniformly distributed over Δ-regular graphs with n vertices, and therefore any event that holds with probability $1 - o(1)$ in $\mathcal{G}_{n,\Delta}$ also holds with probability $1 - o(1)$ over the uniform distribution on Δ-regular graphs with n vertices.

The following lemma guarantees that short cycles are disjoint in a random Δ-regular graph.

Lemma 5.2.1. *Let $\Delta \geq 3$ be an integer. Then, with probability $1 - o(1)$ over the choice of a uniformly random Δ-regular graph with n vertices, any two distinct cycles of length $\leq \frac{1}{5} \log_{\Delta-1} n$ are disjoint, i.e., they do not share any common vertices or edges.*

Proof. For convenience, let $\ell := \lfloor \frac{1}{5} \log_{\Delta-1} n \rfloor$ and $\mathcal{G} := \mathcal{G}_{n,\Delta}$. Let $G \sim \mathcal{G}$ and $\mathcal{E}$ be the event that G contains two distinct cycles of length $\leq \ell$ which are not disjoint. Let also $\mathcal{F}$ be the event that, for some integer $k \in [1, 2\ell]$, G contains a subgraph with k vertices and $k + 1$ edges such that each vertex has at least two incident multiedges (in the subgraph); we call such a subgraph *bad*. When $\mathcal{E}$ occurs, we obtain that $\mathcal{F}$ also occurs, so it suffices to upper bound the probability of the latter.

To bound $\Pr_{\mathcal{G}}[\mathcal{F}]$, we will use a union over all possible bad subgraphs with k vertices with $k \in [1, 2\ell]$. Consider an arbitrary integer $k \in [1, 2\ell]$. There are $\binom{n}{k} \leq (\frac{en}{k})^k$ ways to choose the vertices in the subgraph, at most $\binom{\Delta}{2}^k (\Delta k)^2$ ways to choose the points that are going

to be paired and then $\frac{(2(k+1))!}{(k+1)!2^{k+1}} \leq 10(2k/\mathrm{e})^{k+1}$ ways to pair the points. The probability of a particular pairing of $2(k+1)$ points occurring is equal to $\frac{1}{(\Delta n-1)\cdots(\Delta n-(2k+1))} \leq \frac{2}{(\Delta n)^{k+1}}$ (using that $k \leq 2\ell = o(n^{1/2})$). We therefore obtain that

$$\Pr_{\mathcal{G}}[\mathcal{F}] \leq 20\sum_{k=1}^{2\ell}\binom{n}{k}\frac{(\Delta k)^2\binom{\Delta}{2}^k(2k/\mathrm{e})^{k+1}}{(\Delta n)^{k+1}} \leq \frac{60\Delta}{n}\sum_{k=1}^{2\ell}k^2(\Delta-1)^k \leq \frac{500\Delta\ell^3}{n}(\Delta-1)^{2\ell} = o(1),$$

as needed. This finishes the proof. □

The following lemma guarantees that certain weighted sums over cycles are small; this bound will be used to show that the aggregate error of our samplers is small.

Lemma 5.2.2. *Let $\Delta \geq 3$. Then, for any constant $W > \Delta - 1$ and any constant $\ell_0 > 0$, there exists a constant $\delta > 0$ such that the following holds with probability $1 - O(1/n^{\delta})$ over the choice of $G \sim \mathcal{G}_{n,\Delta}$. Let C_ℓ denote the number of cycles of length ℓ. Then,*

$$\sum_{\ell \geq \ell_0 \log n}\frac{\ell C_\ell}{W^\ell} \leq 1/(2n^{\delta}).$$

Proof. Let $w := (\Delta-1)/W < 1$ and $\delta := \frac{1}{2}\ell_0\log(1/w) > 0$. We will show that, for all sufficiently large n, it holds that

$$\mathbf{E}_{\mathcal{G}}[X] \leq \frac{1}{(1-w)n^{2\delta}}, \quad \text{where } X := \sum_{\ell \geq \ell_0 \log n}\frac{\ell C_\ell}{W^\ell}. \tag{5.2}$$

Once we show (5.2), the result follows from Markov's inequality.

Fix an arbitrary integer $\ell \in [n]$, we will calculate $\mathbf{E}_{\mathcal{G}}[C_\ell]$. In the configuration model, a cycle with ℓ vertices corresponds to an ordered 2ℓ-tuple of points $(u_1, i_1), (u_1, i_2), \ldots, (u_\ell, i_{2\ell-1}), (u_\ell, i_{2\ell})$ where $u_1, \ldots, u_\ell$ are distinct elements of $[n]$ and $i_1, \ldots, i_{2\ell} \in [\Delta]$ such that (u_j, i_{2j}) is paired to (u_{j+1}, i_{2j+1}) for all $j \in [\ell]$ (with the convention that $u_{\ell+1} = u_1$ and $i_{2l+1} = i_1$). There are $\ell!\binom{n}{\ell}$ ways to choose and order $u_1, \ldots, u_\ell$ and $(\Delta(\Delta-1))^\ell$ ways to choose $i_1, \ldots, i_{2\ell}$ for a total of $\ell!\binom{n}{\ell}(\Delta(\Delta-1))^\ell$ possible tuples; this overcounts the number of tuples corresponding to distinct cycles by a factor of 2ℓ (the number of ways to root and orient the 2ℓ-tuple). Now, the pairing corresponding to a tuple occurs with probability $\frac{1}{(\Delta n-1)(\Delta n-3)\cdots(\Delta n-(2\ell-1))}$. Since

$$2(\Delta n-1)(\Delta n-3)\cdots(\Delta n-(2\ell-1)) \geq \Delta^\ell n(n-1)\cdots(n-\ell+1)$$

for all $\ell \in [1, n]$ and $\Delta \geq 3$, it follows that $\mathbf{E}_{\mathcal{G}}[C_\ell] \leq (\Delta - 1)^\ell / \ell$ and hence

$$\mathbf{E}_{\mathcal{G}}[X] = \sum_{\ell \geq \ell_0 \log n} w^\ell \leq w^{\ell_0 \log n}/(1-w),$$

which proves (5.2) and therefore concludes the proof. □

Our next lemma captures the tree-like structure of random Δ-regular graphs that will be relevant for us. In particular, we give a description of the neighbourhood structure around a path. To do this accurately, we will need a few definitions. Let $G = (V, E)$ be a graph. For a vertex $v \in V$ and integer $h \geq 0$, we denote by $\Gamma_h(G, v)$ the set of vertices at distance $\leq h$ from v.

Definition 5.2.3. Let G be a graph and P be a path in G with vertices $u_1, \ldots, u_\ell$. Let $G \setminus P$ be the graph obtained from G by removing the edges of the path P. Then, for an integer $h \geq 0$, the *h-graph-neighbourhood* of the path P is the subgraph of $G \setminus P$ induced by the vertex set $\bigcup_{i \in [\ell]} \Gamma_h(G \setminus P, u_i)$.
A connected component of the h-graph-neighbourhood will be called *isolated* if it contains *exactly one* of the vertices $u_1, \ldots, u_\ell$.

Lemma 5.2.4. *Let $\Delta \geq 3$. Then, for any constant integer $h \geq 0$ and any $\varepsilon > 0$, there exists a constant $\ell_1 > 0$ such that the following holds for all sufficiently large n.*
With probability $1 - O(1/n^2)$ over the choice of $G \sim \mathcal{G}_{n,\Delta}$, every path P in G with ℓ vertices with $\ell_1 \leq \ell \leq n^{9/10}$ has an h-graph-neighbourhood with at least $(1-\varepsilon)\ell$ isolated tree components.

We will use the following version of the well-known Chernoff/Hoeffding inequality.

Lemma 5.2.5 (see, e.g., [31, Theorem 21.6 & Corollary 21.9])**.** *Suppose that $S_n = X_1 + \cdots + X_n$, where $\{X_i\}_{i \in [n]}$ is a collection of independent random variables such that $0 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = \mu_i$ for $i = 1, \ldots, n$. Let $\mu = \mu_1 + \cdots + \mu_n$. Then, for any $c > 1$,*

$$\Pr[S_n \geq c\mu] \leq \exp\big(-c\mu \log(c/\mathrm{e})\big).$$

Proof of Lemma 5.2.4. Fix an arbitrary integer $h \geq 0$ and constant $\varepsilon > 0$. We will show that the lemma holds with $\ell_1 := 200/\varepsilon$.
In the configuration model, a path P with ℓ vertices corresponds to an ordered $2(\ell-1)$-tuple of points $(u_1, i_1), (u_2, i_2), (u_2, i_3), \ldots, (u_{\ell-1}, i_{2(\ell-2)}), (u_{\ell-1}, i_{2\ell-3}), (u_\ell, i_{2(\ell-1)})$ where

$u_1, \dots, u_\ell$ are distinct elements of $[n]$ and $i_1, \dots, i_{2(\ell-1)} \in [\Delta]$ such that (u_j, i_{2j-1}) is paired to (u_{j+1}, i_{2j}) for all $j \in [\ell - 1]$. Fix any such path P with ℓ vertices $u_1, \dots, u_\ell$ and condition on the event that P appears in G. We will next reveal the h-graph-neighbourhood of P in a breadth-first search manner, as follows.

1. Let $U_0 = \{u_1, \dots, u_\ell\}$. Initialize $U_1, \dots, U_h, U_{h+1}$ to be the empty sets.
2. For $t = 0, 1, \dots, h$:
3. Order the vertices in U_t in lexicographic order.
4. For $k = 1, \dots, \ell\Delta^t$:
5. Pick the k-th vertex in U_t, say u (if $k > |U_t|$, set $u = 0$).
6. For $i = 1, \dots, \Delta$:
7. If $u \neq 0$ and the point (u, i) is not already paired,
8. Pair (u, i) with an available point, say (v, i').
9. If $v \notin U_1 \cup \dots \cup U_t \cup U_{t+1}$, add v to U_{t+1}.

By induction, we have that, for all $t \geq 0$, U_t consists of the set of vertices at distance t from a vertex in $\{u_1, \dots, u_\ell\}$. Note also that $|U_t| \leq \ell\Delta^t$. Now, fix arbitrary $t \in \{0, 1, \dots, h\}$, $k \in [\ell\Delta^t]$ and $i \in [\Delta]$. Let $\mathcal{F}_{t,k,i}$ be the pairings that we have revealed about the graph G just before executing lines 7–9. Similarly, let $S_{t,k,i}$ be the set of vertices we have encountered just before executing lines 7–9 (i.e., the union of $U_1, \dots, U_t$ together with the current set U_{t+1}). Let also $Q_{t,k,i}$ be the event that in line 8 all of the following happen: (i) $u \neq 0$, (ii) the point (u, i) is not paired, and (iii) (u, i) is paired to a point in $S_{t,k,i} \times [\Delta]$. There are at least $\Delta(n - |S_{t,k,i}|)$ available points, of which at most $\Delta|S_{t,k,i}|$ belong to $S_{t,k,i} \times [\Delta]$. Therefore,

$$\Pr[Q_{t,k,i} \mid \mathcal{F}_{t,k,i}] \leq \frac{|S_{t,k,i}|}{n - |S_{t,k,i}|} \leq \frac{2\ell\Delta^{h+2}}{n}, \tag{5.3}$$

where in the last inequality we used that

$$|S_{t,k,i}| \leq |U_0| + \dots + |U_{t+1}| \leq \ell\Delta^{t+2} \leq \ell\Delta^{h+2} \text{ and } n - \ell\Delta^{h+2} \geq n/2$$

for all sufficiently large n (using that $\ell_1 \leq \ell \leq n^{9/10}$). Using (5.3), we obtain that the number of events $\{Q_{t,k,i}\}$ that occur is dominated above by a binomial r.v. $X_\ell \sim \mathrm{Bin}(\ell\Delta^{h+2}, \frac{2\ell\Delta^{h+2}}{n})$. By Lemma 5.2.5, we have that

$$\Pr[X_\ell \geq \varepsilon\ell/2] \leq \mathrm{e}^{-\frac{1}{2}\varepsilon\ell\log(c_\ell/\mathrm{e})}, \text{ where } c_\ell = \frac{\varepsilon n}{4\ell\Delta^{2h+4}}.$$

Since $\ell \leq n^{9/10}$, we have $c_\ell \geq \mathrm{e}n^{1/20}$ for all sufficiently large n and hence

$$\Pr[X_\ell \geq \varepsilon\ell] \leq \mathrm{e}^{-\frac{1}{40}\varepsilon\ell\log n}.$$

It follows that for any path P with ℓ vertices, with probability $\geq 1 - \mathrm{e}^{-\frac{1}{40}\varepsilon\ell\log n}$, at most $\varepsilon\ell/2$ of the events $\{Q_{t,k,i}\}$ occur, i.e., the h-graph-neighbourhood of P contains at least $(1-\varepsilon)\ell$ isolated tree components (every event $Q_{t,k,i}$ that occurs decreases the number of isolated tree components by at most two). Since there are at most $n\Delta^\ell$ paths with ℓ vertices, we obtain by a union bound that the probability that there exists a path whose h-graph-neighbourhood contains less than $(1-\varepsilon)\ell$ isolated tree components is upper bounded by

$$\sum_{\ell=\lceil 200/\varepsilon\rceil}^{n^{9/10}} \mathrm{e}^{\log n+\ell\log\Delta-\frac{1}{40}\varepsilon\ell\log n} = O(1/n^2),$$

where the last bound follows by observing that, for all sufficiently large n, the summands are decreasing functions of ℓ and that for $\ell = \lceil 200/\varepsilon\rceil$ we have $\mathrm{e}^{\log n+\ell\log\Delta-\frac{1}{40}\varepsilon\ell\log n} = O(1/n^3)$.
This concludes the proof of Lemma 5.2.4. □

To conclude this section, we clarify a small point relevant to Remark 5.1.1. We will only utilise Lemma 5.2.4 for paths of logarithmic length (despite the fact that the lemma is stated for convenience for much longer paths) and therefore the property can be checked in polynomial time. Analogously, the sum in Lemma 5.2.2 will only be considered for cycles of logarithmic length and therefore the (restricted) inequality can also be checked in polynomial time.

5.3 Sampling algorithm

5.3.1 Connection between Potts on bichromatic classes and the Ising model

In this section, we describe the connection between Potts on bichromatic classes and the Ising model. Recall that the Ising model is the special case $q = 2$ of the Potts model; to distinguish between the models, we will use π_G to denote the Ising distribution on G with parameter β. Sometimes, we will need to replace the binary set of states $\{1, 2\}$ in the Ising model by other binary sets to facilitate the arguments; we use $\pi_G^{c_1,c_2}$ to denote the Ising distribution with binary set of states $\{c_1, c_2\}$ (we will have that $c_1, c_2 \in [q]$).
For a configuration $\sigma : V \to [q]$, we will denote by $\sigma(U)$ the restriction of σ to the set U.
Our sampling algorithm is based on the following simple observation.

Observation 5.3.1. *Let $q \geq 3$ and $\beta > 0$. Let $G = (V, E)$ be a graph, U be a subset of $V$and c, c' be distinct colours in $[q]$. Then, for any configuration $\eta : U \to \{c, c'\}$, it holds that*

$$\mu_G\big(\sigma(U) = \eta \mid \sigma^{-1}(c, c') = U\big) = \pi^{c,c'}_{G[U]}(\eta),$$

i.e., conditioned on U being the (c, c')-colour-class in the Potts distribution μ_G, the marginal distribution on U is the Ising distribution $\pi_{G[U]}$ (with set of states $\{c, c'\}$).

The following definition will be notationally convenient

Definition 5.3.2. Let G be a graph, u, v be vertices in G and c, c' be distinct colours in $[q]$. We write $\pi^{c,c'}_{G,u,v}$ to denote the Ising distribution on G (with set of states $\{c, c'\}$) conditioned on u taking the state c and v the state c'.

5.3.2 Analysing an Ideal ReSample subroutine

In this section, we give a preliminary description and analysis of the RESAMPLE subroutine in Algorithm SAMPLEANTIPOTTS. In fact, for the moment, we will only study an "idealised" version IDEALRESAMPLE; the subroutine is idealised in the sense that it assumes that certain steps can be carried out efficiently. Later, we will modify the subroutine to obtain the actual RESAMPLE subroutine so that the running time is polynomial with respect to the size of the input graph.

The point of analysing first IDEALRESAMPLE is to give the key ideas behind the underlying resampling step without bothering for the moment to make the subroutine computationally efficient. Moreover, the detour is going to be smaller than it might appear since the analysis of the actual RESAMPLE subroutine will follow from the analysis of IDEALRESAMPLE.

The IDEALRESAMPLE subroutine takes as inputs a graph G, two vertices u and v of G and a configuration σ on G such that $\sigma(u) = \sigma(v)$; it outputs a configuration σ' on G. The

details of the algorithm can be found in Algorithm 5.

Algorithm 5: IDEALRESAMPLE subroutine

Parameters: real $\beta \in (0,1)$, integer $q \geq 3$

Input : Graph $G = (V, E)$, vertices $u, v \in V$ with $\{u, v\} \notin E$, configuration $\sigma : V \to [q]$ with $\sigma(u) = \sigma(v)$

Output : A configuration $\sigma' : V \to [q]$

Function IDEALRESAMPLE(G, u, v, σ) **begin**

- $r \leftarrow$ flip a coin with heads probability $\frac{q\beta}{\beta+q-1}$;
- **if** $r =$ head **then**
  - **return** $\sigma' \leftarrow \sigma$;
- **end**
- **else**
  - Pick u.a.r. a colour c' from $[q]/\{c\}$, where $c = \sigma(u) = \sigma(v)$;
  - $U \leftarrow \sigma^{-1}(c, c')$;
  - $H \leftarrow G[U]$;
  - `/* Sample` τ `from Ising on` H `conditioned on` $\tau(u) = c, \tau(v) = c'$ `*/`
  - Sample $\tau : U \to \{c, c'\}$ with $\tau(u) = c$ and $\tau(v) = c'$ so that $\tau \sim \pi^{c,c'}_{H,u,v}$;
  - **return** $\sigma' : V \to [q]$ *s.t.* $\sigma'(w) = \tau(w)$ *for* $w \in U$ *and* $\sigma'(w) = \sigma(w)$ *for* $w \notin U$*;*
- **end**

end

Remark. It is an ideal subroutine; we will later modify this to obtain the actual RESAMPLE subroutine used in Algorithm SAMPLEANTIPOTTS (cf. Algorithm 7).

Definition 5.3.3. Let $\beta > 0$. Suppose that $G = (V, E)$ is a graph and that u, v are vertices in G. For a set $U \subseteq V$ such that $u, v \in U$, let

$$\mathsf{Corr}_G(U, u, v) = \left| \frac{\pi_{G[U]}\big(\eta(u) = 1, \eta(v) = 1\big)}{\pi_{G[U]}\big(\eta(u) = 1, \eta(v) = 2\big)} - 1 \right|$$

i.e., $\mathsf{Corr}_G(U, u, v)$ measures the correlation between u and v in the Ising distribution with parameter β on the subgraph $G[U]$.

Definition 5.3.4. Let $G = (V, E)$ be a graph, u, v be vertices in G and $\sigma : V \to [q]$ be a configuration on G. We let $U_\sigma \subseteq V$ be a bichromatic class under σ which contains u and v, chosen uniformly at random among the set of all such classes if there is more than one. More precisely, if $\sigma(u) \neq \sigma(v)$, then U_σ is the (c_1, c_2)-colour-class in σ where c_1, c_2 are the colours of u and v under σ. If $\sigma(u) = \sigma(v)$, then U_σ is the (c, c')-colour-class in σ, where c is the common colour of u and v under σ and c' is a uniformly random colour from $[q] \setminus \{c\}$.

Lemma 5.3.5. *Let $\beta \in (0,1)$ and $q \geq 3$ be an integer, and consider arbitrary $\varepsilon \in (0,1)$. Let $G = (V,E)$ be a graph and μ_G be the Potts distribution on G with parameter β. Suppose that u, v are vertices in G such that $\{u,v\} \notin E$ and*

$$\mathbf{E}\big[\mathsf{Corr}_G(U_\sigma, u, v)\big] \leq \varepsilon,$$

where the expectation is over the choice of a random configuration $\sigma \sim \mu_G$ and the choice of a random bichromatic class $U_\sigma \subseteq V$ containing u and v under σ (cf. Definition 5.3.4). Consider the graph $G' = (V, E')$ obtained from G by adding the edge $\{u,v\}$. Sample a configuration $\sigma' : V \to [q]$ as follows. First, sample $\sigma : V \to [q]$ according to μ_G. Then, if $\sigma(u) \neq \sigma(v)$, set $\sigma' = \sigma$; otherwise, set $\sigma' = \textsc{IdealReSample}(G, u, v, \sigma)$. Then, the distribution of σ', denoted by $\nu_{\sigma'}$, is within total variation distance $2\varepsilon/\beta$ from the Potts distribution $\mu_{G'}$ on G' with parameter β, i.e.,

$$\|\nu_{\sigma'} - \mu_{G'}\|_{\mathrm{TV}} \leq 2\varepsilon/\beta.$$

Proof. We begin with a few definitions that will be used throughout the proof. Fix distinct colours $c_1, c_2 \in [q]$ and a set $U \subseteq V$ such that $u, v \in U$. Let $\Omega(U, c_1, c_2)$ be the set of configurations such that U is the (c_1, c_2)-colour-class, i.e.,

$$\Omega(U, c_1, c_2) = \big\{\eta : V \to [q] \mid U = \eta^{-1}(c_1, c_2)\big\}.$$

We will be interested in two particular types of configurations in $\Omega(U, c_1, c_2)$, those where u, v take the colours c_1, c_2 and those where u, v take the colours c_1, c_1. Namely, let

$$\Omega_{c_1,c_2}(U, c_1, c_2) = \big\{\eta \in \Omega(U, c_1, c_2) \mid \eta(u) = c_1,\ \eta(v) = c_2\big\},$$
$$\Omega_{c_1,c_1}(U, c_1, c_2) = \big\{\eta \in \Omega(U, c_1, c_2) \mid \eta(u) = c_1,\ \eta(v) = c_1\big\}.$$

For a configuration $\eta \in \Omega(U, c_1, c_2)$, we also denote by

$$\Omega^\eta(U, c_1, c_2) = \big\{\tau \in \Omega(U, c_1, c_2) \mid \tau(V \setminus U) = \eta(V \setminus U)\big\}$$

the set of configurations in $\Omega(U, c_1, c_2)$ that agree with η on $V \setminus U$; we define analogously the sets $\Omega^\eta_{c_1,c_2}(U, c_1, c_2), \Omega^\eta_{c_1,c_1}(U, c_1, c_2)$. Using Observation 5.3.1, we have that for any $\eta \in \Omega(U, c_1, c_2)$ it holds that

$$\mu_G(\eta) = \mu_G\big(\Omega^\eta(U, c_1, c_2)\big)\, \pi^{c_1,c_2}_{G[U]}(\eta(U)), \tag{5.4}$$

and

$$\frac{\mu_G\big(\Omega^{\eta}_{c_1,c_1}(U,c_1,c_2)\big)}{\mu_G\big(\Omega^{\eta}(U,c_1,c_2)\big)} = \pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_1),$$
$$\frac{\mu_G\big(\Omega^{\eta}_{c_1,c_2}(U,c_1,c_2)\big)}{\mu_G\big(\Omega^{\eta}(U,c_1,c_2)\big)} = \pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_2). \tag{5.5}$$

The assumption $\mathbf{E}\big[\mathsf{Corr}_G(U_\sigma,u,v)\big] \le \varepsilon$ translates into

$$\sum_{\substack{c_1,c_2\in[q];\\ c_1\neq c_2}} \sum_{\substack{U\subseteq V;\\ u,v\in U}} \left(\mu_G\big(\Omega_{c_1,c_2}(U,c_1,c_2)\big)+\frac{\mu_G\big(\Omega_{c_1,c_1}(U,c_1,c_2)\big)}{q-1}\right)\left|\frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_2)}-1\right| \le \varepsilon. \tag{5.6}$$

We next proceed to the proof. Recall that $\nu_{\sigma'}$ is the distribution of the output σ', our goal is to show that

$$\|\nu_{\sigma'}-\mu_{G'}\|_{\mathrm{TV}} \le 2\varepsilon/\beta. \tag{5.7}$$

We first show that

$$\forall \eta\in\Omega_{c_1,c_1}(U,c_1,c_2): \quad \nu_{\sigma'}(\eta) = \frac{q\beta}{\beta+q-1}\mu_G(\eta), \tag{5.8}$$
$$\forall \eta\in\Omega_{c_1,c_2}(U,c_1,c_2): \quad \nu_{\sigma'}(\eta) = \mu_G(\eta) + \frac{(1-\beta)\pi^{c_1,c_2}_{G[U],u,v}(\eta(U))}{\beta+q-1}\mu_G\big(\Omega^{\eta}_{c_1,c_1}(U,c_1,c_2)\big). \tag{5.9}$$

To see the expression for $\nu_{\sigma'}(\eta)$ in (5.8), note that we can obtain $\eta\in\Omega_{c_1,c_1}(U,c_1,c_2)$ via the subroutine IDEALRESAMPLE only if $\sigma=\eta$ and the coin flip came up heads (because $\eta(u)=\eta(v)=c_1$). Analogously, to see the expression for $\nu_{\sigma'}(\eta)$ in (5.9), note that we can obtain $\eta\in\Omega_{c_1,c_2}(U,c_1,c_2)$ if one of the following happens (using that $\eta(u)=c_1$ and $\eta(v)=c_2$).

- We started with the configuration $\sigma=\eta$; this happens with probability $\mu_G(\eta)$. Then, we obtain η with probability 1.
- We started with a configuration σ such that $\sigma(u)=\sigma(v)=c_1$, $\sigma^{-1}(c_1,c_2)=U$ and $\sigma(V\setminus U)=\eta(V\setminus U)$, i.e., $\sigma\in\Omega^{\eta}_{c_1,c_1}(U,c_1,c_2)$; this happens with probability $\mu_G\big(\Omega^{\eta}_{c_1,c_1}(U,c_1,c_2)\big)$. Then the coin flip came tails, the colour c' selected was $c'=c_2$, and the Ising configuration τ sampled was $\eta(U)$. All that happens with probability $\frac{(1-\beta)\pi^{c_1,c_2}_{G[U],u,v}(\eta(U))}{\beta+q-1}$.

Using the assumption $\mathbf{E}\big[\mathsf{Corr}_G(U_\sigma, u, v)\big] \le \varepsilon$ (cf. (5.6)), we will later show the following bound for the ratio of the partition functions Z_G, Z'_G:

$$|M| \le \varepsilon/\beta, \text{ where } M := \frac{(\beta+q-1)Z_G}{qZ_{G'}} - 1. \tag{5.10}$$

Let us conclude the proof assuming, for now, (5.10).
For $\eta \in \Omega_{c_1,c_1}(U, c_1, c_2)$, we have $\mu_{G'}(\eta) = \frac{w_{G'}(\eta)}{Z_{G'}} = \frac{\beta w_G(\eta)}{Z_{G'}} = \frac{q\beta(M+1)\mu_G(\eta)}{\beta+q-1}$ and hence, by (5.8),

$$|\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)| = \frac{q\beta}{\beta+q-1}\Big|\mu_G(\eta) - (M+1)\mu_G(\eta)\Big| \le |M|\,\mu_G(\eta), \tag{5.11}$$

where the last inequality follows from $\frac{q\beta}{\beta+q-1} \le 1$ which holds for all $\beta \in (0,1)$.
For $\eta \in \Omega_{c_1,c_2}(U, c_1, c_2)$, we have $\mu_{G'}(\eta) = \frac{w_{G'}(\eta)}{Z_{G'}} = \frac{w_G(\eta)}{Z_{G'}} = \frac{q(M+1)}{\beta+q-1}\mu_G(\eta)$ and hence, by (5.9),

$$\begin{aligned}|\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)| &= \Big|\frac{1-\beta}{\beta+q-1}\big(\mu_G(\eta) - \pi^{c_1,c_2}_{G[U],u,v}(\eta(U))\mu_G(\Omega^\eta_{c_1,c_1}(U,c_1,c_2))\big) + \frac{qM\mu_G(\eta)}{\beta+q-1}\Big| \\ &\le \Big|\mu_G(\eta) - \pi^{c_1,c_2}_{G[U],u,v}(\eta(U))\mu_G\big(\Omega^\eta_{c_1,c_1}(U,c_1,c_2)\big)\Big| + 2|M|\,\mu_G(\eta), \end{aligned} \tag{5.12}$$

where the last inequality follows from the triangle inequality and the inequalities $\frac{1-\beta}{\beta+q-1} \le 1$, $\frac{q}{\beta+q-1} \le 2$. Note that, by the definition of $\pi^{c_1,c_2}_{G[U],u,v}(\cdot)$, for $\eta \in \Omega_{c_1,c_2}(U, c_1, c_2)$ it holds that

$$\pi^{c_1,c_2}_{G[U],u,v}(\eta(U)) = \frac{\pi^{c_1,c_2}_{G[U]}(\eta(U))}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1, \tau(v)=c_2)}$$

and therefore, using (5.4) and (5.5), we have that

$$\Big|\mu_G(\eta) - \pi^{c_1,c_2}_{G[U],u,v}(\eta(U))\mu_G\big(\Omega^\eta_{c_1,c_1}(U,c_1,c_2)\big)\Big| = \mu_G(\eta)\left|\frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_2)} - 1\right|, \tag{5.13}$$

and hence (5.12) gives that, for all $\eta \in \Omega_{c_1,c_2}(U, c_1, c_2)$, it holds that

$$|\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)| \le \left(2|M| + \left|\frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_2)} - 1\right|\right)\mu_G(\eta). \tag{5.14}$$

To proceed, we will decompose $\|\nu_{\sigma'} - \mu_{G'}\|_{\mathrm{TV}}$ as

$$\|\nu_{\sigma'} - \mu_{G'}\|_{\mathrm{TV}} = \frac{1}{2}\sum_{\eta: V\to[q]} |\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)| = \frac{1}{2}\sum_{U;\, u,v\in U}\ \sum_{c_1,c_2\in[q];\, c_1\neq c_2} D(U, c_1, c_2), \tag{5.15}$$

where

$$D(U,c_1,c_2) := \sum_{\eta\in\Omega_{c_1,c_2}(U,c_1,c_2)} |\nu_{\sigma'}(\eta)-\mu_{G'}(\eta)| + \frac{1}{q-1}\sum_{\eta\in\Omega_{c_1,c_1}(U,c_1,c_2)} |\nu_{\sigma'}(\eta)-\mu_{G'}(\eta)|. \tag{5.16}$$

To see the second equality in (5.15), fix any configuration $\eta : V \to [q]$. If $\eta(u) = c_1$ and $\eta(v) = c_2$ for distinct colours $c_1, c_2 \in [q]$, then $|\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)|$ appears only in one term $D(U, c_1, c_2)$ in the summation of (5.15), namely for $U = \eta^{-1}(c_1, c_2)$. If $\eta(u) = \eta(v) = c_1$ for some colour $c_1 \in [q]$, then $|\nu_{\sigma'}(\eta) - \mu_{G'}(\eta)|$ appears in exactly $q-1$ terms $D(U, c_1, c_2)$ in the summation of (5.15), once for each colour $c_2 \neq c_1$ and $U = \eta^{-1}(c_1, c_2)$.
Summing (5.11) and (5.14) over the relevant configurations η, we obtain that, for all $U \subseteq V$ with $u, v \in U$ and distinct colours $c_1, c_2 \in [q]$, it holds that

$$D(U,c_1,c_2) \le \left(2|M| + \left|\frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1,\tau(v)=c_2)} - 1\right|\right)\left(\mu_G\big(\Omega_{c_1,c_2}(U,c_1,c_2)\big) + \frac{\mu_G\big(\Omega_{c_1,c_1}(U,c_1,c_2)\big)}{q-1}\right). \tag{5.17}$$

To conclude the proof of (5.7), note that analogously to (5.15) we have that

$$\sum_{U;\,u,v\in U}\ \sum_{c_1,c_2\in[q];\,c_1\neq c_2}\left(\mu_G\big(\Omega_{c_1,c_2}(U,c_1,c_2)\big) + \frac{\mu_G\big(\Omega_{c_1,c_1}(U,c_1,c_2)\big)}{q-1}\right) = 1. \tag{5.18}$$

It follows from (5.6), (5.15), (5.17) and (5.18) that

$$\|\nu_{\sigma'} - \mu_{G'}\|_{\mathrm{TV}} \le \frac{1}{2}(2|M|+\varepsilon) \le \tfrac{3}{2\beta}\varepsilon \le 2\varepsilon/\beta,$$

where the last inequality follows from $|M| \le \varepsilon/\beta$ (cf. (5.10)). This finishes the proof of (5.7), modulo the proof of (5.10) which is given below.
To prove (5.10), it will be useful to define

$$X(U,c_1,c_2) := \sum_{\eta\in\Omega_{c_1,c_2}(U,c_1,c_2)} w_G(\eta), \quad Y(U,c_1,c_2) := \sum_{\eta\in\Omega_{c_1,c_1}(U,c_1,c_2)} w_G(\eta),$$

and

$$Z(U,c_1,c_2) = X(U,c_1,c_2) + \frac{1}{q-1}Y(U,c_1,c_2). \tag{5.19}$$

Analogously to (5.15) (and (5.18)), we have that

$$Z_G = \sum_{c_1\neq c_2}\ \sum_{U;u,v\in U} Z(U,c_1,c_2). \tag{5.20}$$

Using that $w_{G'}(\eta) = w_G(\eta)$ for $\eta \in \Omega_{c_1,c_2}(U,c_1,c_2)$ and $w_{G'}(\eta) = \beta w_G(\eta)$ for $\eta \in \Omega_{c_1,c_1}(U,c_1,c_2)$, we obtain that

$$Z_{G'} = \sum_{c_1 \neq c_2} \sum_{U : u,v \in U} Z'(U,c_1,c_2), \text{ where } Z'(U,c_1,c_2) := X(U,c_1,c_2) + \frac{\beta}{q-1} Y(U,c_1,c_2). \tag{5.21}$$

We first show that, for all $U \subseteq V$ with $u,v \in U$ and distinct colours $c_1, c_2 \in [q]$, it holds that

$$\left| \frac{(\beta+q-1)Z(U,c_1,c_2)}{qZ'(U,c_1,c_2)} - 1 \right| \leq \frac{1}{q} \cdot \left| \frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1, \tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1, \tau(v)=c_2)} - 1 \right|. \tag{5.22}$$

Observe that

$$\frac{(\beta+q-1)Z(U,c_1,c_2)}{qZ'(U,c_1,c_2)} - 1 = \frac{1-\beta}{q} \frac{Y(U,c_1,c_2) - X(U,c_1,c_2)}{X(U,c_1,c_2) + \frac{\beta}{q-1} Y(U,c_1,c_2)}. \tag{5.23}$$

Observe also that (cf. (5.5))

$$\left| \frac{Y(U,c_1,c_2)}{X(U,c_1,c_2)} - 1 \right| = \left| \frac{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1, \tau(v)=c_1)}{\pi^{c_1,c_2}_{G[U]}(\tau(u)=c_1, \tau(v)=c_2)} - 1 \right|,$$

which combined with (5.23) (and ignoring the positive term $\frac{\beta}{q-1} Y(U,c_1,c_2)$ in the denominator of the r.h.s. in the latter) gives (5.22), using also that $\beta \in (0,1)$.
Using (5.20) and (5.21), we have that

$$\frac{(\beta+q-1)Z_G}{qZ_{G'}} - 1 = \sum_{c_1 \neq c_2} \sum_{U;\, u,v \in U} \frac{qZ'(U,c_1,c_2)}{Z_{G'}} \left(\frac{(\beta+q-1)Z(U,c_1,c_2)}{qZ'(U,c_1,c_2)} - 1 \right). \tag{5.24}$$

Since $\beta \in (0,1)$, we have for all $U \subseteq V$ with $u,v \in U$ and distinct colours $c_1, c_2 \in [q]$ that $Z'(U,c_1,c_2) \leq Z(U,c_1,c_2)$ and $Z_{G'} \geq \beta Z_G$, therefore

$$\frac{Z'(U,c_1,c_2)}{Z_{G'}} \leq \frac{Z(U,c_1,c_2)}{\beta Z_G} = \frac{1}{\beta} \left(\mu_G\big(\Omega_{c_1,c_2}(U,c_1,c_2)\big) + \frac{\mu_G\big(\Omega_{c_1,c_1}(U,c_1,c_2)\big)}{q-1} \right). \tag{5.25}$$

Combining (5.22), (5.24) and (5.25), we obtain by the triangle inequality that (cf. (5.6))

$$\left| \frac{(\beta+q-1)Z_G}{qZ_{G'}} - 1 \right| \leq \mathbf{E}\big[\mathsf{Corr}_G(U_\sigma,u,v)\big]/\beta \leq \varepsilon/\beta,$$

thus completing the proof of (5.10), and therefore the proof of Lemma 5.3.5. □

5.3.3 Sampling the Ising model on graphs with small average growth

Recall that our algorithm for sampling the antiferromagnetic Potts model with parameter β will use as a subroutine a sampling algorithm for the Ising model with parameter β to recolour bichromatic classes. In general, these classes may consist of large bichromatic components (with a linear number of vertices), so to carry out this subroutine efficiently, we need to use an approximate sampling algorithm; our leverage point will be that we can bound the average growth of bichromatic components. Adapting results of [64], we show the following in Section 5.5.2.

Theorem 5.3.6. *Let $\beta \in (0,1)$ and $b > 0$ be constants such that $b\frac{1-\beta}{1+\beta} < 1$, and let $\Delta \geq 3$ be an integer. Then, there exists $M_0 > 0$ such that the following holds for all $M > M_0$.*
There is a polynomial-time algorithm that, on input an n-vertex graph G with maximum degree at most Δ and average growth b up to depth $L = \lceil M \log n \rceil$, outputs a configuration $\tau : V \to \{1,2\}$ whose distribution ν_τ is within total variation distance $1/n^{10}$ from the Ising distribution on G with parameter β, i.e.,

$$\left\|\nu_\tau - \pi_G\right\|_{\mathrm{TV}} \leq 1/n^{10}.$$

Moreover, the algorithm, when given as additional input two vertices u and v in G, outputs a configuration $\tau : V \to \{1,2\}$ such that $\tau(u) = 1$ and $\tau(v) = 2$, and whose distribution ν_τ satisfies

$$\left\|\nu_\tau - \pi_{G,u,v}(\cdot)\right\|_{\mathrm{TV}} \leq 1/n^{10},$$

where $\pi_{G,u,v}$ is the Ising distribution on G conditioned on u having state 1 and v having state 2.

In addition, we will use the following spatial mixing result to analyse the accuracy of our algorithm for the antiferromagnetic Potts model. The proof is given in Section 5.5.3.

Lemma 5.3.7. *Let $\beta \in (0,1)$ and $b > 0$ be constants such that $b\frac{1-\beta}{1+\beta} < 1$. Then, there exists $M_0' > 0$ such that the following holds for all $M > M_0'$.*
Let G be an n-vertex graph with average growth b up to depth $L = \lceil M \log n \rceil$, and let u, v be distinct vertices in G. Then

$$\left|\pi_G(\sigma(u) = 1 \mid \sigma(v) = 1) - \pi_G(\sigma(u) = 1 \mid \sigma(v) = 2)\right| \leq \frac{1}{n^{10}} + \sum_{\ell=1}^{L} P_\ell(G,u,v)\Big(\frac{1-\beta}{1+\beta}\Big)^\ell$$

where $P_\ell(G,u,v)$ is the number of paths with ℓ vertices in G that connect u and v.

We will also need the following crude bound.

Lemma 5.3.8. *Let $\beta \in (0,1)$ and $\Delta \geq 3$ be an integer. Suppose that G is a graph of maximum degree at most Δ and let u be a vertex and Λ be a set of vertices in G such that $u \notin \Lambda$. Then, for every configuration $\tau : \Lambda \to \{1,2\}$ and $s \in \{1,2\}$*

$$\pi_G(\sigma(u) = s \mid \sigma(\Lambda) = \tau) \geq \frac{\beta^\Delta}{1+\beta^\Delta}\,.$$

Proof. Without loss of generality, we assume that $s = 1$. Let D be the number of neighbours of u and $u_1, \ldots, u_d$ be the neighbours of u in $G \setminus \Lambda$ and note that $d \leq D \leq \Delta$. Let d_0 be the number of $v \in \Lambda$ such that v is u's neighbour and $\tau(v) = 1$. Let $s_1, \ldots, s_d \in \{1,2\}$ be arbitrary and let d_1 be the number of the s_i's that are equal to 1. Then we have that

$$\pi_G(\sigma(u) = 1 \mid \sigma(u_1) = s_1, \ldots, \sigma(u_d) = s_d, \sigma(\Lambda) = \tau) = \frac{\beta^{d_2}}{\beta^{d_2} + \beta^{D-d_2}},$$

where $d_2 = d_0 + d_1 \leq d_0 + d \leq D$. Since $u_1, \ldots, u_d \notin \Lambda$ we have that $\pi_G(\sigma(u_1) = s_1, \ldots, \sigma(u_d) = s_d \mid \sigma(\Lambda) = \tau) \neq 0$ for any choice of $s_1, \ldots s_d$ and therefore by the law of total probability we obtain

$$\pi_G(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau) \geq \min_{d_2 \in \{d_0, \ldots, d_0+d\}} \left\{ \frac{\beta^{d_2}}{\beta^{d_2} + \beta^{D-d_2}} \right\}.$$

The function $\frac{\beta^x}{\beta^x + \beta^{D-x}}$ is decreasing for $x \in [0, D]$ (since $\beta \in (0,1)$) and hence

$$\min_{d_2 \in \{d_0, \ldots, d_0+d\}} \left\{ \frac{\beta^{d_2}}{\beta^{d_2} + \beta^{D-d_2}} \right\} = \frac{\beta^D}{1+\beta^D}.$$

Using that $\beta \in (0,1)$ and $D \leq \Delta$, we obtain the inequalities in the statement of the lemma. □

We are now able to describe our RESAMPLE subroutine, which works for graphs with small average growth.

Definition 5.3.9. Let $b, M > 0$ be constants. Let $G = (V, E)$ be an n-vertex graph and let σ be a configuration on G. We say that a bichromatic component in σ is (b, M)-*good* if it has average growth b up to depth $L = \lceil M \log n \rceil$; we say that it is (b, M)-*bad* otherwise. Analogously, we say that σ is (b, M)-*good* if *all* bichromatic components in σ are good; otherwise, we say that σ is (b, M)-bad.

Our ReSample takes as inputs a graph G, two vertices u and v of G and a configuration σ on G such that $\sigma(u) = \sigma(v)$, see Algorithm 6.

Algorithm 6: ReSample subroutine used in SampleAntiPotts (cf. Algorithm 7).

Parameters: real $\beta \in (0,1)$, $b, M > 0$ satisfying Theorem 5.3.6, integer $q \geq 3$

Input : Graph $G = (V, E)$, vertices $u, v \in V$ with $\{u, v\} \notin E$, configuration $\sigma : V \to [q]$ with $\sigma(u) = \sigma(v)$

Output : A configuration $\sigma' : V \to [q]$

Function $\text{ReSample}_{b,M}(G, u, v, \sigma)$
begin
  if *σ contain a bichromatic component which is (b, M)-bad* **then**
    return *$\sigma' : V \to [q]$ selected uniformly at random;*
  end
  else
    $r \leftarrow$ flip a coin with heads probability $\frac{q\beta}{\beta+q-1}$;
    if $r =$ heads **then**
      return $\sigma' \leftarrow \sigma$;
    end
    else
      Pick u.a.r. a colour c' from $[q]/\{c\}$, where $c = \sigma(u) = \sigma(v)$;
      $U \leftarrow \sigma^{-1}(c, c')$;
      $H \leftarrow G[U]$;
      `/* Use algorithm of Theorem 5.3.6 to sample from Ising on` H `*/`
      Sample $\tau : U \to \{c, c'\}$ with $\tau(u) = c$ and $\tau(v) = c'$ so that $\|\nu_\tau - \pi^{c,c'}_{H,u,v}\| \leq \frac{1}{n^{10}}$
    end
    return *$\sigma' : V \to [q]$ s.t. $\sigma'(w) = \tau(w)$ for $w \in U$ and $\sigma'(w) = \sigma(w)$ for $w \notin U$;*
  end
end

To utilise Theorem 5.3.6 and Lemma 5.3.7 for our sampling algorithms, we will need to bound the average growth of bichromatic components in a typical Potts configuration on a random regular graph. Our key lemma to achieve this will bound the probability that a path is bichromatic[1] in uniqueness, provided that the local neighbourhood around the path (in the sense of Definition 5.2.3) has a tree-like structure. The following lemma quantifies this probability bound and is proved in Section 5.4.2.

Lemma 5.3.10. *Let $\Delta, q \geq 3$ be integers, and $\beta \in (0,1)$ be in the uniqueness regime of the*

[1] Let $G = (V, E)$ be a graph and $\sigma : V \to [q]$. We call a path P bichromatic under σ if there exist colours $c_1, c_2 \in [q]$ such that every vertex u of P satisfies $\sigma(u) \in \{c_1, c_2\}$.

$(\Delta-1)$-ary tree with $\beta \neq (\Delta-q)/\Delta$. Then, for any $\varepsilon' > 0$, there exists a positive constant $K < \frac{1+\beta}{\beta+q-1} + \varepsilon'$ and $\varepsilon > 0$ such that the following holds for all sufficiently large integers ℓ and h.

Let G be a graph of maximum degree Δ and P be a path with ℓ vertices whose h-graph-neighbourhood contains $(1-\varepsilon)\ell$ isolated tree components. Let μ_G be the Potts measure on G with parameter β. Then,

$$\mu_G(\textit{path } P \textit{ is bichromatic}) \leq K^{\ell}.$$

For a random Δ-regular graph G, paths do have the tree-like structure of Lemma 5.3.10 (cf. Lemma 5.2.4), and hence we can aggregate over all paths emanating from an arbitrary vertex (roughly $(\Delta-1)^{\ell}$ of them) and get a bound of roughly $(\Delta-1)K < \frac{(\Delta-1)(1+\beta)}{\beta+q-1}$ for the average growth of bichromatic components in a typical configuration σ. This will allow us to use the ReSample subroutine from Algorithm 6.

5.3.4 Analysis of the Potts algorithm

Algorithm 7: SAMPLEANTIPOTTS(G) sampling a Potts configuration in the case $\beta \in (0,1)$

Parameters: real $\beta \in (0,1)$, $b, M > 0$ satisfying Theorem 5.3.6, integers $q, \Delta \geq 3$

Input : Graph $G = (V, E)$
Output : Either FAIL or an assignment $\sigma : V \to [q]$

Function SAMPLEANTIPOTTS(G)
begin
  if $G' = (V, E')$ *contains a component which is neither a cycle nor an isolated vertex* **then**
    return FAIL;
  end
  else
    Sample a configuration $\sigma' : V \to [q]$ on G' (according to $\mu_{G'}$);
    $e_1, e_2, \ldots, e_t \leftarrow$ edges in $E \setminus E'$;
    $G_t \leftarrow G'$;
    $\sigma_t \leftarrow \sigma'$;
    for $j = t$ **downto** 1 **do**
      $(u_j, v_j) \leftarrow e_j$;
      if $\sigma_j(u_j) = \sigma_j(v_j)$ **then**
        $\sigma_{j-1} \leftarrow \text{RESAMPLE}_{b,M}(G_j, u_j, v_j, \sigma_j)$;
      end
      else
        $\sigma_{j-1} = \sigma_j$;
      end
      $G_{j-1} \leftarrow$ adding the edge $\{u_j, v_j\}$ in G_j;
    end
    return $\sigma \leftarrow \sigma_0$;
  end
end

In this Chapter, we will present the proof of Theorem 5.1.1. Specifically, our sampling algorithm is describe in Algorithm 7 and thus Theorem 5.1.1 follows directly from Theorem 5.3.11.

Theorem 5.3.11. *Let $\Delta \geq 3$, $q \geq 3$ and $\beta \in (0,1)$ be in the uniqueness regime of the $(\Delta - 1)$-ary tree. Then, there exists constants $b, M, \delta > 0$ such that, for all sufficiently large n, the following holds with probability $1 - o(1)$ over the choice of a random Δ-regular graph $G = (V, E)$ with n vertices.*

The output of the algorithm SAMPLEANTIPOTTS(G) *(cf. Algorithm 7) is an assignment*

$\sigma : V \to [q]$ whose distribution ν_σ is within total variation distance $O(1/n^\delta)$ from the Potts distribution μ_G with parameter β, i.e.,

$$\|\nu_\sigma - \mu_G\|_{\mathrm{TV}} = O(1/n^\delta).$$

Proof. Since β is in the uniqueness regime of the $(\Delta-1)$-ary tree and $\beta \neq \frac{\Delta-q}{\Delta}$, we have that $\beta > \frac{\Delta-q}{\Delta}$. It follows that $\frac{1-\beta}{\beta+q-1} < \frac{1}{\Delta-1}$ and therefore there exists $\varepsilon' > 0$ such that

$$\Big(\frac{1+\beta}{\beta+q-1} + 3\varepsilon'\Big)\frac{1-\beta}{1+\beta} < \frac{1}{\Delta-1}. \tag{5.26}$$

Let $K < \frac{1+\beta}{\beta+q-1} + \varepsilon'$ and $\varepsilon > 0$ be the constants in Lemma 5.3.10 corresponding to ε', and let h', ℓ' be positive constants such that Lemma 5.3.10 applies for all integers $h \geq h'$ and $\ell \geq \ell'$. Fix h to be any integer greater than h'. Let $\ell_1 > 0$ be the constant in Lemma 5.2.4 corresponding to the values of ε and h. Let $\delta > 0$ be the constant in Lemma 5.2.2 corresponding to $\ell_0 := 1/(5\log(\Delta-1))$ and $W := 1/(K\frac{1-\beta}{1+\beta})$ (note that (5.26) guarantees that $W > \Delta - 1$). Let also

$$b' := (\Delta-1)\Big(\frac{1+\beta}{\beta+q-1} + 2\varepsilon'\Big), \text{ and } b := (\Delta-1)\Big(\frac{1+\beta}{\beta+q-1} + 3\varepsilon'\Big).$$

Let M_0, M_0' be the constants in Theorem 5.3.6 and Lemma 5.3.7, respectively. Let M be sufficiently large so that $M > \max\{2\ell_0, 2M_0, 2M_0'\}$ and the following inequalities hold (for all sufficiently large n):

$$\Delta\big((\Delta-1)K\big)^{\lceil M\log n\rceil} \leq (b')^{\lceil M\log n\rceil} \text{ and } (b'/b)^{\lceil M\log n\rceil} \leq 1/n^{11}. \tag{5.27}$$

Note that such an M exists since $(\Delta-1)K < b' < b$. Finally, set

$$L_0 := \lceil \ell_0 \log n\rceil = \lceil \tfrac{1}{5}\log_{\Delta-1} n\rceil, \quad L := \lceil M\log n\rceil.$$

Taking a union bound over Lemmas 5.2.2 and 5.2.4, we have that, for all sufficiently large n, a uniformly random Δ-regular graph $G = (V, E)$ with n vertices satisfies the following with probability $1 - o(1)$ over the choice of the graph:

1. $\displaystyle\sum_{\ell=L_0}^{L} \ell C_\ell \Big(K\frac{1-\beta}{1+\beta}\Big)^\ell \leq 1/(2n^\delta)$, where C_ℓ is the number of cycles of length ℓ in G.
2. every path P in G with ℓ vertices where $\ell_1 \leq \ell \leq L$ has an h-graph-neighbourhood with at least $(1-\varepsilon)\ell$ isolated tree components.

Fix any Δ-regular graph G. The theorem will follow by showing that the output of SAMPLEANTIPOTTS(G) (cf. Algorithm 7) is an assignment $\sigma : V \to [q]$ whose distribution ν_σ is within total variation distance $O(1/n^\delta)$ from the Potts measure μ_G with parameter β, i.e.,

$$\|\nu_\sigma - \mu_G\|_{\mathrm{TV}} = O(1/n^\delta).$$

To do this, as in the algorithm SAMPLEANTIPOTTS(G), let $e_1 = \{u_1, v_1\}, \ldots, e_t = \{u_t, v_t\}$ be the edges of G that do not belong to short cycles (i.e., cycles of length $\leq \ell_0 \log n$). For $j \in \{0, 1, \ldots, t\}$, let G_j be the subgraph $G \setminus \{e_1, \ldots, e_j\}$ and μ_j be the Potts distribution on G_j with parameter β; note that the graphs G_j are defined exactly just as in the algorithm SAMPLEANTIPOTTS(G). We will use $\hat{\sigma}_j$ to denote a random configuration distributed according to μ_j; note that σ_j is used to denote the configuration considered by the algorithm at the beginning of the step j of the algorithm and, as we shall see soon, its distribution is close to that of $\hat{\sigma}_j$ on G_j.

For an integer $\ell \geq 1$, denote by $\mathcal{P}_{\ell,j}$ the set of paths of length ℓ that connect u_j and v_j in G_j and by $P_{\ell,j} = |\mathcal{P}_{\ell,j}|$ the number of all such paths. Note that

$$\textstyle\sum_{j=1}^{t} P_{\ell,j} \leq \ell C_\ell, \tag{5.28}$$

since every path with ℓ vertices connecting the endpoints of an edge $\{u_j, v_j\}$ in G_j maps to a cycle with ℓ vertices in the initial graph G (by adding the edge $\{u_j, v_j\}$), and each cycle with ℓ vertices in G can potentially arise at most ℓ times under this mapping. Let also

$$\varepsilon_j := \frac{6}{n^{10}} + \sum_{\ell=L_0}^{L} P_{\ell,j} \Big(K \frac{1-\beta}{1+\beta}\Big)^\ell. \tag{5.29}$$

Let also $\Omega_j(b, M)$ be the set of all (b, M)-bad configurations on G_j. Using the fact that G satisfies Item 2, we will show that for all $j = 1, \ldots, t$ it holds that

$$\mu_{G_j}\big(\Omega_j(b, M)\big) \leq 1/n^{10}, \tag{5.30}$$

$$\mathbf{E}_j\big[\mathsf{Corr}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j)\big] \leq 2\varepsilon_j/\beta^\Delta, \tag{5.31}$$

where the expectation in (5.31) is over the choice of a random configuration $\hat{\sigma}_j$ distributed according to μ_j and over the choice of the random bichromatic class $U_{\hat{\sigma}_j}$ containing u_j and v_j under $\hat{\sigma}_j$.

We will prove (5.30) and (5.31) shortly, but let us assume them for now and conclude the proof of the theorem. We will prove by induction that for all $j \in \{0, 1, \ldots, t\}$ it holds

that

$$\left\|\nu_{\sigma_j} - \mu_{G_j}\right\|_{\mathrm{TV}} \leq \frac{6(t-j)}{n^{10}} + \frac{4}{\beta^{\Delta+1}} \sum_{j'=j+1}^{t} \varepsilon_{j'}. \tag{5.32}$$

For $j = t$, the result holds trivially since σ_t is distributed as μ_{G_t} (exactly). Assume that (5.32) holds for j where $j \in \{1, \ldots, t\}$, we will also show that it holds for $j-1$. To do this, let us consider the configuration σ'_{j-1} defined as follows:

$$\text{if } \hat{\sigma}_j(u_j) \neq \hat{\sigma}_j(v_j), \text{ then } \sigma'_{j-1} = \hat{\sigma}_j \text{ else } \sigma'_{j-1} = \textsc{IdealReSample}(G_j, u_j, v_j, \hat{\sigma}_j),$$

i.e., σ'_{j-1} is obtained using at the j-th step of the algorithm the random configuration $\hat{\sigma}_j$ distributed according to μ_{G_j} (exactly). In contrast, note that

$$\text{if } \sigma_j(u_j) \neq \sigma_j(v_j), \text{ then } \sigma_{j-1} = \sigma_j \text{ else } \sigma_{j-1} = \textsc{ReSample}_{b,M}(G_j, u_j, v_j, \sigma_j).$$

Let $\nu_{\sigma'_{j-1}}$ denote the distribution of σ'_{j-1}. Then, by Lemma 5.3.5 applied to the graph G_j and the inequality in (5.31) we have that

$$\left\|\nu_{\sigma'_{j-1}} - \mu_{G_{j-1}}\right\|_{\mathrm{TV}} \leq \frac{2}{\beta}\mathbf{E}_j[\mathsf{Corr}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j)] \leq \frac{4}{\beta^{\Delta+1}}\varepsilon_j. \tag{5.33}$$

Since $\hat{\sigma}_j$ is distributed according to μ_{G_j}, by the Coupling Lemma [1], there exists a coupling $\Pr[\cdot]$ of σ_j and $\hat{\sigma}_j$ such that

$$\Pr[\sigma_j \neq \hat{\sigma}_j] = \left\|\nu_{\sigma_j} - \mu_{G_j}\right\|_{\mathrm{TV}}. \tag{5.34}$$

Note also that, for a configuration $\eta \notin \Omega_j(b, M)$, conditioned on $\sigma_j = \eta$ and $\hat{\sigma}_j = \eta$, we can couple σ_{j-1} and σ'_{j-1} so that $\sigma_{j-1} \neq \sigma'_{j-1}$ with probability at most $1/n^{10}$. To see this, if $\eta(u_j) = \eta(v_j)$ then we trivially have $\sigma'_{j-1} = \sigma_{j-1} = \eta$. Otherwise, σ_{j-1} and σ'_{j-1} are produced by first choosing a random bichromatic class containing u_j and v_j under η, which we can couple so that it is the same in both $\textsc{ReSample}_{b,M}(G_j, u_j, v_j, \eta)$ and $\textsc{IdealReSample}(G_j, u_j, v_j, \eta)$. Denote this class by U and let $c_1 = \eta(u_j)$, $c_2 = \eta(v_j)$. Then, from the definition of $\textsc{IdealReSample}(G_j, u_j, v_j, \eta)$ we have that the distribution of $\sigma'_{j-1}(U)$ is given by $\pi^{c_1,c_2}_{G_j[U],u_j,v_j}$, while from $\textsc{ReSample}_{b,M}(G_j, u_j, v_j, \eta)$ we have that the distribution of $\sigma_{j-1}(U)$ is $1/n^{10}$-close to $\pi^{c_1,c_2}_{G_j[U],u_j,v_j}$. We therefore have that

$$\Pr\left[\sigma_{j-1} \neq \sigma'_{j-1} \mid \sigma_j = \hat{\sigma}_j \notin \Omega_j(b, M)\right] \leq 1/n^{10}. \tag{5.35}$$

Invoking the Coupling Lemma [1] again, we therefore obtain that

$$\begin{aligned}\left\|\nu_{\sigma_{j-1}} - \nu_{\sigma'_{j-1}}\right\|_{\mathrm{TV}} &\leq \Pr[\sigma_j \neq \hat{\sigma}_j] + \Pr[\sigma_j = \hat{\sigma}_j \in \Omega_j(b, M)] \\ &\qquad + \Pr\left[\sigma_{j-1} \neq \sigma'_{j-1} \mid \sigma_j = \hat{\sigma}_j \notin \Omega_j(b, M)\right] \\ &\leq \left\|\nu_{\sigma_j} - \mu_{G_j}\right\|_{\mathrm{TV}} + \mu_{G_j}(\Omega_j(b, M)) + 1/n^{10}, \end{aligned} \tag{5.36}$$

where in the last inequality we used (5.34), (5.35) and

$$\Pr[\sigma_j = \hat{\sigma}_j \in \Omega_j(b, M)] \leq \Pr[\hat{\sigma}_j \in \Omega_j(b, M)] = \mu_{G_j}(\Omega_j(b, M)).$$

Combining (5.36) with the inductive hypothesis (5.32) and (5.30), we obtain

$$\left\|\nu_{\sigma_{j-1}} - \nu_{\sigma'_{j-1}}\right\|_{\mathrm{TV}} \leq \frac{6(t-j+1)}{n^{10}} + \frac{4}{\beta^{\Delta+1}} \sum_{j'=j+1}^{t} \varepsilon_{j'}.$$

Using this and (5.33) completes (via the triangle inequality) the inductive step, therefore completing the proof of (5.32) for all $j \in \{0, 1, \ldots, t\}$ as wanted.

We have $G_0 = G$ and $\sigma_0 = \sigma$, so (5.32) for $j = 0$ gives

$$\left\|\nu_\sigma - \mu_G\right\|_{\mathrm{TV}} \leq \frac{6t}{n^{10}} + \frac{4}{\beta^{\Delta+1}} \sum_{j=1}^{t} \varepsilon_j \leq \frac{1}{n^8} + \frac{4}{\beta^{\Delta+1}} \sum_{j=1}^{t} \sum_{\ell=L_0}^{L} P_{\ell,j} \Big(K \frac{1-\beta}{1+\beta}\Big)^{\ell},$$

where the last inequality holds for all sufficiently large n using the values of ε_j from (5.29) (and the crude bound $t \leq |E| \leq \Delta n$). Using (5.28), we therefore have that

$$\left\|\nu_\sigma - \mu_G\right\|_{\mathrm{TV}} \leq \frac{1}{n^8} + \frac{4}{\beta^{\Delta+1}} \sum_{\ell=L_0}^{L} \ell C_\ell \Big(K \frac{1-\beta}{1+\beta}\Big)^{\ell} \leq \frac{10}{\beta^{\Delta+1} n^{\delta}},$$

where the last inequality follows from our assumption that G satisfies Item 1 (and by assuming that n is sufficiently large). This completes the proof of Theorem 5.3.11, modulo the proofs of (5.30) and (5.31) which are given below.

To prove (5.30) and (5.31), fix any value $j \in \{1, \ldots, t\}$. Since G_j is a subgraph of G and G satisfies Item 2 every path P in G_j with ℓ vertices where $\ell \in [\ell_1, L]$ has an h-graph-neighbourhood with at least $(1-\varepsilon)\ell$ isolated tree components. By Lemma 5.3.10 which, recall, applies for all $\ell \geq \ell'$ (where ℓ' is the constant specified in the beginning of the proof), we therefore have that

$$\mu_{G_j}(P \text{ is bichromatic}) \leq K^{\ell} \text{ for any path } P \text{ in } G_j \text{ with } \ell \text{ vertices, } \ell \in [\max\{\ell_1, \ell'\}, L]. \tag{5.37}$$

We are now ready to prove (5.30). For a vertex v, let X_v be the r.v. that counts the number of bichromatic paths with L vertices that start from v in a random configuration $\sigma \sim \mu_{G_j}$. Since G_j has maximum degree $\leq \Delta$, there are at most $\Delta(\Delta-1)^{L-2}$ paths starting from v with L vertices and each of them is bichromatic with probability at most K^L from (5.37). Hence, the expectation of X_v is at most

$$\Delta(\Delta-1)^{L-2} K^L \leq (b')^L,$$

where the last inequality follows from the choice of M (cf. (5.27)). By Markov's inequality we therefore have that for all $v \in V$ it holds that

$$\mu_{G_j}(X_v > b^L) \leq (b'/b)^L \leq 1/n^{11},$$

where the last inequality also follows from the choice of M (cf. (5.27)). Note that for (b, M)-bad configurations, i.e., configurations in $\Omega_j(b, M)$, at least one of the events $X_v > b^L$ occurs for some $v \in V$, and therefore by a union bound over $v \in V$ we have that

$$\mu_{G_j}\big(\Omega_j(b, M)\big) \leq \sum_{v \in V} \mu_{G_j}\big(X_v > b^L\big) \leq 1/n^{10}.$$

This finishes the proof of (5.30).

To prove (5.31), for a set $U \subseteq V$ such that $u_j, v_j \in U$, let

$$\widehat{\mathsf{Corr}}_{G_j}(U, u_j, v_j) := \big|\pi_{G_j[U]}(\eta(u_j) = 1 \mid \eta(v_j) = 1) - \pi_{G_j[U]}(\eta(u_j) = 1 \mid \eta(v_j) = 2)\big|$$

Note, by the symmetry of the states in the Ising model, we have that $\pi_{G[U]}\big(\tau(v_j) = 1\big) = \pi_{G[U]}\big(\tau(v_j) = 2\big) = 1/2$ and therefore

$$\mathsf{Corr}_{G_j}(U, u_j, v_j) = \left|\frac{\pi_{G[U]}\big(\tau(u_j) = 1, \tau(v_j) = 1\big)}{\pi_{G[U]}\big(\tau(u_j) = 1, \tau(v_j) = 2\big)} - 1\right| = \left|\frac{\pi_{G[U]}\big(\tau(u_j) = 1 \mid \tau(v_j) = 1\big)}{\pi_{G[U]}\big(\tau(u_j) = 1 \mid \tau(v_j) = 2\big)} - 1\right|.$$

By Lemma 5.3.8 and since G_j has maximum degree $\leq \Delta$, we obtain that

$$\mathsf{Corr}_{G_j}(U, u_j, v_j) \leq \frac{2}{\beta^\Delta} \widehat{\mathsf{Corr}}_{G_j}(U, u_j, v_j),$$

and therefore, to prove (5.31) we only need to show that

$$\mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j)\big] \leq \varepsilon_j, \tag{5.38}$$

For convenience, let $\mathcal{F}_j$ be the event that the Potts configuration $\hat{\sigma}_j$ is (b, M)-bad, i.e., $\mathcal{F}_j = \{\hat{\sigma}_j \in \Omega_j(b, M)\}$, and denote by $\overline{\mathcal{F}_j}$ the event that $\hat{\sigma}_j \notin \Omega_j(b, M)$.

$$\begin{aligned}\mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j)\big] &= \mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \mid \mathcal{F}_j\big]\mu_j(\mathcal{F}_j) + \mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \mid \overline{\mathcal{F}_j}\big]\mu_j(\overline{\mathcal{F}_j}) \\ &\leq \frac{2}{n^{10}} + \mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \mid \overline{\mathcal{F}_j}\big], \end{aligned} \tag{5.39}$$

where the inequality follows from the bounds $\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \leq 2$ and $\mu_j\big(\mathcal{F}_j\big) \leq 1/n^{10}$ from (5.30). Recall that $\mathcal{P}_{\ell,j}$ denotes the set of paths of length ℓ that connect u_j and v_j in G_j. Consider also the r.v.

$$Y_j = \frac{1}{n^{10}} + \sum_{\ell=L_0}^{L} \sum_{P \in \mathcal{P}_{\ell,j}} \mathbf{1}\{P \text{ is bichromatic}\} \times \Big(\frac{1-\beta}{1+\beta}\Big)^\ell.$$

Recall that $U_{\hat{\sigma}_j}$ is a bichromatic class under $\hat{\sigma}_j$ which contains u_j and v_j (chosen uniformly at random among the set of all such classes if there is more than one). Conditioned on the event $\hat{\sigma}_j \notin \Omega_j(b, M)$, we have that every bichromatic component under $\hat{\sigma}_j$ has average growth b up to depth $L = \lceil M \log n \rceil$ and therefore, irrespective of the random choice of $U_{\hat{\sigma}_j}$, we obtain by Lemma 5.3.7 that $\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \leq Y_j$. Note, in applying Lemma 5.3.7, we used that $P_{\ell,j} = 0$ for all $\ell \in [1, L_0]$; this holds because the edge $\{u_j, v_j\}$ does not belong to a short cycle in G (and therefore in G_{j-1} as well). We thus have that

$$\mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \mid \overline{\mathcal{F}_j}\big] \leq \mathbf{E}_j\big[Y_j \mid \overline{\mathcal{F}_j}\big] \leq \frac{\mathbf{E}_j[Y_j]}{\mu_{G_j}(\overline{\mathcal{F}_j})} \leq \Big(1 + \frac{2}{n^{10}}\Big)\mathbf{E}_j[Y_j], \tag{5.40}$$

where the last inequality follows from $1/\mu_{G_j}(\overline{\mathcal{F}_j}) \leq 1/(1-1/n^{10}) \leq 1+2/n^{10}$ (cf. (5.30)). Using (5.37) and the fact that $L_0 \geq \max\{\ell_1, \ell'\}$ for all sufficiently large n, we have that

$$\mathbf{E}_j[Y_j] \leq \frac{1}{n^{10}} + \sum_{\ell=L_0}^{L} P_{\ell,j}\Big(K\frac{|\beta-1|}{\beta+1}\Big)^{\ell}. \tag{5.41}$$

Since G satisfies Item 1, we have that the sum in (5.41) is less than $1/n^{\delta}$ and hence (5.40) and (5.41) give that for all sufficiently large n it holds that

$$\mathbf{E}_j\big[\widehat{\mathsf{Corr}}_{G_j}(U_{\hat{\sigma}_j}, u_j, v_j) \mid \overline{\mathcal{F}_j}\big] \leq \frac{1}{n^{10}} + \sum_{\ell=L_0}^{L} P_{\ell,j}\Big(K\frac{|\beta-1|}{\beta+1}\Big)^{\ell} + \frac{2}{n^{10}}\Big(\frac{1}{n^{10}} + \frac{1}{n^{\delta}}\Big).$$

Combining this with (5.39) yields (5.38), therefore concluding the proof of (5.31). This finishes the proof of Theorem 5.3.11. □

5.4 Analysing Potts on graphs with tree-like structure

5.4.1 Analysing Potts on trees

Fix an integer $\Delta \geq 3$. Note that use $\mathbb{T}_\Delta$ denote the infinite $(\Delta-1)$-ary tree with root vertex ρ. For integer $h \geq 0$, let $T_h = (V_h, E_h)$ denote the subtree of $\mathbb{T}_\Delta$ induced by the vertices at distance $\leq h$ from ρ and let L_h denote the leaves of T_h.

Recall that uniqueness on $\mathbb{T}_\Delta$ implies that root-to-leaves correlations on T_h tend to 0 as $h \to \infty$, cf. Definition 1.2.2. The following lemma extends these decay properties to arbitrary subtrees of T_h; this was proved in [29] in the case of the colourings model and the proof for the Potts model is analogous.

Lemma 5.4.1. *Let $\Delta, q \geq 3$ be integers, and $\beta > 0$ be in the uniqueness regime of the $(\Delta-1)$-ary tree. There exists a function $\vartheta : \mathbb{N} \to \mathbb{R}_{\geq 0}$ with $\vartheta(h) \to 0$ as $h \to \infty$ such that the following holds for any integer $h \geq 0$.*
Let T'_h be an arbitrary subtree of T_h containing the root ρ and let L'_h be the set of vertices in T'_h at distance exactly h from ρ. Then, for any colour $c \in [q]$ and any configuration $\tau : L'_h \to [q]$, it holds that

$$\left|\mu_{T'_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) - \frac{1}{q}\right| \leq \vartheta(h).$$

Proof. We will show that the statement of the lemma holds with the function $\vartheta(\cdot)$ given by $\vartheta(0) = 1$ and

$$\vartheta(h) := \max_{c \in [q],\ \tau : L_h \to [q]} \left|\mu_{T_h}[\sigma(\rho) = c \mid \sigma(L_h) = \tau] - \frac{1}{q}\right|.$$

Note that, since β is assumed to be in the uniqueness regime of the $(\Delta-1)$-ary tree, we have by definition that $\vartheta(h) \to 0$ as $h \to \infty$.
We first show by induction on h that, for all $h \geq 0$, for all subtrees T'_h of T_h containing the root ρ, for all configurations $\tau : L'_h \to [q]$ and any colour $c \in [q]$, it holds that

$$\mu_{T'_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) = \mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau). \tag{5.42}$$

For $h = 0$ the result is trivial. Suppose that $h \geq 1$ and that the result holds for all integers less than h, we prove the result for h as well. Let $U = \{u_1, \ldots, u_{\Delta-1}\}$ be the children of ρ in T_h and let $U' \subseteq U$ be the children of ρ in T'_h. For a vertex $u \in U$, denote by $T_h(u)$ the subtree of T_h rooted at u and by $L_h(u)$ the vertices in L_h that belong to the tree $T_h(u)$. Similarly, for a vertex $u \in U'$, denote by $T'_h(u)$ the subtree of T'_h rooted at u and by $L'_h(u)$ the vertices in L'_h that belong to the tree $T'_h(u)$. Using standard tree recursions (see for example Lemma 3.1.3 in Chapter 3), we have that

$$\mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) = \frac{\prod_{u \in U}\left(1 - (1-\beta)\,\mu_{T_h(u)}(\sigma(u) = c \mid \sigma(L'_h(u)) = \tau(L'_h(u)))\right)}{\sum_{c'=1}^{q}\prod_{u \in U}\left(1 - (1-\beta)\,\mu_{T_h(u)}(\sigma(u) = c' \mid \sigma(L'_h(u)) = \tau(L'_h(u)))\right)} \tag{5.43}$$

and

$$\mu_{T'_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) = \frac{\prod_{u \in U'}\left(1 - (1-\beta)\,\mu_{T'_h(u)}(\sigma(u) = c \mid \sigma(L'_h(u)) = \tau(L'_h(u)))\right)}{\sum_{c'=1}^{q}\prod_{u \in U'}\left(1 - (1-\beta)\,\mu_{T'_h(u)}(\sigma(u) = c' \mid \sigma(L'_h(u)) = \tau(L'_h(u)))\right)} \tag{5.44}$$

For every $u \in U \setminus U'$ we have that $L'_h(u') = \emptyset$ and therefore, using the symmetry among the colours, we have that for every $c' \in [q]$ it holds that

$$\mu_{T_h(u)}(\sigma(u) = c' \mid \sigma(L'_h(u)) = \tau(L'_h(u))) = \mu_{T_h(u)}(\sigma(u) = c') = 1/q \tag{5.45}$$

For $u \in U'$, we have that $T_h(u)$ is isomorphic to T_{h-1} and $T'_h(u)$ is a subtree of $T_h(u)$, so by the induction hypothesis we have that for any colour $c \in [q]$ it holds that

$$\mu_{T'_h(u)}(\sigma(u) = c' \mid \sigma(L'_h(u)) = \tau(L'_h(u))) = \mu_{T_h(u)}(\sigma(u) = c' \mid \sigma(L_h(u)) = \tau(L_h(u))). \tag{5.46}$$

Combining (5.43), (5.44), (5.45) and (5.46) yields (5.42), thus completing the induction. To complete the proof it remains to observe that for any colour $c \in [q]$, it holds that

$$\mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) = \sum_{\eta : L_h \setminus L'_h \to [q]} \mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau, \sigma(L_h \setminus L'_h) = \eta) \cdot \mu_{T_h}(\sigma(L_h \setminus L'_h) = \eta \mid \sigma(L'_h) = \tau)$$

Note that, by the definition of the function $\vartheta(\cdot)$, for any $\eta : L_h \setminus L'_h \to [q]$ we have that

$$\left|\mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau, \sigma(L_h \setminus L'_h) = \eta) - \frac{1}{q}\right| \leq \vartheta(h)$$

and $\sum_{\eta : L_h \setminus L'_h \to [q]} \mu_{T_h}(\sigma(L_h \setminus L'_h) = \eta \mid \sigma(L'_h) = \tau) = 1$, which gives that

$$\left|\mu_{T_h}(\sigma(\rho) = c \mid \sigma(L'_h) = \tau) - \frac{1}{q}\right| \leq \vartheta(h),$$

Combining this with (5.42) yields the lemma. □

5.4.2 Analysing Potts on disjoint trees whose roots are connected via a path

To prove Lemma 5.3.10, we will need the following lemmas.

Lemma 5.4.2. *Let $q \geq 3$ and $\beta \in (0,1)$, and set $\chi := \frac{1+\beta}{\beta+q-1}$. Let P be a path with $\ell \geq 2$ vertices and endpoints u, v. Then, for arbitrary colours $c, c' \in [q]$, it holds that*

$$\mu_P(P \text{ is bichromatic} \mid \sigma(u) = c, \sigma(v) = c') \leq (4q/\beta)\chi^{\ell-2}.$$

Proof. Let z be the neighbour of v (note that if $\ell = 2$ then $z = u$). We have that

$$\mu_P(\sigma(v) = c' \mid \sigma(u) = c) = \sum_{c'' \in [q]} \mu_P(\sigma(v) = c' \mid \sigma(z) = c'', \sigma(u) = c) \times \mu_P(\sigma(z) = c'' \mid \sigma(u) = c). \tag{5.47}$$

Note also that for arbitrary colours $c, c', c'' \in [q]$ we have that

$$\mu_P(\sigma(v) = c' \mid \sigma(z) = c'', \sigma(u) = c) = \mu_P(\sigma(v) = c' \mid \sigma(z) = c'') \geq \frac{\beta^{\mathbf{1}\{c'=c''\}}}{\beta + q - 1} \geq \beta/(\beta+q-1).$$

Combining this with (5.47), we obtain that $\mu_P(\sigma(v) = c' \mid \sigma(u) = c) \geq \beta/(\beta + q - 1)$ and hence

$$\mu_P(\sigma(u) = c, \sigma(v) = c') = \frac{1}{q}\mu_P(\sigma(v) = c' \mid \sigma(u) = c) \geq \frac{\beta}{q(\beta + q - 1)}. \tag{5.48}$$

Let $c_1, c_2 \in [q]$ be distinct colours in $[q]$ and $\mathcal{E}(c_1, c_2)$ be the event that every vertex in P is coloured with c_1, c_2. The lemma will follow by showing that

$$\mu_P(\mathcal{E}(c_1, c_2)) = 2\chi^{\ell-1}/q. \tag{5.49}$$

Let us briefly conclude the lemma assuming (5.49). Indeed, if $c \neq c'$, then applying (5.49) for $c_1 = c$ and $c_2 = c'$ and using the lower bound in (5.48) we obtain that

$$\mu_P(P \text{ is bichromatic} \mid \sigma(u) = c, \sigma(v) = c') \leq (4/\beta)\chi^{\ell-2},$$

while, if $c = c'$, we obtain by summing (5.49) for $c_1 = c$ and the $q - 1$ possible values of c_2 that

$$\mu_P(P \text{ is bichromatic} \mid \sigma(u) = c, \sigma(v) = c') \leq (4q/\beta)\chi^{\ell-2}.$$

It remains to prove (5.49). For convenience, denote by $w_1, \ldots, w_\ell$ the vertices of P in order so that $u = w_1$ and $v = w_\ell$ and let $C = \{c_1, c_2\}$. Note that

$$\mu_P(\mathcal{E}(c_1, c_2)) = \mu_P(\sigma(w_1) \in C) \prod_{i=2}^{\ell} \mu_P\big(\sigma(w_i) \in C \mid \sigma(w_{i-1}) \in C, \ldots, \sigma(w_1) \in C\big).$$

We have $\mu_P(\sigma(w_1) \in C) = 2/q$. For $i = 2, \ldots, n$, let P_i be the path induced by the vertices $w_{i-1}, \ldots, w_n$. We have that

$$\mu_P\big(\sigma(w_i) \in C \mid \sigma(w_{i-1}) \in C, \ldots, \sigma(w_1) \in C\big) = \mu_{P_i}\big(\sigma(w_i) \in C \mid \sigma(w_{i-1}) \in C\big).$$

Since $\mu_{P_i}(\sigma(w_i) \in C) = \mu_{P_i}(\sigma(w_{i-1}) \in C) = 2/q$, we have by the Bayes' rule that

$$\mu_{P_i}(\sigma(w_i) \in C \mid \sigma(w_{i-1}) \in C) = \mu_{P_i}(\sigma(w_{i-1}) \in C \mid \sigma(w_i) \in C) = \frac{1 + \beta}{\beta + q - 1} = \chi.$$

Combining these, we obtain (5.49), therefore concluding the proof of Lemma 5.4.2. □

We will use the following corollary of Lemma 5.4.2.

Corollary 5.4.3. *Let $q \geq 3$ and $\beta \in (0,1)$, and set $\chi := \frac{1+\beta}{\beta+q-1}$. Let P be a path with $\ell \geq 2$ vertices and let Λ be a subset of the vertices which includes the endpoints of the path. Then, for any configuration $\tau : \Lambda \to [q]$, it holds that*

$$\mu_P(P \text{ is bichromatic} \mid \sigma(\Lambda) = \tau) \leq (4q/\beta)^{|\Lambda|}\chi^{\ell-2|\Lambda|}.$$

Proof. Let $t := |\Lambda|$ and denote the set of vertices in Λ by $u_1, \ldots, u_t$ in the order that they appear in the path. Note that u_1 and u_t are the endpoints of P (since by assumption Λ includes the endpoints of P). For $i = 1, 2, \ldots, t-1$, let P_i be the path induced by the vertices between u_i and u_{i+1} and let ℓ_i be the number of vertices in P_i. Then, we have

$$\begin{aligned}\mu_P(P \text{ is bichromatic} \mid \sigma(\Lambda) = \tau) &\leq \mu_P(P_i \text{ is bichromatic for } i = 1, \ldots, t-1 \mid \sigma(\Lambda) = \tau) \\ &= \prod_{i=1}^{t-1} \mu_{P_i}(P_i \text{ is bichromatic} \mid \sigma(u_i) = \tau(u_i), \sigma(u_{i+1}) = \tau(u_{i+1})).\end{aligned}$$

By Lemma 5.4.2, we have that

$$\mu_{P_i}(P_i \text{ is bichromatic} \mid \sigma(u_i) = \tau(u_i), \sigma(u_{i+1}) = \tau(u_{i+1})) \leq (4q/\beta)\chi^{\ell_i - 2}$$

and therefore, since $\sum_i^t \ell_i = \ell$ and $\chi, \beta \in (0,1)$, we obtain that

$$\mu_P(P \text{ is bichromatic} \mid \sigma(\Lambda) = \tau) \leq (4q/\beta)^{|\Lambda|-1}\chi^{\ell-2(|\Lambda|-1)} \leq (4q/\beta)^{|\Lambda|}\chi^{\ell-2|\Lambda|}.$$

This finishes the proof. □

We are now ready to prove Lemma 5.3.10, which we restate here for convenience.

Lemma 5.3.10. *Let $\Delta, q \geq 3$ be integers, and $\beta \in (0,1)$ be in the uniqueness regime of the $(\Delta-1)$-ary tree with $\beta \neq (\Delta - q)/\Delta$. Then, for any $\varepsilon' > 0$, there exists a positive constant $K < \frac{1+\beta}{\beta+q-1} + \varepsilon'$ and $\varepsilon > 0$ such that the following holds for all sufficiently large integers ℓ and h.*
Let G be a graph of maximum degree Δ and P be a path with ℓ vertices whose h-graph-neighbourhood contains $(1-\varepsilon)\ell$ isolated tree components. Let μ_G be the Potts measure on G with parameter β. Then,

$$\mu_G(\text{path } P \text{ is bichromatic}) \leq K^{\ell}.$$

Proof. Let $\chi := \frac{1+\beta}{\beta+q-1}$ and consider arbitrary $\varepsilon' > 0$. We begin by specifying the constants K, ε and how large ℓ and h need to be. In particular, let K be any constant satisfying $\chi < K < \chi + \varepsilon'$ and $\varepsilon > 0$ be a small constant such that

$$(4q/\chi\beta)^{2\varepsilon}\Big(\frac{1/q+\varepsilon}{1/q-\varepsilon}\chi\Big)^{1-2\varepsilon} \leq K/(1+\varepsilon). \tag{5.50}$$

Note that such an ε exists by considering the limit $\varepsilon \to 0$. Let ℓ be sufficiently large so that

$$\varepsilon\ell \geq 2 \quad \text{and} \quad (1+\varepsilon)^{\ell} \geq q^2. \tag{5.51}$$

Finally, let $\vartheta(\cdot)$ be the function in Lemma 5.4.1, so that for all sufficiently large h it holds that

$$\vartheta(h) \leq \varepsilon. \tag{5.52}$$

We are now ready to proceed to the proof of the lemma. Let G be a graph and consider a path P in G, with vertices $u_1, \ldots, u_\ell$, whose h-graph neighbourhood contains at least $(1-\varepsilon)\ell$ isolated tree components. By definition, an isolated tree component contains exactly one of the vertices $u_1, \ldots, u_\ell$ and therefore the set

$$U = \big\{u_i \,\big|\, u_i \text{ belongs to an isolated tree component and } u_i \neq u_1, u_\ell\big\},$$

satisfies $|U| \geq (1-\varepsilon)\ell - 2 \geq (1-2\varepsilon)\ell$. Note that we exclude the endpoints of the path P from U, even if they belong to isolated tree components. Let also $\overline{U}$ denote the set $\{u_1, \ldots, u_\ell\} \setminus U$, so that $|\overline{U}| \leq 2\varepsilon\ell$. Recall that T_h is the subtree of the $(\Delta-1)$-ary tree consisting of vertices at distance at most h from the root. By definition of the set U, we have that

$$\text{for } u \in U,\ \Gamma_h(G \setminus P, u) \text{ induces a subgraph in } G \text{ which is a subtree of } T_h. \tag{5.53}$$

For $u \in U$, let for convenience $V_u = \Gamma_h(G \setminus P, u)$, F_u be the subgraph of G induced on V_u and L_u be the set of vertices that are at distance h from u in F_u. Note that F_u is a tree rooted at u all of whose vertices are at distance at most h from u; L_u is thus the set of all leaves in F_u which are at distance h from u (note that there could be other leaves which are closer to u but these will not matter).

Let F be the union of the subgraphs F_u for $u \in U$; note that F is a disjoint union of copies of trees (each of which is a subtree of T_h from (5.53)). Let also $L := \bigcup_{u\in U} L_u$. We will also denote by V_P, E_P the vertex and edge set of the path P. Note that E_F is disjoint

from E_P (though V_P and V_F intersect at V_P). Finally, let $H = (V_H, E_H)$ be the subgraph of G with vertex set $V_P \cup V_F$ and edge set $E_P \cup E_F$.
To bound the probability that P is bichromatic, we will condition on a worst case boundary configuration on L and $\overline{U}$. In particular, we have the bound

$$\mu_G(P \text{ is bichromatic}) \leq \max_{\tau: L \to [q], \tau': \overline{U} \to [q]} \mu_G(P \text{ is bichromatic} \mid \sigma(L) = \tau,\, \sigma(\overline{U}) = \tau').$$

To bound the r.h.s., fix arbitrary configurations $\tau : L \to [q], \tau' : \overline{U} \to [q]$ and note that

$$\mu_G(P \text{ is bichromatic} \mid \sigma(L) = \tau,\, \sigma(\overline{U}) = \tau') = \mu_H(P \text{ is bichromatic} \mid \sigma(L) = \tau,\, \sigma(\overline{U}) = \tau'),$$

so the lemma will follow (since τ, τ' are arbitrary) by showing that, for arbitrary colours $c_1, c_2 \in [q]$, it holds that

$$\mu_H\big(\mathcal{E}(c_1, c_2) \mid \sigma(L) = \tau,\, \sigma(\overline{U}) = \tau'\big) \leq K^\ell / q^2, \tag{5.54}$$

where $\mathcal{E}(c_1, c_2)$ is the event that each vertex in P is coloured with either c_1 or c_2. Since $\overline{U}$ includes the endpoints of the path P, by Corollary 5.4.3 we have that

$$\mu_P(\mathcal{E}(c_1, c_2) \mid \sigma(\overline{U}) = \tau') \leq (4q/\beta)^{|\overline{U}|} \chi^{\ell - 2|\overline{U}|} = (4q/\chi\beta)^{|\overline{U}|} \chi^{|U|}. \tag{5.55}$$

Let Z_P be the partition function of the path P and $Z_P(c_1, c_2)$ be the contribution to Z_P from configurations such that P is coloured with c_1 or c_2, i.e.,

$$Z_P = \sum_{\sigma: V_P \to [q];\ \sigma(\overline{U}) = \tau'} w_P(\sigma), \qquad Z_P(c_1, c_2) = \sum_{\sigma: V_P \to \{c_1, c_2\};\ \sigma(\overline{U}) = \tau'} w_P(\sigma),$$

Then, (5.55) translates into

$$Z_P(c_1, c_2)/Z_P \leq (4q/\chi\beta)^{|\overline{U}|} \chi^{|U|}.$$

The trees hanging from vertices in U reweigh the probability that the path is bichromatic but do not cause significant distortion since the states of the roots are roughly uniformly distributed (because of uniqueness on the tree). To quantify this, for $u \in U$ and a colour $c \in [q]$, let

$$Z_u(c) = \sum_{\substack{\sigma: V_u \to [q]; \\ \sigma(L_u) = \tau(L_u),\ \sigma(u) = c}} w_{F_u}(\sigma).$$

Let also $Z_u = \sum_{c\in[q]} Z_u(c)$. By Lemma 5.4.1 and the choice of h in (5.52), we have that for all $c \in [q]$ it holds that

$$\Big|\frac{Z_u(c)}{Z_u} - \frac{1}{q}\Big| \le \vartheta(h) \le \varepsilon. \tag{5.56}$$

We can now write $\mu_H\big(\mathcal{E}(c_1,c_2) \mid \sigma(L)=\tau,\, \sigma(\overline{U})=\tau'\big)$ as

$$\mu_H\big(\mathcal{E}(c_1,c_2) \mid \sigma(L)=\tau,\, \sigma(\overline{U})=\tau'\big) = \frac{\sum_{\sigma:V_P\to\{c_1,c_2\};\ \sigma(\overline{U})=\tau'} w_P(\sigma)\prod_{u\in U} Z_u(\sigma(u))}{\sum_{\sigma:V_P\to[q];\ \sigma(\overline{U})=\tau'} w_P(\sigma)\prod_{u\in U} Z_u(\sigma(u))}.$$

Dividing both numerator and denominator by $\prod_{u\in U} Z_u$, we obtain using (5.56) that

$$\mu_H\big(\mathcal{E}(c_1,c_2) \mid \sigma(L)=\tau,\, \sigma(\overline{U})=\tau'\big) \le \Big(\frac{1/q+\varepsilon}{1/q-\varepsilon}\Big)^{|U|}\frac{Z_P(c_1,c_2)}{Z_P} \le (4q/\chi\beta)^{|\overline{U}|}\Big(\frac{1/q+\varepsilon}{1/q-\varepsilon}\chi\Big)^{|U|}.$$

We have $|U| \ge (1-2\varepsilon)\ell$ and $|\overline{U}| \le 2\varepsilon\ell$, so from the choice of ε and ℓ in (5.50) and (5.51), we obtain that

$$\mu_H\big(\mathcal{E}(c_1,c_2) \mid \sigma(L)=\tau,\, \sigma(\overline{U})=\tau'\big) \le K^\ell/(1+\varepsilon)^\ell \le K^\ell/q^2,$$

which is exactly (5.54), as wanted. This concludes the proof of Lemma 5.3.10. □

5.5 Correlation decay and sampling for antiferromagnetic Ising

In this section, we prove Theorem 5.3.6 and Lemma 5.3.7. The proofs follow relatively easily from correlation decay bounds on trees appearing in [64]; the bounds in there are stated for the case of the ferromagnetic Ising model, but there is a simple translation of these bounds to the antiferromagnetic case which allows us to conclude the desired results. In Section 5.5.1, we import the results from the literature that we need. We then give the proof of Theorem 5.3.6 in Section 5.5.2 and the proof of Lemma 5.3.7 in Section 5.5.3.

5.5.1 Preliminaries

Following [64], for a graph G and a vertex u in G, we consider the self-avoiding walk tree $T = T_{\text{SAW}}(G,u)$, which consists of all paths starting from u and not intersecting themselves, except possibly at the terminal vertex of the path; note that each vertex in T corresponds to a vertex in G in the natural way (in a many-to-one fashion). The following lemma originates in the work of Weitz for the hard-core model [78]; it is well-known that the lemma holds more generally for any 2-state system. The particular version we state here is close to [64, Lemma 13].

Lemma 5.5.1 (see, e.g., [64, Lemma 13]). *Let $G = (V, E)$ be a graph and u be a vertex in G. Consider the self-avoiding walk tree $T = T_{\mathrm{SAW}}(G, u)$ starting from u and denote by A the leaves of the tree. Then, there is a configuration $\eta : A \to \{1, 2\}$ (described in [78]) such that the following holds for any set $\Lambda \subseteq V$ and any configuration $\tau : \Lambda \to \{1, 2\}$.*
Let U_Λ be the set of vertices in T which correspond to vertices in Λ and let $\eta' : U_\Lambda \setminus A \to \{1, 2\}$ be the configuration where each vertex in $U_\Lambda \setminus A$ inherits the state of the corresponding vertex in Λ under τ. Then,

$$\pi_G(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau) = \pi_T(\sigma(u) = 1 \mid \sigma(A) = \eta, \sigma(U_\Lambda \setminus A) = \eta').$$

The following lemma follows from a strong spatial mixing result in [64] for the ferromagnetic Ising model, which in turn builds upon a lemma from [5].

Lemma 5.5.2. *Let $\beta \in (0, 1)$. Let $T = (V, E)$ be a tree, $\Lambda \subseteq V$ be a subset of the vertices and u be an arbitrary vertex. Let $\tau_1, \tau_2 : \Lambda \to \{1, 2\}$ be two configurations on Λ which differ only on a subset $U \subseteq \Lambda$. Then,*

$$\left|\pi_T(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau_1) - \pi_T(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau_2)\right| \leq \sum_{v \in U} \left(\frac{1-\beta}{1+\beta}\right)^{\mathrm{dist}(u,v)},$$

where $\mathrm{dist}(u, v)$ *denotes the distance between u and v in T.*

Proof. It is well-known that, on bipartite graphs $G = (V, E)$, there is a measure-preserving bijection between configurations of the antiferromagnetic Ising model with parameter $\beta \in (0, 1)$ and configurations of the ferromagnetic Ising model with parameter $1/\beta$, obtained by flipping the states of each vertex on one side of the bipartition. The desired inequality therefore follows from the strong spatial mixing result for the ferromagnetic Ising model given in [64, Lemma 14]. To translate the parametrisation of that result, note that in [64] the weight of a (ferromagnetic) Ising configuration σ is parametrised to be proportional to $\exp(2\beta m(\sigma))$ – therefore $1/\beta = e^{2\beta}$, so $\tanh \beta = \frac{1-\beta}{1+\beta}$. Using this translation, we obtain the desired inequality. □

5.5.2 Proof of Theorem 5.3.6

To prove Theorem 5.3.6, we will use the following algorithm that allows us to compute conditional marginal probabilities with very small absolute error.

Lemma 5.5.3. *Let $\beta \in (0,1)$ and $b > 0$ be constants such that $b\frac{1-\beta}{1+\beta} < 1$, and let $\Delta \geq 3$ be an integer. Then, there exists $M_0 > 0$ such that the following holds for all $M > M_0$.*
There is a polynomial-time algorithm that, on input: (i) an n-vertex graph $G = (V, E)$ with maximum degree at most Δ and average growth b up to depth $L = \lceil M \log n \rceil$, (ii) a subset $\Lambda \subseteq V$ with a configuration $\tau : \Lambda \to \{1, 2\}$, and (iii) a vertex $u \in V \setminus \Lambda$, outputs a number $\hat{p} \in [0, 1]$ such that

$$\left|\hat{p} - p\right| \leq \frac{\beta^{\Delta}}{2n^{11}(1+\beta^{\Delta})} \quad \text{where } p := \pi_G(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau),$$

i.e., $\hat{p}$ is within absolute error $O(1/n^{11})$ from the marginal probability that $\sigma(u) = 1$ conditioned on $\sigma(\Lambda) = \tau$, where σ is from the Ising distribution π_G on G with parameter β.

Proof. Let M_0 be a large constant so that for $n \geq 2$ and $L_0 := \lceil M_0 \log n \rceil$, it holds that

$$b^{L_0}\Big(\frac{1-\beta}{1+\beta}\Big)^{L_0} \leq \frac{\beta^{\Delta}}{2n^{11}(1+\beta^{\Delta})}. \tag{5.57}$$

Note that such an M_0 exists since $b\frac{1-\beta}{1+\beta} < 1$. Fix M to be an arbitrary constant larger than M_0.
Let G be an arbitrary n-vertex graph with average growth b up to depth $L = \lceil M \log n \rceil$, $\Lambda \subseteq V$ be a subset of the vertices, $\tau : \Lambda \to \{1, 2\}$ be a configuration on Λ, and $u \in V \setminus \Lambda$ be a vertex. Consider the self-avoiding walk tree $T = T_{\mathrm{SAW}}(G, u)$ starting from u, and denote by A the leaves of the tree. By Lemma 5.5.1, we have that there is a configuration $\eta : A \to \{1, 2\}$ such that

$$\pi_G(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau) = \pi_T(\sigma(u) = 1 \mid \sigma(A) = \eta, \sigma(U \setminus A) = \eta') \tag{5.58}$$

where $U := U_\Lambda$ is the set of vertices in T that correspond to some vertex in Λ and η' is the assignment on $U \setminus A$ inherited by τ (see Lemma 5.5.1 for details). For convenience, let $F := U \cup A$ and $\zeta : F \to \{1, 2\}$ be the configuration which agrees with η on A and with η' on $U \setminus A$, so that (5.58) can be rewritten as

$$\pi_G(\sigma(u) = 1 \mid \sigma(\Lambda) = \tau) = \pi_T(\sigma(u) = 1 \mid \sigma(F) = \zeta).$$

Let $T' = (V', E')$ be the subtree of T induced by vertices at distance at most $L - 1$ from u and let $F' = F \cap V'$. Let

$$p := \pi_T(\sigma(u) = 1 \mid \sigma(F) = \zeta) \quad \text{and} \quad \hat{p} := \pi_{T'}(\sigma(u) = 1 \mid \sigma(F') = \zeta(F')).$$

Note that $\hat{p}$ can be computed in polynomial time since the tree T' and the configuration $\zeta_{F'}$ can be constructed in polynomial time (since T' is a tree of size $O(\Delta^L)$). So, the lemma will follow by showing that

$$\left|\hat{p} - p\right| \leq \frac{\beta^\Delta}{2n^{11}(1+\beta^\Delta)}. \tag{5.59}$$

To prove this, let J be the set of vertices in $V' \setminus F'$ whose distance from u is exactly $L-1$ in T and note that $|J| \leq b^L$ since G has average growth b up to depth L. Note also that, conditioned on the configurations on J and F', the probability that $\sigma(u) = 1$ depends only on T' (and not on the configuration on rest of the tree T). Using the law of total probability, we can therefore expand p as

$$p = \sum_{\iota : J \to \{1,2\}} \pi_{T'}(\sigma(u) = 1 \mid \sigma(F') = \zeta(F'), \sigma(J) = \iota) \times \pi_T(\sigma(J) = \iota \mid \sigma(F) = \zeta(F)). \tag{5.60}$$

Therefore, (5.59) will follow by showing that, for any configuration $\iota : J \to \{1,2\}$ it holds that

$$\left|\pi_{T'}\big(\sigma(u) = 1 \mid \sigma(F') = \zeta(F'),\, \sigma(J) = \iota\big) - \hat{p}\right| \leq \frac{\beta^\Delta}{2n^{11}(1+\beta^\Delta)}. \tag{5.61}$$

Note that we can expand $\hat{p}$ analogously to (5.60) by conditioning on the configuration on J, so to prove (5.60) it suffices to show that for any two configurations $\iota_1, \iota_2 : J \to [q]$ it holds that

$$\kappa \leq \frac{\beta^\Delta}{2n^{11}(1+\beta^\Delta)}$$

where $\kappa := \left|\pi_{T'}\big(\sigma(u) = 1 \mid \sigma(F') = \zeta_{F'},\, \sigma(J) = \iota_1\big) - \pi_{T'}\big(\sigma(u) = 1 \mid \sigma(F') = \zeta_{F'},\, \sigma(J) = \iota_2\big)\right|$. By the strong spatial mixing result of Lemma 5.5.2, we have that

$$\kappa \leq \sum_{v \in J} \Big(\frac{1-\beta}{\beta+1}\Big)^{\mathrm{dist}(u,v)} = |J|\Big(\frac{1-\beta}{\beta+1}\Big)^L \leq b^L\Big(\frac{1-\beta}{1+\beta}\Big)^L.$$

Combining this with the choice of M_0 (cf. (5.57)), we obtain (5.61), thus concluding the proof of the lemma. □

Using Lemma 5.5.3, the proof of Theorem 5.3.6 follows by standard techniques.

Proof of Theorem 5.3.6. Denote by $v_1, v_2, \ldots, v_n$ the vertices of G. The algorithm will sample the state s_i of vertex v_i sequentially for $i = 1, \ldots, n$. We just give the details

for the first part of the algorithm, the proof of the second part is completely analogous (namely, it suffices to assume in the following that we first set $v_1 = u, v_2 = v$, we then fix the states $s_1 = 1, s_2 = 2$ and finally sample the states s_i for $i = 3, \dots, n$.)

Assume that, at some time $i = 1, \dots, n$, we have sampled the states $s_1, \dots, s_{i-1}$ (which can take arbitrary values in $\{1, 2\}$). Using the algorithm of Lemma 5.5.3, we obtain in polynomial time numbers $\hat{p}_i(1), \hat{p}_i(2) \in [0, 1]$ such that $\hat{p}_i(1) + \hat{p}_i(2) = 1$ and, for $s \in \{1, 2\}$, it holds that

$$\left|a_i(s) - \hat{p}_i(s)\right| \leq \frac{\beta^\Delta}{2n^{11}(1 + \beta^\Delta)}, \tag{5.62}$$

where $a_i(s) := \pi_G(\sigma(v_i) = s \mid \sigma(v_1) = s_1, \dots, \sigma(v_{i-1}) = s_{i-1})$. We then sample the state s_i by letting $s_i = 1$ with probability $\hat{p}_i(1)$, or else $s_i = 2$ (note that $s_i = 2$ with probability $\hat{p}_i(2)$). Denote by τ the final configuration and by ν_τ its distribution.

We will show that, for any configuration $\eta : V \to \{1, 2\}$, it holds that

$$|\nu_\tau(\eta) - \pi_G(\eta)| \leq \frac{2}{n^{10}} \pi_G(\eta), \tag{5.63}$$

so by summing over η we obtain

$$\left\|\nu_\tau - \pi_G\right\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\eta \colon V \to \{1,2\}} \left|\nu_\tau(\eta) - \pi_G(\eta)\right| \leq \frac{1}{n^{10}} \sum_{\eta \colon V \to \{1,2\}} \pi_G(\eta) = 1/n^{10},$$

which proves the theorem. To prove (5.63), fix an arbitrary configuration $\eta : V \to \{1, 2\}$ and let

$$p_{i,\eta} := \pi_G(\sigma(v_i) = \eta(v_i) \mid \sigma(v_1) = \eta(v_1), \dots, \sigma(v_{i-1}) = \eta(v_{i-1})).$$

Note that

$$\pi_G(\eta) = \prod_{i=1}^{n} p_{i,\eta} \quad \text{and} \quad \nu_\tau(\eta) = \prod_{i=1}^{n} \hat{p}_i(\eta(v_i)). \tag{5.64}$$

Moreover, from (5.62), we have that for all $i = 1, \dots, n$ it holds that

$$\left|p_{i,\eta} - \hat{p}_i(\eta(v_i))\right| \leq \frac{\beta^\Delta}{2n^{11}(1 + \beta^\Delta)} \leq \frac{p_{i,\eta}}{2n^{11}},$$

where the last inequality follows from the lower bound of Lemma 5.3.8. Using the inequalities $1 - x \leq \mathrm{e}^{-x} \leq 1 - x/2$ which hold for all $x \in [0, 1/2]$, we obtain that, for $\varepsilon = 1/n^{11}$, it holds that

$$\mathrm{e}^{-\varepsilon} p_{i,\eta} \leq \hat{p}_i(\eta(v_i)) \leq \mathrm{e}^{\varepsilon} p_{i,\eta}.$$

Multiplying over $i = 1, \dots, n$ and combining with (5.64) gives (5.63), thus completing the proof of Theorem 5.3.6. □

5.5.3 Proof of Lemma 5.3.7

In this section, we give the proof of Lemma 5.3.7, which we restate here for convenience.

Lemma 5.3.7. *Let $\beta \in (0,1)$ and $b > 0$ be constants such that $b\frac{1-\beta}{1+\beta} < 1$. Then, there exists $M_0' > 0$ such that the following holds for all $M > M_0'$.*
Let G be an n-vertex graph with average growth b up to depth $L = \lceil M \log n \rceil$, and let u, v be distinct vertices in G. Then

$$\Big|\pi_G(\sigma(u) = 1 \mid \sigma(v) = 1) - \pi_G(\sigma(u) = 1 \mid \sigma(v) = 2)\Big| \le \frac{1}{n^{10}} + \sum_{\ell=1}^{L} P_\ell(G, u, v)\Big(\frac{1-\beta}{1+\beta}\Big)^\ell$$

where $P_\ell(G, u, v)$ is the number of paths with ℓ vertices in G that connect u and v.

Proof of Lemma 5.3.7. Let M_0' be sufficiently large so that for all $n \ge 2$ and $L_0 := \lceil M_0' \log n \rceil$, it holds that

$$b^{L_0}\Big(\frac{1-\beta}{1+\beta}\Big)^{L_0} \le 1/(2n^{10}). \tag{5.65}$$

Note that such a constant exists since $b\frac{1-\beta}{1+\beta} < 1$. Fix arbitrary $M > M_0'$. Let G be an arbitrary graph with average growth b up to depth $L = \lceil M \log n \rceil$, $\Lambda \subseteq V$ be a subset of the vertices, $\tau : \Lambda \to \{1,2\}$ be a configuration on Λ, and $u \in V \setminus \Lambda$ be a vertex.
Consider the self-avoiding walk tree $T = T_{\mathrm{SAW}}(G, u)$ starting from u, and denote by A the leaves of the tree and by U the set of vertices in T that correspond to v. For a subset of vertices W of the tree we denote by $\sigma(W) = 1$ the event that all vertices in W have the state 1 and analogously for $\sigma(W) = 2$. By Lemma 5.5.1, we have that there is a configuration $\eta : A \to \{1,2\}$ such that for $s \in \{1,2\}$ it holds that

$$\pi_G(\sigma(u) = 1 \mid \sigma(v) = s) = p_s \text{ where } p_s := \pi_T\big(\sigma(u) = 1 \mid \sigma(A) = \eta,\, \sigma(U \setminus A) = s\big). \tag{5.66}$$

Let $T' = (V', E')$ be the subtree of T induced by vertices at distance at most $L-1$ from u and let $A' = A \cap V'$, $U' = U \cap V'$. For $s \in \{1,2\}$, let

$$p_s' := \pi_{T'}(\sigma(u) = 1 \mid \sigma(A') = \eta(A'), \sigma(U' \setminus A') = s).$$

We will show that

$$|p_1' - p_2'| \le \sum_{\ell=1}^{L} P_\ell(G, u, v)\Big(\frac{1-\beta}{1+\beta}\Big)^\ell, \tag{5.67}$$

$$|p_1 - p_1'| \le 1/(2n^{10}), \quad |p_2 - p_2'| \le 1/(2n^{10}). \tag{5.68}$$

Assuming these for the moment, we obtain by (5.66) and the triangle inequality that

$$|\pi_G(\sigma(u)=1 \mid \sigma(v)=1)-\pi_G(\sigma(u)=1 \mid \sigma(v)=2)| = |p_1-p_2| \le \frac{1}{n^{10}}+\sum_{\ell=1}^{L} P_\ell(G,u,v)\Big(\frac{1-\beta}{1+\beta}\Big)^{\ell},$$

thus proving the lemma. It thus remains to prove (5.67) and (5.68).
To prove (5.67), note that by Lemma 5.5.2, we have that

$$|p_1'-p_2'| \le \sum_{w\in U'\setminus A'} \Big(\frac{1-\beta}{1+\beta}\Big)^{\mathrm{dist}(u,w)} \le \sum_{\ell=1}^{L} P_\ell(G,u,v)\Big(\frac{1-\beta}{1+\beta}\Big)^{\ell},$$

where the last inequality follows from observing that each vertex $w \in U' \setminus A'$ with $\mathrm{dist}(u,w)=\ell$ corresponds to a distinct path with $\ell+1$ vertices between u and v in G. This proves (5.67).
To prove (5.68), we focus on showing that $|p_1-p_1'| \le 1/(2n^{10})$, the other inequality being completely analogous. We follow closely a similar argument which was presented in the proof of Lemma 5.5.3. Let J be the set of vertices in $V' \setminus (U' \cup A')$ whose distance from u is exactly $L-1$ in T and note that $|J| \le b^L$ since G has average growth b up to depth L. Note also that, conditioned on the configurations on J and $U' \cup A'$, the probability that $\sigma(u)=1$ depends only on T' (and not on the configuration on rest of the tree T), i.e., we can expand p as

$$p_1 = \sum_{\iota: J\to\{1,2\}} \pi_{T'}(\sigma(u)=1 \mid \sigma(A')=\eta(A'), \sigma(U'\setminus A')=1, \sigma(J)=\iota) \\ \times \pi_T(\sigma(J)=\iota \mid \sigma(A)=\eta, \sigma(U\setminus A)=1). \tag{5.69}$$

Therefore, $|p_1-p_1'| \le 1/(2n^{10})$ will follow by showing that, for any configuration $\iota: J \to \{1,2\}$ it holds that

$$\big|\pi_{T'}\big(\sigma(u)=1 \mid \sigma(A')=\eta(A'), \sigma(U'\setminus A')=1,\ \sigma(J)=\iota\big) - p_1'\big| \le 1/(2n^{10}). \tag{5.70}$$

We can expand p_1' analogously to (5.69) by conditioning on the configuration on J, so to prove (5.69) it suffices to show that for any two configurations $\iota_1, \iota_2 : J \to [q]$ it holds that $\kappa \le 1/(2n^{10})$ where

$$\kappa := \big|\pi_{T'}\big(\sigma(u)=1 \mid \sigma(A')=\eta(A'), \sigma(U'\setminus A')=1,\ \sigma(J)=\iota_1\big) \\ - \pi_{T'}\big(\sigma(u)=1 \mid \sigma(A')=\eta(A'), \sigma(U'\setminus A')=1,\ \sigma(J)=\iota_2\big)\big|$$

By the strong spatial mixing result of Lemma 5.5.2, we have that

$$\kappa \leq \sum_{v \in J} \Big(\frac{1-\beta}{\beta+1}\Big)^{\mathrm{dist}(u,v)} = |J|\Big(\frac{1-\beta}{\beta+1}\Big)^{L} \leq b^{L}\Big(\frac{1-\beta}{1+\beta}\Big)^{L}.$$

Combining this with the choice of M_0' (cf. (5.65)), we obtain $\kappa \leq 1/(2n^{10})$, thus concluding the proof of $|p_1 - p_1'| \leq 1/(2n^{10})$ and therefore completing the proof of Lemma 5.3.7. □

Chapter 6

Classification of Complexity : Counting Boolean Bounded-Degree CSP

This chapter is based on the joint work with Andreas Galanis and Leslie Goldberg [36].

6.1 Introduction

Constraint Satisfaction Problems (CSPs) was originally defined in artificial intelligence. The paradigm provides a general framework of problems from diverse areas such as operations research, database theory, scheduling and graph theory. Moreover, the problem of counting solutions to CSPs also captures partition function problems from statistical physics. In particular, Boolean CSPs can be viewed as a generalisation of hypergraph 2-state spin models. In this chapter, we will be concerned with the complexity of approximate counting Boolean CSPs.

An instance $I = (V, \mathcal{C})$ of a Boolean counting CSP consists of a set V of *variables* and a set $\mathcal{C}$ of constraints. An assignment $\sigma : V \to \{0, 1\}$ assigns a Boolean value (which corresponds to a "spin" state) to each variable. Each constraint associates a tuple $(v_1, \ldots, v_k)$ of variables with a Boolean relation which constrains the spins that can be assigned to $v_1, \ldots, v_k$. An assignment σ is said to "satisfy" the constraint if the tuple $(\sigma(v_1), \ldots, \sigma(v_k))$ is in the corresponding relation, and it is said to be "satisfying" if it satisfies all constraints. A Constraint Satisfaction Problem comes with two important parameters — the constraint language Γ is the set of all relations that may be used in constraints and the degree Δ is the maximum number of times that any variable $v \in V$

may be used in constraints in any instance. The number of satisfying assignments is denoted Z_I. The computational problem $\#\mathsf{CSP}_\Delta(\Gamma)$ is the problem of computing Z_I, given a CSP instance I with constraints in Γ and degree at most Δ. We use $\#\mathsf{CSP}(\Gamma)$ to denote the version of the problem in which the degree of instances is unconstrained.

6.1.1 Definitions and statement of our classification result

Before giving formal definitions of the problems that we study, we introduce some notation. We use boldface letters to denote Boolean vectors. A *pseudo-Boolean* function is a function of the form $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ for some positive integer k, which is called the *arity* of f.

Definition 6.1.1. Given a pseudo-Boolean function $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$, we use the notation R_f to denote the relation $R_f = \{\mathbf{x} \in \{0,1\}^k \mid f(\mathbf{x}) > 0\}$, which is the relation underlying f.

If the range of f is $\{0,1\}$ then f is said to be a *Boolean function* and of course in that case $R_f = \{\mathbf{x} \in \{0,1\}^k \mid f(\mathbf{x}) = 1\}$.
In order to allow consistency with obvious generalisations, our formal definition of the Boolean Constraint Satisfaction Problem is in terms of Boolean functions (rather than, equivalently, using the underlying relations).
A *Constraint language* Γ is a set of pseudo-Boolean functions. It is called a *Boolean constraint language* if all of the functions in it are Boolean functions. An instance $I = (V, \mathcal{C})$ of a CSP with constraint language Γ consists of a set V of variables and a set $\mathcal{C}$ of constraints. Each constraint $C_i \in \mathcal{C}$ is of the form $f_i(v_{i,1}, \ldots, v_{i,k_i})$ where f_i is an arity-k_i function in Γ and $(v_{i,1}, \ldots, v_{i,k_i})$ is a tuple of (not necessarily distinct) variables in V. The constraint C_i is said to be "Repeat-Free" if all of the variables are distinct. Similar to the definition of spins systems, we can also define the weight, partition function and the Gibbs distribution of I as follows. Each assignment $\sigma : V \to \{0,1\}$ of Boolean values to the variables in V has a weight

$$w_I(\sigma) := \prod_{f_i(v_{i,1},\ldots,v_{i,k_i}) \in \mathcal{C}} f_i(\sigma(v_{i,1}), \ldots, \sigma(v_{i,k_i})).$$

The *partition function* maps the instance I to the quantity

$$Z_I := \sum_{\sigma : V \to \{0,1\}} w_I(\sigma) = \sum_{\sigma : V \to \{0,1\}} \prod_{f_i(v_{i_1},\ldots,v_{i,k_i}) \in \mathcal{C}} f_i(\sigma(v_{i_1}), \ldots, \sigma(v_{i,k_i})).$$

If Γ is a Boolean constraint language then it is easy to see that $w_I(\sigma) = 1$ if the assignment is satisfying and $w_I(\sigma) = 0$, otherwise. Thus, Z_I is the number of satisfying assignments of I.

When $Z_I > 0$, we will use $\mu_I(\cdot)$ to denote the Gibbs distribution corresponding to Z_I, which is given by

$$\mu_I(\sigma) = \frac{w_I(\sigma)}{Z_I} \text{ for all } \sigma : V \to \{0,1\}.$$

The *degree* $d_v(C)$ of a variable v in a constraint C is the number of times that the variable v appears in the tuple corresponding to C and the degree d_v of the variable is $d_v = \sum_{C \in \mathcal{C}} d_v(C)$. Finally, the degree of the instance I is $\max_{v \in V} d_v$.

Definition 6.1.2. $\#\mathsf{CSP}_\Delta(\Gamma)$ is the problem of computing Z_I, given a CSP instance I with constraints in Γ and degree at most Δ. $\#\mathsf{CSP}(\Gamma)$ is the version of the problem in which the degree of instances is unconstrained. $\#\mathsf{CSP}_{\Delta,c}(\Gamma)$ has an extra parameter $c > 1$ that captures the desired accuracy. The problem is to compute a number $\widehat{Z}$ such that $c^{-n} Z_I \leq \widehat{Z} \leq c^n Z_I$, where n is the number of variables in the instance I. The problems $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$, $\#\mathsf{NoRepeatCSP}(\Gamma)$ and $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ are defined similarly, except that inputs are restricted so that all constraints are Repeat-Free.

Definition 6.1.3. A Boolean function $f : \{0,1\}^k \to \{0,1\}$ is *affine* if there is a $k \times k$ Boolean matrix $\mathbf{A}$ and a length-k Boolean vector $\mathbf{b}$ such that R_f is equal to the set of solutions $\mathbf{x}$ of $\mathbf{Ax} = \mathbf{b}$ over $\mathrm{GF}(2)$.

Definition 6.1.4 (The set of functions IM_2)**.** A Boolean function $f : \{0,1\}^k \to \{0,1\}$ is in IM_2 if $f(x_1, \ldots, x_k)$ is logically equivalent to a conjunction of (any number of) predicates of the form x_i, $\neg x_i$ or $x_i \Rightarrow x_j$.

We have now defined all of the terms in our main theorem apart from some well-known concepts from complexity theory, which we discuss next.

Definition 6.1.5 (FP)**.** FP is the class of computational problems (with numerical output) that can be solved in polynomial time.

Definition 6.1.6 (AP-reduction)**.** For two counting problems $\#\mathsf{A}$ and $\#\mathsf{B}$, we say that $\#\mathsf{A}$ is $\#\mathsf{B}$-easy if there is an approximation-preserving (AP)-reduction from $\#\mathsf{A}$ to $\#\mathsf{B}$. The formal definition of an AP-reduction can be found in [25]. It is a randomised Turing reduction that yields close approximations to $\#\mathsf{A}$ when provided with close approximations to $\#\mathsf{B}$.

The definition of AP-reduction relates to the definition of FPRAS in the sense that the existence of an FPRAS for #B implies the existence of an FPRAS for #A. We say that #A is #B-hard if there is an AP-reduction from #B to #A. Finally, we say that #A is #B-equivalent if #A is both #B-easy and #B-hard.
The problem of counting satisfying assignments of a Boolean formula is denoted by #SAT. Every counting problem in #P is AP-reducible to #SAT, so #SAT is said to be complete for #P with respect to AP-reductions. It is known that there is no FPRAS for #SAT unless RP $=$ NP. The problem of counting independent sets in a bipartite graph is denoted by #BIS. The problem #BIS appears to be of intermediate complexity: there is no known FPRAS for #BIS (and it is generally believed that none exists) but there is no known AP-reduction from #SAT to #BIS. Indeed, #BIS is complete with respect to AP-reductions in the complexity class $\#\mathsf{RH}\Pi_1$.
Given all of these definitions, we now formally state our main result.

Theorem 6.1.7. *Let Γ be a Boolean constraint language. Then,*

1. *If every function in Γ is affine then $\#\mathsf{CSP}(\Gamma)$ and $\#\mathsf{NoRepeatCSP}(\Gamma)$ are both in* FP.
2. *Otherwise, if $\Gamma \subseteq \mathsf{IM}_2$, then there exists an integer Δ_0 such that for all $\Delta \geq \Delta_0$, $\#\mathsf{CSP}_\Delta(\Gamma)$ and $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ are both #BIS-equivalent under* AP*-reductions, and*
3. *Otherwise, there exists an integer Δ_0 such that for all $\Delta \geq \Delta_0$, there exists a real number $c > 1$ such that $\#\mathsf{CSP}_{\Delta,c}(\Gamma)$ and $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ are both* NP*-hard.*

6.1.2 Prior works

Given the usefulness of CSPs, the study of the complexity of CSPs is an extremely active area in computational complexity (for example, see [13] and the references therein). For every $\Delta \geq 3$, the work of Cai, Lu and Xia [15] completely classifies the complexity of exactly solving $\#\mathsf{CSP}_\Delta(\Gamma)$, depending on the parameter Γ. If every relation in Γ is affine, then $\#\mathsf{CSP}_\Delta(\Gamma)$ is solvable in polynomial time (so the problem in the complexity class called FP). Otherwise, it is #P-complete. The characterisation of Cai, Lu and Xia is exactly the same classification that was obtained for the unbounded problem $\#\mathsf{CSP}(\Gamma)$ by Creignou and Hermann [17]. Thus, as far as exact counting is concerned, the degree-bound Δ does not affect the complexity as long as $\Delta \geq 3$.

However, much less is known about the complexity of *approximately* solving $\#\mathsf{CSP}_\Delta(\Gamma)$. In fact, even the decision problem $\mathsf{CSP}_\Delta(\Gamma)$ is still open. For $\Delta \geq 3$, Dalmau and Ford [20] have solved the special case where Γ includes both of the unary relations $R_{\delta_0} = \{0\}$ and $R_{\delta_1} = \{1\}$. This special case is known as the "conservative case" in the CSP literature. For $\Delta \geq 6$, Dyer et al. [27] have classified the difficulty of the approximate counting problem, also in the conservative case.

- If every relation in Γ is affine, then $\#\mathsf{CSP}_\Delta(\Gamma \cup \{R_{\delta_0}, R_{\delta_1}\})$ is in FP.
- Otherwise, if every relation in Γ is in a class called IM_2 then $\#\mathsf{CSP}_\Delta(\Gamma \cup \{R_{\delta_0}, R_{\delta_1}\})$ is equivalent under approximation-preserving (AP) reductions to the counting problem $\#\mathsf{BIS}$ (the problem of counting independent sets in bipartite graphs).
- Otherwise, there is no FPRAS for $\#\mathsf{CSP}_\Delta(\Gamma \cup \{R_{\delta_0}, R_{\delta_1}\})$ unless $\mathsf{NP} = \mathsf{RP}$.

However, the restriction that R_{δ_0} and R_{δ_1} are contained in the constraint language is a severe one. It does not apply to many natural applications, such as the hypergraph independent set problem on k-uniform hypergraphs.[1] The hypergraph independent set problem on k-uniform hypergraphs is equivalent to the problem $\mathsf{CSP}(\{\mathsf{OR}_k\})$ where OR_k is the relation $\{0,1\}^k \setminus \{(0,0,\ldots,0)\}$. It is known that

1. there is an FPRAS for $\#\mathsf{CSP}(\{\mathsf{OR}_k\})$ if $\Delta = O(2^{k/2})$ [45];
2. it is NP-hard to approximate $\#\mathsf{CSP}(\{\mathsf{OR}_k\})$ when $\Delta = \Omega(2^{k/2})$ [6].

In contrast, $\#\mathsf{CSP}_\Delta(\{\mathsf{OR}_k\} \cup \{R_{\delta_0}, R_{\delta_1}\})$ is NP-hard to approximate as long as $\Delta \geq 6$ [27]. Our result removed this severe restriction that R_{δ_0} and R_{δ_1} are contained in Γ from the classification result of $\#\mathsf{CSP}_\Delta(\Gamma)$ from [27]. Note that even for the well-studied case of $\#\mathsf{CSP}_\Delta(\{\mathsf{OR}_k\})$, we do not know its approximation complexity. Therefore, in our result, we focused on whether, for all Boolean constraint languages Γ, there is a "barrier" value Δ_0 such that for all $\Delta \geq \Delta_0$, the problem $\#\mathsf{CSP}_\Delta(\Gamma)$ is computationally hard to approximate.

[1] In a k-uniform hypergraph, all hyperedges have size k.

6.1.3 CSPs and hypergraph spin systems

We will often be interested in the case where Γ contains a single function $f : \{0,1\} \to \mathbb{R}_{\geq 0}$. In this case, the problem can be represented by a hypergraph 2-state spin system. So, for convenience, we make the following definitions.

A *k-tuple hypergraph* $H = (V, \mathcal{F})$ consists of a set V of vertices, together with a set $\mathcal{F}$ of *hyperarcs*, where every hyperarc in $\mathcal{F}$ is a k-tuple of distinct vertices in V. The degree of H is the maximum, over all vertices $v \in V$, of the number of hyperarcs that contain v. Given a function $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$, we let $I_f(H)$ denote the instance of $\#\mathsf{NoRepeatCSP}(\{f\})$ whose constraints correspond to the hyperarcs of H. Given an assignment $\sigma\colon V \to \{0,1\}$ we define $w_{f;H}(\sigma) := \prod_{(v_1,\ldots,v_k)\in\mathcal{F}} f(\sigma(v_1),\ldots,\sigma(v_k))$ and $Z_{f;H} := \sum_{\sigma:V\to\{0,1\}} w_{f;H}(\sigma)$, so $Z_{I_f(H)} = Z_{f;I_f(H)}$. By analogy to the Gibbs distribution on satisfying assignments, when $Z_{f;H} > 0$, we use $\mu_{f;H}(\cdot)$ to denote the probability distribution in which, for all assignments $\sigma : V \to \{0,1\}$, $\mu_{f;H}(\sigma) = w_{f;H}(\sigma)/Z_{f;H}$. Given a function $f : \{0,1\} \to \mathbb{R}_{\geq 0}$, a positive integer Δ, and a real number $c > 1$, the following computational problems are equivalent to $\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$ and $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\{f\})$, respectively.

Name $\#\mathsf{Hyper2Spin}_\Delta(f)$.

Instance A k-tuple hypergraph H with degree at most Δ.

Output The partition function $Z_{f;H}$.

Name $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$.

Instance An n-vertex k-tuple hypergraph H with degree at most Δ.

Output A number $\widehat{Z}$ such that $c^{-n} Z_{f;H} \leq \widehat{Z} \leq c^n Z_{f;H}$.

The name $\#\mathsf{Hyper2Spin}_\Delta(f)$ indicates that the problem is to compute the partition function of a 2-spin system with multi-body interactions specified by f and degree-bound Δ.

6.2 Proof sketch

6.2.1 Overview of the proof of our classification result

In this section, we give a non-technical overview of the proof of Theorem 6.1.7. Our objective is to illustrate the main ideas and obstacles without delving into the more

detailed definitions. Our focus in this section will be on the case where Γ consists of a single Boolean function $f : \{0,1\}^k \to \{0,1\}$. As will be clear in Section 6.6, this case is the main ingredient in the proof of the theorem.

For a Boolean function $f : \{0,1\}^k \to \{0,1\}$, we consider the complexity of the problems $\#\mathsf{Hyper2Spin}_\Delta(f)$ and $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$. Classifying the complexity of these problems is the most important step in the proof of Theorem 6.1.7. Namely, to obtain Theorem 6.1.7, it suffices to show that for every non-affine function f, we have that:

- If f is in IM_2, then for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_\Delta(f)$ is $\#\mathsf{BIS}$-equivalent.
- If f is not in IM_2, then for all sufficiently large Δ, there exists a real number $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.

A typical approach for showing that $\#\mathsf{Hyper2Spin}_\Delta(f)$ and $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ are intractable is to use an instance of the hypergraph spin system to build a "gadget" which simulates an intractable binary 2-spin constraint. This was the approach used in [35], which proved the intractability of $\#\mathsf{Hyper2Spin}_\Delta(f)$ ($\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$) for any *symmetric* non-affine Boolean function f by constructing an instance I of $\#\mathsf{Hyper2Spin}_\Delta(f)$ ($\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$), along with variables x and y, such that for all spins $s_x \in \{0,1\}$ and $s_y \in \{0,1\}$ the marginal distribution $\mu_I(x,y)$ satisfies

$$\mu_I(\sigma(x) = s_x, \sigma(y) = s_y) = \frac{g(s_x, s_y)}{g(0,0) + g(0,1) + g(1,0) + g(1,1)}, \tag{6.1}$$

where g is a binary function that codes up the interaction of an intractable anti-ferromagnetic 2-spin system. We will not need to give detailed definitions of 2-spin systems in this paper. Instead, we give a sufficient condition for intractability.

Definition 6.2.1. A binary function $g : \{0,1\}^2 \to \mathbb{R}_{\geq 0}$ is said to be "*hard*" if all of the following hold:

$$\begin{aligned}
&g(0,0) + g(1,1) > 0,\\
&\min\{g(0,0), g(1,1)\} < \sqrt{g(0,1)g(1,0)},\\
&\max\{g(0,0), g(1,1)\} \leq \sqrt{g(0,1)g(1,0)}.
\end{aligned}$$

It was established in [35] that the ability to "simulate" a hard function g in the sense of (6.1) ensures that $\#\mathsf{Hyper2Spin}_\Delta(f)$ ($\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$) is NP-hard to approximate, even within an exponential factor.

A key feature of *symmetric* Boolean functions f which facilitated such simulation in [35] was the fact that the class of relevant hard function g is well-behaved, and it turned out that it suffices to encode such a hard binary function with only ε-accuracy, for some sufficiently small $\varepsilon > 0$, and this was enough to ensure the NP-hardness of $\#\mathsf{CSP}_{\Delta,c}(\{f\})$. The main obstacle in adapting the approach of [35] to the case where f need not be symmetric in its arguments arises when f is in IM_2. It is unlikely that such a function f can simulate a hard function g in the sense of (6.1) — indeed such a simulation would prove the (very surprising) result that #BIS does not have an FPRAS (unless NP = RP). Thus, for $f \in \mathsf{IM}_2$, we need instead to encode a binary function which will allow us to connect the problem $\#\mathsf{Hyper2Spin}_{\Delta}(f)$ ($\#\mathsf{NoRepeatCSP}_{\Delta}(\{f\})$) to #BIS.

Now consider the binary Boolean function Implies whose underlying relation $R_{\mathsf{Implies}} = \{(0,0),(0,1),(1,1)\}$ contains all (x,y) satisfying $x \Rightarrow y$. Obviously, Implies is not symmetric, and it is not hard according to Definition 6.2.1. On bipartite instances, however, the symmetry can be restored by interpreting differently the spins 0 and 1 on the two parts of the graph, and this leads to a connection with #BIS. In particular, it is well-known [26] that $\#\mathsf{CSP}(\{\mathsf{Implies}\})$ is equivalent to #BIS under AP-reductions. This connection was extended to the bounded-degree setting by [14].

Unfortunately, the symmetrisation which connects $\#\mathsf{CSP}(\{\mathsf{Implies}\})$ to #BIS is not very robust. For example, suppose that a (non-symmetric) Boolean function f can be used to simulate, in the sense of (6.1), a binary function g which is very close to Implies. In particular, suppose that for some $\varepsilon > 0$ and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ satisfying $|\varepsilon_i| \leq \varepsilon$ for $i = 1, 2, 3, 4$, we have

$$\begin{array}{ll} g(0,0) = 1 + \varepsilon + 1, & g(0,1) = 1 + \varepsilon_2, \\ g(1,0) = \varepsilon_3, & g(1,1) = 1 + \varepsilon_4. \end{array}$$

Such a close approximation is about the best that can be expected using the kind of approximate encodings that are available. However, the complexity of asymmetric 2-spin systems is not sufficiently well understood to exploit such a simulation. Surprisingly, for *any* arbitrarily small constant $\varepsilon > 0$, it is not known even whether the unbounded degree version $\#\mathsf{CSP}(\{g\})$ is #BIS-hard, and certainly nothing is known in our bounded-degree setting! The trouble is that the symmetrisation that works for Implies (i.e., when $\varepsilon_i = 0$ for $i = 1, 2, 3, 4$) is no longer guaranteed to symmetrise the imperfect version with the ε_i's, so the swapping of spin-0 and spin-1 values on one side of the bipartite graph leads to an *asymmetric* 2-spin system on bipartite graphs and this does not fall into the scope of

known results [14] concerning bounded-degree bipartite 2-spin systems.
Our approach to handle this problem for $f \in \mathsf{IM}_2$ is to carefully ensure that there is no accuracy error ε in encoding the function Implies. In other words, we show that, using $f \in \mathsf{IM}_2$, we can encode Implies *perfectly*, a task which is surprisingly intricate in the repeat-free setting. Our main technical theorem, Theorem 6.2.9, achieves this goal. Namely, it shows that, for every non-affine Boolean function f, either f simulates a hard function (with arbitrarily small accuracy-error ε, which leads to the desired intractability of $\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$) or else f "supports perfect equality" — a concept which will be defined later, but essentially means that f can be used to perfectly simulate the binary function EQ with underlying relation $R_{\mathsf{EQ}} = \{(0,0),(1,1)\}$. Using EQ, it is possible to simulate repeated variables in constraints, so the $\#\mathsf{BIS}$-hardness of $\#\mathsf{CSP}_\Delta(\{f\})$ follows from [26]. When $f \notin \mathsf{IM}_2$ but f supports perfect equality, instead of reducing to the work in [26], we work somewhat harder to make sure that we also get the strong (exponential factor) inapproximability given in Theorem 6.1.7.

6.2.2 Supporting pinning and equality

Let k be a positive integer and let $H = (V, \mathcal{F})$ be a k-tuple hypergraph. Given a configuration $\sigma : V \to \{0,1\}$ and a subset $T \subseteq V$, we will use σ_T to denote the restriction of σ to vertices in T. For a vertex $v \in V$, we will also use σ_v to denote the spin $\sigma(v)$ of vertex v in σ. The following definitions are generalisations of definitions from [35].

Definition 6.2.2. Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. Suppose that $\varepsilon \geq 0$ and $s \in \{0,1\}$. The k-tuple hypergraph H is an *ε-realisation of pinning-to-s* if there exists a vertex v of H such that $\mu_{f;H}(\sigma_v = s) \geq 1 - \varepsilon$.

Definition 6.2.3. Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ and $s \in \{0,1\}$. We say that f supports *pinning-to-s* if, for every $\varepsilon > 0$, there is a k-tuple hypergraph which is an ε-realisation of pinning-to-s. We say that f supports *perfect* pinning-to-s if there is a k-tuple hypergraph which is a 0-realisation of pinning-to-s.

We now define what it means for a function f to support (perfect) equality which was already discussed in Section 6.2.1.

Definition 6.2.4. Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ and $\varepsilon \geq 0$. The k-tuple hypergraph H is an *ε-realisation of equality* if there exist distinct vertices v_1 and v_2 of H such that, for each

$s \in \{0,1\}$,

$$\mu_{f;H}(\sigma_{v_1} = \sigma_{v_2} = s) \geq (1-\varepsilon)/2.$$

Definition 6.2.5. Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. The function f *supports equality* if, for every $\varepsilon > 0$, there is a k-tuple hypergraph which is an ε-realisation of equality. The function f *supports perfect equality* if there is a k-tuple hypergraph which is a 0-realisation of equality.

6.2.3 Realising conditional distributions induced by pinning and equality

Given a set S of vertices, it will be convenient to follow [35] as follows. We write $\sigma_S = \mathbf{0}$ to denote the event that all vertices in S are assigned the spin 0 under the assignment σ. We similarly write $\sigma_S = \mathbf{1}$ to denote the event that all vertices in S are assigned the spin 1 under the assignment σ. Finally, we use use σ_S^{eq} to denote the event that all vertices in S have the same spin under σ (the spin could be 0 or 1). The following definition is a generalisation of Definition 16 of [35] except that we have changed the notation slightly for convenience.

Definition 6.2.6 ([35, Definition 16])**.** Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. Let $H = (V, \mathcal{F})$ be a k-tuple hypergraph. Let $\mathcal{V} = (V_{\mathsf{pin}0}, V_{\mathsf{pin}1}, \mathcal{V}_{\mathsf{eq}})$ where $V_{\mathsf{pin}0}$ and $V_{\mathsf{pin}1}$ are disjoint subsets of V and $\mathcal{V}_{\mathsf{eq}}$ is a (possibly empty) set of disjoint subsets of $V \backslash (V_{\mathsf{pin}0} \cup V_{\mathsf{pin}1})$. Suppose that: (i) $V_{\mathsf{pin}0} = \emptyset$ if f does not support pinning-to-0, (ii) $V_{\mathsf{pin}1} = \emptyset$ if f does not support pinning-to-1, (iii) $\mathcal{V}_{\mathsf{eq}} = \emptyset$ if f does not support equality, (iv) it holds that $\mu_{f;H}(\sigma_{V_{\mathsf{pin}0}} = \mathbf{0}, \sigma_{V_{\mathsf{pin}1}} = \mathbf{1}, \bigcap_{W \in \mathcal{V}_{\mathsf{eq}}} \sigma_W^{\mathsf{eq}}) > 0$. We will then say that "$\mathcal{V}$ is *admissible* for H with respect to f" and we will denote by $\mu_{f;H}^{\mathrm{cond}(\mathcal{V})}$ the probability distribution $\mu_{f;H}(\cdot \mid \sigma_{V_{\mathsf{pin}0}} = \mathbf{0}, \sigma_{V_{\mathsf{pin}1}} = \mathbf{1}, \bigcap_{W \in \mathcal{V}_{\mathsf{eq}}} \sigma_W^{\mathsf{eq}})$.

Remark 6.2.1. Frequently, instead of formally specifying $\mathcal{V}$, we will specify $\mathcal{V}$ implicitly by just saying "consider the conditional distribution $\mu_{f;H}^{\mathrm{cond}(\mathcal{V})}$ where the vertices in $V_{\mathsf{pin}0}$ are pinned to 0, the vertices in $V_{\mathsf{pin}1}$ are pinned to 1 and for all $W \in \mathcal{V}_{\mathsf{eq}}$, all the vertices in W are forced to be equal".

6.2.4 Simulating hard functions and inapproximability results

We can now give a formal definition of "simulation", along the lines that was informally discussed in Section 6.2.1 (Equation (6.1)).

Definition 6.2.7. Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ and $g : \{0,1\}^t \to \mathbb{R}_{\geq 0}$. The function f simulates the function g if there is a k-tuple hypergraph H, an admissible set $\mathcal{V}$ for H with respect to f, and t vertices $v_1, v_2, \ldots, v_t$ of H such that, for all $(s_1, s_2, \ldots, s_t) \in \{0,1\}^t$,

$$\mu_{f;H}^{\mathrm{cond}(\mathcal{V})}(\sigma(v_1) = s_1, \sigma(v_2) = s_2, \ldots, \sigma(v_t) = s_t) = \frac{g(s_1, s_2, \ldots, s_t)}{\sum_{(s_1', s_2', \ldots, s_t') \in \{0,1\}^t} g(s_1', s_2', \ldots, s_t')}.$$

If $\mathcal{V} = (\emptyset, \emptyset, \emptyset)$, then we say that f *perfectly simulates* g. More generally, we say that f simulates a set of functions $\mathcal{G}$ if f simulates every $g \in \mathcal{G}$.

The connection between "hard" as defined in Definition 6.2.1 and intractability is given in the following lemma.

Lemma 6.2.8 ([35, Lemma 18])**.** *Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. If f simulates a hard function, then for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.* □

Remark 6.2.2. [35, Lemma 18] is stated for symmetric functions, but the proof in [35] also works for asymmetric functions.

6.2.5 Technical lemmas

Our main technical theorem is the following classification of Boolean functions, which asserts that every non-affine function either supports perfect equality or simulates a hard function.

Theorem 6.2.9. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function. Then at least one of three following propositions is true:*

1. *f is affine;*
2. *f supports perfect equality;*
3. *f simulates a hard function.*

Theorem 6.2.9 is proved in Section 6.4. When f simulates a hard function, using Lemma 6.2.8, we can immediately conclude that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. As we already discussed in Section 6.2.1, it is important that, in the case where f does not simulate a hard function, Theorem 6.2.9

guarantees that f supports *perfect* equality (rather than simple imperfect equality); this allows us to recover the connection to #BIS for those $f \in \mathsf{IM}_2$. In fact, when f supports perfect equality, we can effectively carry out (a strengthening of) the analysis in [26] to obtain the following classification which perfectly aligns with Theorem 6.1.7.

Theorem 6.2.10. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function that is not affine. Suppose that f supports perfect equality.*

1. *If f is in IM_2, then for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_\Delta(f)$ is #BIS-equivalent.*
2. *If f is not in IM_2, then for all sufficiently large Δ, there exists a real number $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.*

Theorem 6.2.10 is proved in Section 6.5. Thus, Theorems 6.2.9 and 6.2.10 together achieve the desired classification of $\#\mathsf{Hyper2Spin}_\Delta(f)$ when $f \in \mathsf{IM}_2$ as well as the strong inapproximability results when $f \notin \mathsf{IM}_2$. Before delving into the proofs of Theorems 6.2.9 and 6.2.10 however, it will be instructive to give the main ideas behind the proofs, especially of the more critical Theorem 6.2.9.

To prove Theorem 6.2.9, our proof departs from the previous approaches in the related works [26] and [35]. In these works, f was used to directly encode a binary function which was feasible because of the presence of equality in [26] and the symmetry of f in [35]. Instead, we take a much more painstaking combinatorial approach by using induction on the arity of the function f.

The base case of the induction (proving Theorem 6.2.9 for arity-2 functions) is fairly simple to handle, so let us focus on the induction step. The rough idea, to put the induction hypothesis to work, is to study whether f supports pinning-to-0 or pinning-to-1; then, provided that at least one these pinnings is available, we need to pin appropriately some arguments of f to obtain a function h of smaller-arity. Our goal is then to ensure that h is non-affine; then, we can invoke the induction hypothesis and obtain that h either supports perfect equality or simulates a hard function. From there, since h was obtained by pinning some arguments of f, we will obtain by a transitivity argument (cf. Lemma 6.4.8) that f either supports perfect equality or f simulates the same hard function as h. (A detail here is that, in the case where h supports perfect equality, to conclude that f supports perfect equality from Lemma 6.4.8, we need to ensure that the pinnings of f used to obtain h were perfect.)

Determining which arguments of f need to be pinned is the most challenging aspect of this scheme. Our method for reducing the number of functions under consideration is to symmetrise f in a natural way and obtain a new function f^* which is now symmetric (see Definition 6.4.1). Then, it turns out that there are seven possibilities for the function f^* which we need to consider in detail (the functions are given in Definition 6.3.2). That is, when the symmetrisation of f is one of these seven functions, we have to figure out whether f supports perfect equality and, if not, work out the combinatorial structure of f and pinpoint which arguments are suitable to be pinned. The details of the argument can be found in Section 6.4.2.

The proof of Theorem 6.2.10, where f supports perfect equality, basically follows the approach of [26]. However, to get the stronger inapproximability results, we have to take a detour studying self-dual functions (functions whose value does not change when we complement their arguments). We show that if f is self-dual then it simulates a hard function (Theorem 6.5.6). The problem with self-dual functions is that they do not support pinning-to-0 or pinning-to-1, so we are not able to use the relevant results from [26]. After proving Theorem 6.5.6 and doing some preparatory work in Section 6.5.1 to ensure that "implementations in CSPs" work in the repeat-free setting when f supports perfect equality (see Lemma 6.5.2), the techniques of [26] can be adapted to get Theorem 6.2.10.

6.3 Properties of affine, IM_2, and symmetric functions

For a vector $\mathbf{x}$, we use x_i to denote the i'th entry of $\mathbf{x}$. Further, for vectors $\mathbf{x}$ and $\mathbf{y}$ of the same length, $\mathbf{x} \oplus \mathbf{y}$ will denote the coordinate-wise addition of $\mathbf{x}$ and $\mathbf{y}$ over $\mathrm{GF}(2)$. More generally, for any binary Boolean operator $\otimes$, we will denote by $\mathbf{x} \otimes \mathbf{y}$ the vector whose i-th entry is given by $x_i \otimes y_i$. We will use $\mathbf{0}, \mathbf{1}$ to denote the vectors whose entries are all zeros and all ones, respectively (the length of these vectors will be clear from context). Finally, for a Boolean vector $\mathbf{x}$, $\overline{\mathbf{x}}$ will denote the coordinate-wise "negation" of $\mathbf{x}$, i.e., $\overline{\mathbf{x}} = \mathbf{x} \oplus \mathbf{1}$.

6.3.1 Affine functions

The following well-known characterisation of affine functions (cf. Definition 6.1.3) is instructive and will be useful later. For a proof, see, for example, Lemma 4.10 of [19] or Lemma 11 of [26].

Lemma 6.3.1. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function. Then:*

1. *f is affine iff for every $\mathbf{a}, \mathbf{b}, \mathbf{c} \in R_f$, it holds that $\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c} \in R_f$.*
2. *If f is not affine, then for every $\mathbf{a} \in R_f$, there are $\mathbf{b}, \mathbf{c} \in R_f$ such that $\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c} \notin R_f$.* □

The set of affine *symmetric* Boolean functions $f : \{0,1\}^k \to \{0,1\}$ is given by the following set $\mathsf{EASY}(k)$.

Definition 6.3.2. For $k \geq 2$, let $\mathsf{EASY}(k)$ be the set containing the following seven functions.

$$f^{(k)}_{\mathsf{zero}}(x_1, \ldots, x_k) = 0, \quad f^{(k)}_{\mathsf{one}}(x_1, \ldots, x_k) = 1, \quad f^{(k)}_{\mathsf{allzero}}(x_1, \ldots, x_k) = \mathbf{1}\{x_1 = \ldots = x_k = 0\},$$
$$f^{(k)}_{\mathsf{allone}}(x_1, \ldots, x_k) = \mathbf{1}\{x_1 = \ldots = x_k = 1\}, \quad f^{(k)}_{\mathsf{EQ}}(x_1, \ldots, x_k) = \mathbf{1}\{x_1 = \ldots = x_k\},$$
$$f^{(k)}_{\mathsf{even}}(x_1, \ldots, x_k) = \mathbf{1}\{x_1 \oplus \cdots \oplus x_k = 0\}, \quad f^{(k)}_{\mathsf{odd}}(x_1, \ldots, x_k) = \mathbf{1}\{x_1 \oplus \cdots \oplus x_k = 1\}.$$

6.3.2 A characterisation of IM_2

In the language of universal algebra, Creignou, Kolaitis, and Zanuttini [19] have shown that IM_2 (see Definition 6.1.4) is precisely the "co-clone" corresponding to the "clone" M_2 in Post's lattice (see [8]). Defining clones and co-clones would be a bit of a distraction from this paper, but the only fact that we need is the following (which follows directly from the definitions of clones and co-clones and from the fact that IM_2 is the co-clone corresponding to M_2).

Lemma 6.3.3. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function. Then, the function f is in IM_2 iff for every $\mathbf{x}, \mathbf{y} \in R_f$ it holds that $\mathbf{x} \vee \mathbf{y} \in R_f$ and $\mathbf{x} \wedge \mathbf{y} \in R_f$.* □

6.3.3 The case where f is symmetric: extensions to the asymmetric case

In this section, we state a few results from [35] which were stated for the case where f is a symmetric Boolean function, but whose proof works just as well even when f is asymmetric.

The following lemma, which is Lemma 12 of [35], gives sufficient conditions for pinning-to-0, pinning-to-1 and equality. The statement of the lemma in [35] is restricted to symmetric functions f, but the proof applies to all functions (with the trivial modification

that the vertices in the hyperarcs in the constructed k-tuple hypergraph H must be ordered appropriately).

Lemma 6.3.4 ([35, Lemma 12]). *Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ and let H be a k-tuple hypergraph.*

1. *If there is a vertex v in H such that $\mu_{f;H}(\sigma_v = 0) > \mu_{f;H}(\sigma_v = 1)$, then f supports pinning-to-0.*

2. *If there is a vertex v in H such that $\mu_{f;H}(\sigma_v = 1) > \mu_{f;H}(\sigma_v = 0)$, then f supports pinning-to-1.*

3. *If there are vertices x, y in H such that $\mu_{f;H}(\sigma_x = \sigma_y = 0) = \mu_{f;H}(\sigma_x = \sigma_y = 1)$ and $\mu_{f;H}(\sigma_x = \sigma_y) > \mu_{f;H}(\sigma_x \neq \sigma_y)$, then f supports equality.* □

Lemma 6.3.5 ([35, Lemma 17]). *Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. Let $H = (V, \mathcal{F})$ be a k-tuple hypergraph and let S be a subset of V. Let $\mathcal{V}$ be admissible for H with respect to f. Then, for every $\varepsilon > 0$, there is a k-tuple hypergraph $H' = (V', \mathcal{F}')$ with $V \subseteq V'$ and $\mathcal{F} \subseteq \mathcal{F}'$ such that, for every $\tau : S \to \{0,1\}$, it holds that*

$$\left|\mu_{f;H'}(\sigma_S = \tau) - \mu_{f;H}^{\mathrm{cond}(\mathcal{V})}(\sigma_S = \tau)\right| \leq \varepsilon,$$

where $\mu_{f;H}^{\mathrm{cond}(\mathcal{V})}(\cdot)$ is as in Definition 6.2.6. □

We will also use the following result from [35] which applies to *symmetric* Boolean functions.

Lemma 6.3.6 ([35, Proof of Theorem 3]). *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a* symmetric *Boolean function such that $f \notin \mathsf{EASY}(k)$. Then either f simulates a hard function or f supports perfect equality (or both).*

Proof. We briefly overview the proof in [35], the relevant parts are in [35, Section 4].

1. [35, Lemma 13] shows that every function $f \notin \mathsf{EASY}(k)$ supports one of pinning-to-0, pinning-to-1 or equality.

2. In [35, Section 4.1], the case where f supports both pinning-to-0 and pinning-to-1 is considered. Then, [35] shows that f simulates a hard function.

3. In [35, Section 4.2], the case where f supports equality but neither pinning-to-0 nor pinning-to-1 is considered. The proof splits into cases depending on whether $f(\mathbf{0}) = 0$ or $f(\mathbf{0}) = 1$. When $f(\mathbf{0}) = 0$ ([35, Section 4.2.2]), [35] shows that f supports perfect equality ([35, Lemma 28]). When $f(\mathbf{0}) = 1$ ([35, Section 4.2.3]), [35] shows that f simulates a hard function.

4. In [35, Section 4.3], the case where f supports pinning-to-0 is considered. Then, [35] shows that f simulates a hard function. (The case where f supports pinning-to-1 is identical by switching the spins 0 and 1.)

Thus, for every symmetric function $f : \{0,1\}^k \to \{0,1\}$ such that $f \notin \mathsf{EASY}(k)$, the results of [35] show that either f simulates a hard function or f supports perfect equality. □

6.4 Non-affine Boolean functions either support perfect equality or simulate a hard function

In this section, we prove Theorem 6.2.9, i.e., that every non-affine Boolean function either supports perfect equality or simulates a hard function. Before proving the theorem, we will need a few technical lemmas.

6.4.1 A few preparatory lemmas

Definition 6.4.1 (The symmetrisation f^*)**.** Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. We denote by f^* the symmetrisation of f obtained as follows. Let S_k denote the set of all permutations $\pi : [k] \to [k]$. Then $f^* : \{0,1\}^k \to \mathbb{R}_{\geq 0}$ is the function defined by

$$f^*(x_1, \ldots, x_k) = \prod_{\pi \in S_k} f(x_{\pi(1)}, \ldots, x_{\pi(k)}).$$

Lemma 6.4.2. *Let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f^* = f_{\mathsf{allone}}$. Then f supports perfect pinning-to-1.*

Proof. Let $H = (V, \mathcal{F})$ be the k-tuple hypergraph with $V = \{v_1, v_2, \ldots, v_k\}$ and $\mathcal{F} = \{e_\pi \mid \pi \in S_k\}$ where $e_\pi = (v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(k)})$. Since $f^* = f_{\mathsf{allone}}$, we have that for all $\sigma : V \to \{0,1\}$ it holds that $w_{f,H}(\sigma) > 0$ if and only if $\sigma(v_1) = \sigma(v_2) = \cdots = \sigma(v_k) = 1$. Thus, f supports perfect pinning-to-1. □

Completely analogously, we have the following pinning lemma when $f^* = f_{\mathsf{allzero}}$.

Lemma 6.4.3. *Let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f^* = f_{\mathsf{allzero}}$. Then f supports perfect pinning-to-0.* □

For any function f such that $f^* = f_{\mathsf{zero}}$, we have the following pinning lemma.

Lemma 6.4.4. *Let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f \neq f_{\mathsf{zero}}$ and $f^* = f_{\mathsf{zero}}$. Then at least one of the following two propositions is true:*

1. *f supports perfect pinning-to-0 and perfect pinning-to-1;*
2. *f supports perfect equality.*

Proof. Note that the conditions in the lemma imply that $k \geq 2$. Let S_k denote the set of all permutations $\pi : [k] \to [k]$ and let $\mathrm{id} \in S_k$ denote the identity permutation. For any subset $A \subseteq S_k$, let f_A be the function defined by $f_A(w_1, \ldots, w_k) := \prod_{\pi \in A} f(w_{\pi(1)}, \ldots, w_{\pi(k)})$. Note that $f_{S_k} = f^* = f_{\mathsf{zero}}$. Also, for any $\pi \in S_k$ we have $f_{\{\pi\}} \neq f_{\mathsf{zero}}$ (since $f \neq f_{\mathsf{zero}}$). By iteratively removing permutations from S_k we will thus obtain a subset $T \subseteq S_k$ with $|T| > 1$ such that $f_T = f_{\mathsf{zero}}$ and, for every $\pi \in T$, it holds that $f_{T \setminus \{\pi\}} \neq f_{\mathsf{zero}}$. By renaming the variables if necessary, we may assume that $\mathrm{id} \in T$.
Let $H_0 = (V_0, \mathcal{F}_0)$ be the k-tuple hypergraph with vertex set $V_0 = \{x_1, \ldots, x_k\}$ and hyperarc set $\mathcal{F}_0 = \cup_{\pi \in T \setminus \{\mathrm{id}\}} \{(x_{\pi(1)}, \ldots, x_{\pi(k)})\}$. By the choice of the set T, we have that $Z_{f;H_0} > 0$. For $i = 1, \ldots, k$, let $H_i = (V_i, \mathcal{F}_i)$ be the k-tuple hypergraph with vertex set $V_i = V_0 \cup \{y_{i+1}, \ldots, y_k\}$ and hyperarc set $\mathcal{F}_i = \mathcal{F}_0 \cup \{(x_1, \ldots, x_i, y_{i+1}, \ldots, y_k)\}$. Again by the choice of the set T we have that $Z_{f;H_k} = 0$. Thus, there exists $0 \leq j < k$ such that $Z_{f;H_j} > 0$ and $Z_{f;H_{j+1}} = 0$. Note that for every assignment $\sigma : V_j \to \{0,1\}$ with $w_{f;H_j}(\sigma) > 0$ it holds that $\sigma(x_{j+1}) \neq \sigma(y_{j+1})$; otherwise, for the assignment $\sigma' = \sigma_{V_{j+1}}$ (i.e., the restriction of the assignment σ to the set V_{j+1}), it would hold that $w_{f;H_{j+1}}(\sigma') > 0$, contradicting that $Z_{f;H_{j+1}} = 0$.
For $s_1, s_2 \in \{0,1\}$, let

$$Z_{s_1,s_2} := \sum_{\substack{\sigma : V_j \to \{0,1\}; \\ \sigma(x_{j+1}) = s_1, \sigma(y_{j+1}) = s_2}} w_{f;H_j}(\sigma)\,.$$

By the argument above, we have that $Z_{0,0} = Z_{1,1} = 0$. Since $Z_{f;H_j} > 0$, we have that at least one of $Z_{0,1}$ and $Z_{1,0}$ is non-zero. In fact, we may assume that both are non-zero, since otherwise f supports both perfect pinning-to-0 and perfect pinning-to-1 so

proposition 1 in the statement of the lemma is satisfied (for example, if $Z_{1,0} = 0$, then $\mu_{f;H_j}(\sigma(x_{j+1}) = 0) = 1$ and $\mu_{f;H_j}(\sigma(y_{j+1}) = 1) = 1$).
Let J_1, J_2 be two disjoint copies of H_j. Denote by u_1, u_2 the vertices corresponding to x_{j+1} in J_1, J_2, respectively. Also, denote by v_1, v_2 the vertices corresponding to y_{j+1} in J_1, J_2. Let $J = (V, \mathcal{F})$ be the k-tuple hypergraph obtained by taking the union of J_1 and J_2 and identifying the vertices u_2 and v_1 into a new vertex w (i.e., we merge the vertex corresponding to x_{j+1} in J_2 and the vertex corresponding to y_{j+1} in J_1).
For $s_1, s_2 \in \{0, 1\}$, let

$$Z'_{s_1,s_2} := \sum_{\substack{\sigma:V\to\{0,1\};\\ \sigma(u_1)=s_1,\,\sigma(v_2)=s_2}} w_{f;J}(\sigma).$$

By considering the spin of the vertex w, we obtain that

$$Z'_{s_1,s_2} = Z_{s_1,0}Z_{0,s_2} + Z_{s_1,1}Z_{1,s_2},$$

which gives that

$$Z'_{0,0} = Z_{0,1}Z_{1,0}, \quad Z'_{0,1} = 0, \quad Z'_{1,0} = 0, \quad Z'_{1,1} = Z_{1,0}Z_{0,1}.$$

Since $Z_{0,1}, Z_{1,0} \neq 0$, we obtain that

$$\mu_{f;J}(\sigma(u_1) = \sigma(v_2) = 0) = \mu_{f;J}(\sigma(u_1) = \sigma(v_2) = 1) = \frac{1}{2},$$

and hence f supports perfect equality. □

For any function f, we can show that if f^* is f_{EQ}, f_{odd} or f_{even} then f supports perfect equality.

Lemma 6.4.5. *Let $f : \{0, 1\}^k \to \{0, 1\}$ be a k-ary Boolean function. Suppose that $f^* = f_{\mathsf{EQ}}$. Then f supports perfect equality.*

Proof. Let $H = (V, \mathcal{F})$ be the k-tuple hypergraph with $V = \{v_1, v_2, \ldots, v_k\}$ and hyperarc set $\mathcal{F} = \{e_\pi \mid \pi \in S_k\}$ where $e_\pi = (v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(k)})$. Since $f^* = f_{\mathsf{EQ}}$, we have that for all $\sigma : V \to \{0, 1\}$ it holds that $w_{f,H}(\sigma) > 0$ iff $\sigma(v_1) = \sigma(v_2) = \cdots = \sigma(v_k) = 1$ or $\sigma(v_1) = \sigma(v_2) = \cdots = \sigma(v_k) = 0$. Thus, f supports perfect equality. □

Lemma 6.4.6. *Let $f : \{0, 1\}^k \to \{0, 1\}$ be a k-ary Boolean function. Suppose that $f^* \in \{f_{\mathsf{odd}}, f_{\mathsf{even}}\}$. Then f supports perfect equality.*

Proof. Let $H = (V, \mathcal{F})$ be the k-tuple hypergraph with $V = \{v_1, v_2, \ldots, v_{k+1}\}$ and $\mathcal{F} = \{e_\pi \mid \pi \in S_k\} \cup \{e'_\pi \mid \pi \in S_k\}$ where $e_\pi = (v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(k)})$ and $e'_\pi = (v_{\pi(1)+1}, v_{\pi(2)+1}, \ldots, v_{\pi(k)+1})$ (note that H has $k+1$ vertices and $2k!$ hyperarcs). Since f^* is either f_{odd} or f_{even}, for all $\sigma : V \to \{0,1\}$ with $w_{f,H}(\sigma) > 0$, we have that the parity of number of ones among $\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_k)$ and the parity of number of ones among $\sigma(v_2), \sigma(v_3), \ldots, \sigma(v_{k+1})$ must be the same and thus $\sigma(v_1) = \sigma(v_{k+1})$. Furthermore, for $s \in \{0,1\}$, there are exactly 2^{k-1} assignments $\sigma : V \to \{0,1\}$ such that $w_{f;H}(\sigma) > 0$, $\sigma(v_1) = \sigma(v_{k+1})$ and $\sigma(v_2) \oplus \sigma(v_3) \oplus \cdots \oplus \sigma(v_k) = s$. It follows that

$$\mu_{f;H}(\sigma(v_1) = \sigma(v_{k+1}) = 0) = \mu_{f;H}(\sigma(v_1) = \sigma(v_{k+1}) = 1) = \frac{1}{2},$$

which means that f supports perfect equality. □

By the above lemmas, we can show that some functions can be either dealt with directly, or reduced to other functions with smaller arity.

Definition 6.4.7. For $s \in \{0,1\}$, let $\delta_s : \{0,1\} \to \{0,1\}$ be the Boolean function defined by $\delta_s(s) = 1$ and $\delta_s(1 \oplus s) = 0$. Define $f_{i \to s}$ to be the function obtained from f by pinning the i-th argument of f to s, i.e.

$$f_{i \to s}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = \sum_{x_i} f(x_1, \ldots, x_k) \cdot \delta_s(x_i).$$

Similarly, for $S, T \subseteq [k]$, let $f_{S \to 0, T \to 1}$ be the $(k - |S \cup T|)$-ary function obtained from f by pinning the arguments in S to 0 and the arguments in T to 1, i.e.,

$$f_{S \to 0, T \to 1} = \sum_{x_i : i \in S \cup T} f(x_1, \ldots, x_k) \cdot \prod_{i \in S} \delta_0(x_i) \cdot \prod_{j \in T} \delta_1(x_j).$$

If $S = \emptyset$ or $T = \emptyset$, we will omit $S \to 0$ or $T \to 1$ from notation.

Using Definition 6.4.7, we have the following lemma:

Lemma 6.4.8. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function. Suppose that S_0 and S_1 are disjoint subsets of $[k]$ such that, for $a \in \{0,1\}$, S_a is empty if f does* not *support perfect pinning-to-a. Let $h = f_{S_0 \to 0, S_1 \to 1}$.*

1. *If h supports equality, then f also supports equality. Further, if h supports perfect equality, then f also supports perfect equality.*

2. *If h supports pinning-to-s for some $s \in \{0,1\}$, then f also supports pinning-to-s.*

3. *If h simulates a function $g : \{0,1\}^2 \to \mathbb{R}_{\geq 0}$ that is not $f^{(2)}_{\mathsf{zero}}$ then f simulates g as well. Also, if h perfectly simulates g then f perfectly simulates g as well.*

Proof. Without loss of generality, we assume that the arity of h is n, and that $S_0 \cup S_1 = \{n+1, n+2, \ldots, k\}$. For each $a \in \{0,1\}$, if S_a is non-empty, then by assumption f supports perfect pinning-to-a, so there exists a k-tuple hypergraph $H_a = (V_a, \mathcal{F}_a)$ with a vertex $w_a \in V_a$ such that $\mu_{f;H_a}(\sigma_{w_a} = a) = 1$.

We now give a general construction which takes any n-tuple hypergraph $H = (V, \mathcal{F})$ and produces a new k-tuple hypergraph $H' = (V', \mathcal{F}')$. To do this, we take $k - n$ new vertices $v'_{n+1}, \ldots, v'_k$ that are not in V and let $V' = V \cup \{v'_{n+1}, \ldots, v'_k\}$. The hyperarcs of H' are in one-to-one correspondence with those in H: For each hyperarc $(u_1, u_2, \ldots, u_n)$ in H, we add the hyperarc $(u_1, u_2, \ldots, u_n, v'_{n+1}, v'_{n+2}, \ldots, v'_k)$ to H'. Moreover, for $i \in S_0$, add a distinct copy of H_0 to H' by identifying v'_i with the vertex w_0 in H_0. Also, for $i \in S_1$, add a distinct copy of H_1 to H' by identifying v'_i with the vertex w_1 in H_1.

Say that an assignment $\sigma : V' \to \{0,1\}$ is *relevant* if, for each $a \in \{0,1\}$ and each $i \in S_a$, $\sigma(v_i) = a$. The copies of H_0 and H_1 ensure that, for every assignment $\sigma : V' \to \{0,1\}$ with $w_{f;H'}(\sigma) > 0$, σ is relevant. The definition of h ensures that, for any relevant assignment σ,

$$w_{f;H'}(\sigma) = w_{h;H}(\sigma_V). \tag{6.2}$$

We now use (6.2) to establish the three items in the statement of the lemma.

1. Suppose that h supports equality. For any $\varepsilon \in (0,1)$, there is an n-tuple hypergraph $H = (V, \mathcal{F})$ and two vertices x and y of H such that, for every $s \in \{0,1\}$, $\mu_{h;H}(\sigma_x = \sigma_y = s) \geq (1-\varepsilon)/2$. Construct H' from H using the general construction above. From (6.2), we conclude that, for any $s \in \{0,1\}$, $\mu_{f;H'}(\sigma_x = \sigma_y = s) = \mu_{h;H}(\sigma_x = \sigma_y = s) \geq (1-\varepsilon)/2$, so f supports equality. If h supports *perfect* equality, then we can take $\varepsilon = 0$ in this argument, obtaining the conclusion that f also supports *perfect* equality.

2. Suppose that h supports pinning-to-s. For any $\varepsilon \in (0,1)$ there is an n-tuple hypergraph $H = (V, \mathcal{F})$ and a vertex x of H such that $\mu_{h;H}(\sigma_x = s) \geq 1 - \varepsilon$. Construct H' from H using the general construction above. From (6.2), we conclude that, $\mu_{f;H'}(\sigma_x = s) = \mu_{h;H}(\sigma_x = s) \geq 1 - \varepsilon$, so f supports pinning-to-s.

3. Let $g : \{0,1\}^2 \to \mathbb{R}_{\geq 0}$ be a function that is not $f_{\mathsf{zero}}^{(2)}$. Suppose first that h simulates g. By the definition of "simulates", there exists an n-tuple hypergraph H with admissible $\mathcal{V}$ (with respect to h) and two vertices u and v in H such that, for every $s, t \in \{0,1\}$, it holds that

$$\mu_{h;H}^{\mathrm{cond}(\mathcal{V})}(\sigma(u) = s, \sigma(v) = t) = \frac{g(s,t)}{\sum_{i,j\in\{0,1\}} g(i,j)}. \tag{6.3}$$

Since $g \neq f_{\mathsf{zero}}^{(2)}$, the expression in (6.3) is well-defined.

Construct H' from H using the general construction above. From Items 1 and 2 of the lemma, if h supports equality or pinning-to-0 or pinning-to-1 then so does f. Thus, $\mathcal{V}$ is admissible for H' with respect to f. It follows from (6.2) that

$$\begin{aligned}\mu_{h;H}^{\mathrm{cond}(\mathcal{V})}(\sigma(u) = s, \sigma(v) = t) &= \mu_{f;H'}^{\mathrm{cond}(\mathcal{V})}(\sigma(u) = s, \sigma(v) = t \mid \wedge_{i\in S_0}\sigma(v_i) = 0, \wedge_{i\in S_1}\sigma(v_i) = 1) \\ &= \mu_{f;H'}^{\mathrm{cond}(\mathcal{V})}(\sigma(u) = s, \sigma(v) = t),\end{aligned}$$

so, using (6.3), we obtain that f simulates g, as wanted. If h *perfectly* simulates g then we can take $\mathcal{V} = (\emptyset, \emptyset, \emptyset)$, so the argument shows that f perfectly simulates g.

□

6.4.2 Proof that every non-affine Boolean functions either supports perfect equality or simulates a hard function

Definition 6.4.9 (Ω_f, χ_S)**.** Let $f : \{0,1\}^k \to \mathbb{R}_{\geq 0}$. For $S \subseteq [k]$, χ_S denotes the characteristic vector of S, which is the length-k Boolean vector such that, for all $i \in [k]$, the i-th bit of χ_S is 1 if and only if $i \in S$. Finally, $\Omega_f = \{S \subseteq [k] \mid \chi_S \in R_f\}$, where R_f is the relation underlying f, defined at the beginning of Section 6.1.1.

Definition 6.4.10. A function $f : \{0,1\}^k \to \{0,1\}$ is *semi-trivial* if and only if there exists a set S such that $\Omega_f = \{T \mid S \subseteq T \subseteq [k]\}$ or $\Omega_f = \{T \mid T \subseteq S\}$.

Remark 6.4.1. Every semi-trivial function is affine.

Lemma 6.4.11. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f \neq f_{\mathsf{allone}}$ and $f^* = f_{\mathsf{allone}}$. Let S be a set in Ω_f such that $S \neq [k]$. Then at least one of the following propositions is true:*

1. *$\forall T \supseteq S$, we have $T \in \Omega_f$;*

2. *f supports perfect equality;*

3. *f simulates a hard function.*

Proof. Without loss of generality (by re-numbering the variables), let $S = \{n+1, n+2, \ldots, k\}$ for some integer $n \geq 1$. By Lemma 6.4.2, f supports perfect pinning-to-1. Let $h(x_1, x_2, \ldots, x_n) = f(x_1, \ldots, x_k)_{S \to 1}$. Note that $h^*(\mathbf{0}) = 1$ and $h^*(\mathbf{1}) = 1$. We may assume that $n \geq 2$ (otherwise $\forall T \supseteq S$, we have $T \in \Omega_f$).

Case 1. $h^* \notin \mathsf{EASY}(n)$. In this case, Lemma 6.3.6 ensures that either h simulates a hard function or h supports perfect equality (or both). If h simulates a hard function, then by Item 3 of Lemma 6.4.8, f also simulates a hard function. If h supports perfect equality, then, by Item 1 of Lemma 6.4.8, f also supports perfect equality.

Case 2. $h^* \in \mathsf{EASY}(n)$. Then $h^* \in \{f_{\mathsf{one}}, f_{\mathsf{even}}, f_{\mathsf{EQ}}\}$ since $h^*(\mathbf{0}) = 1$ and $h^*(\mathbf{1}) = 1$. If $h^* = f_{\mathsf{one}}$, we have that $h(\mathbf{x}) = 1$ for all $\mathbf{x} \in \{0,1\}^n$. Since $h(\mathbf{x}) = f_{S \to 1}$, we obtain that $T \in \Omega_f$ for all $T \supseteq S$.

If $h^* \in \{f_{\mathsf{even}}, f_{\mathsf{EQ}}\}$, then h supports perfect equality by Lemmas 6.4.5 and 6.4.6. Since f supports perfect pinning-to-1, by Item 1 of Lemma 6.4.8 we obtain that f supports perfect equality as well.

This concludes the proof. □

Lemma 6.4.12. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f \neq f_{\mathsf{allone}}$ and $f^* = f_{\mathsf{allone}}$. Then at least one of the four following propositions is true:*

1. *f is semi-trivial;*

2. *there exists $t \in [k]$ such that $f_{t \to 1}$ is not affine;*

3. *f supports perfect equality;*

4. *f simulates a hard function.*

Proof. If $k = 2$, since $f \neq f_{\mathsf{allone}}$ and $f^* = f_{\mathsf{allone}}$, we have $f(1,1) = 1$, $f(0,0) = 0$ and exactly one of $f(0,1)$ and $f(1,0)$ is one, so f is semi-trivial. Thus, for the rest of the proof we may assume that $k \geq 3$.
Since $f^* = f_{\mathsf{allone}}$, we have that $f(\mathbf{1}) = 1$ and $f(\mathbf{0}) = 0$. Further, since $f \neq f_{\mathsf{allone}}$, there exists $S \in \Omega_f$ with $|S| < k$.

Case 1. Every $S \in \Omega_f$ satisfies $|S| \geq k-1$.

If there is only one set S in Ω_f with $|S| = k-1$, then we have that f is semi-trivial (since $f(\mathbf{1}) = 1$). Otherwise, there are distinct sets $S, S' \in \Omega_f$ with $|S| = |S'| = k-1$, so $|S \cap S'| = k-2$ and thus $S \cap S' \neq \emptyset$ and $S \cap S' \notin \Omega_f$. Let $t \in S \cap S'$. We claim that $h = f_{t \to 1}$ is not affine; to see this, note that $f(\chi_S) = f(\chi_{S'}) = f(\chi_{[k]}) = 1$ and $f(\chi_S \oplus \chi_{S'} \oplus \chi_{[k]}) = f(\chi_{S \cap S'}) = 0$. Since $h = f_{t \to 1}$ and $t \in S \cap S'$, we obtain that

$$S \backslash \{t\},\ S' \backslash \{t\},\ [k] \backslash \{t\} \in \Omega_h \text{ but } (S \cap S') \backslash \{t\} \notin \Omega_h.$$

By Item 1 of Lemma 6.3.1, it thus follows that h is not affine, as wanted.

Case 2. There exists $S \in \Omega_f$ with $|S| \leq k-2$.

Let S be a set in Ω_f with the smallest cardinality among the sets in Ω_f. By Lemma 6.4.11, either f satisfies proposition 3 or 4, in the statement of the lemma (so we are finished), or every $Q \supseteq S$ satisfies $Q \in \Omega_f$. Thus, for the rest of the proof we may assume that for every $Q \supseteq S$ it holds that $Q \in \Omega_f$.

Let $\Psi = \{W \in \Omega_f \mid S \setminus W \neq \emptyset\}$. If Ψ is empty then f is semi-trivial, so it satisfies proposition 1 in the statement of the lemma (and we are finished). So assume that Ψ is non-empty and choose $T \in \Psi$ with cardinality as small as possible.

By the choice of S, T cannot be a strict subset of S, so $T \setminus S$ is not empty. Applying Lemma 6.4.11 to the set T, we may assume that $\forall\, Q \supseteq T$ it holds that $Q \in \Omega_f$ (otherwise, f will satisfy proposition 3 or 4, in the statement of the lemma, so we are finished). Since $f(\mathbf{0}) = 0$ and S has minimum cardinality among the sets in Ω_f, we have $1 \leq |S| \leq |T|$.

Case $2a$. $|T| = 1$, which implies $|S| = 1$. Suppose $S = \{s\}$ and $T = \{t\}$. Consider a set $Q \subseteq [k]$ with $|Q| = k-2$. By the above assumptions, we have that if $s \in Q$ or $t \in Q$ then $Q \in \Omega_f$. This accounts for all but one sets $Q \subseteq [k]$ with $|Q| = k-2$; for the remaining set $Q = [k] \backslash \{s,t\}$, it must be the case that $Q \notin \Omega_f$, otherwise all sets Q with $|Q| = k-2$ are in Ω_f, which contradicts the fact that $f^* = f_{\mathsf{allone}}$. Now let's consider. The number of sets $W \in \Omega_f$ which contain both s and t is 2^{k-2}. Similarly, the number of sets $W \in \Omega_f$ which contain s but not t is 2^{k-2} and the number of sets $W \in \Omega_f$ which contain t but not s is 2^{k-2}. But the number of sets $W \in \Omega_f$ which contain neither s

nor t is less than 2^{k-2}. So the k-tuple hypergraph with the single hyperarc $(v_1, \ldots, v_k)$ induces a hard function on the two vertices v_s and v_t and therefore f simulates a hard function.

Case $2b$. $|T| \geq 2$ and $S \cap T \neq \emptyset$. Since $S \setminus T \neq \emptyset$, we have $|S| > |S \cap T|$ and thus $S \cap T \notin \Omega_f$ by the minimality of S. Let $r \in S \cap T$. Now we know that $S \in \Omega_f, T \in \Omega_f$ and $S \cup T \in \Omega_f$ by the assumptions above. But $S \cap T \notin \Omega_f$ and $\chi_S \oplus \chi_T \oplus \chi_{S \cup T} = \chi_{S \cap T}$, so by Item 1 of Lemma 6.3.1, $f_{r \to 1}$ is not affine.

Case $2c$. $|T| \geq 2$ and $S \cap T = \emptyset$. Since $T \setminus S \neq \emptyset$, let $r \in T \setminus S$. By the above assumptions, we have that $S \cup \{r\}$ and $S \cup T$ are in Ω_f. Note that $\{r\} \notin \Omega_f$; otherwise, we would obtain a contradiction to the choice of the set T, since $T' = \{r\}$ satisfies $T' \in \Omega_f$, $S \setminus T' = S \neq \emptyset$ and $|T'| < |T|$. Now we know that $S \cup \{r\}, T, S \cup T \in \Omega_f$ and $\{r\} \notin \Omega_f$. Note that since $S \cap T = \emptyset$, it holds that $\chi_{S \cup \{r\}} \oplus \chi_T \oplus \chi_{S \cup T} = \chi_{\{r\}}$, so by Item 1 of Lemma 6.3.1 we have that $f_{r \to 1}$ is not affine.

This concludes the proof of Lemma 6.4.12. □

Similarly, by switching the spins 0 and 1, we obtain the following lemma when $f^* = f_{\mathsf{allzero}}$.

Lemma 6.4.13. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f \neq f_{\mathsf{allzero}}$ and $f^* = f_{\mathsf{allzero}}$. Then at least one of the four following propositions is true:*

1. *f is semi-trivial;*
2. *there exists $t \in [k]$ such that $f_{t \to 0}$ is not affine;*
3. *f supports perfect equality;*
4. *f simulates a hard function.*

Proof. Suppose f is a Boolean function such that $f^* = f_{\mathsf{allzero}}$ and $f \neq f_{\mathsf{allzero}}$. Let g be the function defined by $g(\mathbf{x}) = f(\overline{\mathbf{x}})$ for all $\mathbf{x} \in \{0,1\}^k$. Now it holds that $g^* = f_{\mathsf{allone}}$ and $g^* \neq f_{\mathsf{allone}}$. So g satisfies one of the four propositions in Lemma 6.4.12. We then have

1. If g is semi-trivial, f is semi-trivial.
2. If $g_{t \to 1}$ is not affine for some $t \in [k]$, $f_{t \to 0}$ is not affine either.

3. If g supports perfect equality, f supports perfect equality too.

4. If g simulates a hard function, f simulates the bitwise complement of the hard function, which is also hard. □

For every function f such that $f^* = f_{\mathsf{zero}}$ and $f \neq f_{\mathsf{zero}}$, we still have a similar reduction lemma, but the proof is more complicated.

Lemma 6.4.14. *Let $k \geq 3$ and let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f^* = f_{\mathsf{zero}}$ and $f \neq f_{\mathsf{zero}}$. Let $S \in \Omega_f$. Then, at least one of the four following propositions is true:*

1. *$h = f_{\overline{S}\to 0}$ is semi-trivial;*

2. *there exists $T \subseteq [k]$ such that $f_{\overline{S}\to 0, T\to 1}$ is not affine;*

3. *f supports perfect equality;*

4. *f simulates a hard function.*

Proof. By Lemma 6.4.4, we have that either f supports perfect equality or f supports both perfect pinning-to-0 and perfect pinning-to-1. We assume that the latter holds (otherwise we are done).
Let $h = f_{\overline{S}\to 0}$. Since $f(\mathbf{0}) = f^*(\mathbf{0}) = 0$ and $S \in \Omega_f$, we have that $h(\mathbf{0}) = 0$ and $h(\mathbf{1}) = 1$. Note that h has arity $q := |S|$. We may assume that $q > 1$; otherwise, h is semi-trivial (proposition 1 in the statement of the lemma). There are two cases to consider: $h^* \notin \mathsf{EASY}(q)$ or $h^* \in \mathsf{EASY}(q)$.

- Case 1. $h^* \notin \mathsf{EASY}(q)$.

  Case 1a. $q = 2$**.**

  In this case, $h^*(0,0) = 0$ and $h^*(0,1) = h^*(1,0) = h^*(1,1) = 1$, so $h^* = \mathsf{OR}$ which is a hard function. We have already assumed (in the first line of the proof) that f supports perfect pinning-to-0. Also, by definition, h perfectly simulates itself. By Item 3 of Lemma 6.4.8, f perfectly simulates h as well, so f simulates a hard function (proposition 4 in the statement of the lemma).

  Case 1b. $q > 2$**.**

By Lemma 6.3.6, either h simulates a hard function or h supports perfect equality (or both). If h simulates a hard function then by Item 3 of Lemma 6.4.8, f simulates the same hard function (proposition 4 in the statement of the lemma). On the other hand, if h supports perfect equality then by Item 1 of Lemma 6.4.8 so does f (proposition 3 in the statement of the lemma).

- Case 2. $h^* \in \mathsf{EASY}(q)$. Since $h(\mathbf{0}) = 0$ and $h(\mathbf{1}) = 1$, we have that h^* is f_{odd} or f_{allone}.

  Case 2a. $h^* = f_{\mathsf{odd}}$. By Lemma 6.4.6, h supports perfect equality and thus f supports perfect equality by Item 1 of Lemma 6.4.8 (proposition 3 in the statement of the lemma).

  Case 2b. $h^* = f_{\mathsf{allone}}$. If $h = f_{\mathsf{allone}}$, h is semi-trivial (proposition 1 in the statement of the lemma). Otherwise, note that $q > 1$, so by Lemma 6.4.12, h is semi-trivial (proposition 1 in the statement of the lemma), or there exists $t \in [k]$ such that $h_{t\to 1}$ is not affine or h supports perfect equality or h simulates a hard function. If there exists $t \in [k]$ such that $h_{t\to 1}$ is not affine then taking $T = \{t\}$, f satisfies proposition 2 in the statement of the lemma. Finally, if h supports perfect equality then so does f (like Case 2a) and if h simulates a hard function, then so does f (like Case 1b). □

Lemma 6.4.15. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a k-ary Boolean function. Suppose that $f \neq f_{\mathsf{zero}}$ and $f^* = f_{\mathsf{zero}}$. Then at least one of the four following propositions is true:*

1. *f is affine;*
2. *there exist $S, T \subseteq [k]$ such that $f_{S\to 0, T\to 1}$ is not affine;*
3. *f supports perfect equality;*
4. *f simulates a hard function.*

Proof. If $k = 2$, we have $f(0,0) = f(1,1) = 0$, so $|\Omega_f| \leq 2$ and thus f is affine (cf. Item 1 of Lemma 6.3.1) so it satisfies proposition 1 in the statement of the lemma.
Now suppose $k \geq 3$. By Lemma 6.4.14, we can assume that for all $W \in \Omega_f$, $f_{\overline{W}\to 0}$ is a semi-trivial function (otherwise f satisfies at least one of propositions 2, 3 or 4).

Choose $S \in \Omega_f$ such that $|S|$ is as large as possible. Let $h = f_{\overline{S}\to 0}$. Since h is semi-trivial (by taking $W = S$ above), we claim that there is a T satisfying $\emptyset \subset T \subseteq S$ such that $\Omega_h = \{U \mid T \subseteq U \subseteq S\}$. (To see this, note that the definition of semi-trivial implies that there is a subset T of S such that either $\Omega_h = \{U \mid U \subseteq T\}$ or $\Omega_h = \{U \mid T \subseteq U \subseteq S\}$. The former is impossible since $\emptyset \notin \Omega_h$ since $h(\mathbf{0}) = f(\mathbf{0})$ and $f(\mathbf{0}) = 0$ since $f^* = f_{\mathsf{zero}}$. Also, in the latter case, T is not empty because, once again, $\emptyset \notin \Omega_h$.)

Case 1. *Suppose that* $\forall X \in \Omega_f$, $T \subseteq X$*:* Recall that T is non-empty. Also, for every $i \in T$, $\Omega_{f_{i\to 1}} = \Omega_f$ so either f is affine (proposition 1 in the statement of the lemma) or $f_{i\to 1}$ is not affine (proposition 2 in the statement of the lemma).

Now, if Case 1 does not hold then there is an $X \in \Omega_f$ such that $T \setminus X$ is non-empty. Since $\Omega_h = \{U \mid T \subseteq U \subseteq S\}$ we conclude that $X \notin \Omega_h$. Since $h = f_{\overline{S}\to 0}$ we conclude that $X \setminus S$ is non-empty. Thus, the only other case to consider is as follows.

Case 2. *Suppose that there is an* $X \in \Omega_f$ *such that* $T \setminus X$ *and* $X \setminus S$ *are both non-empty:* Let $\Psi = \{X \in \Omega_f \mid T \setminus X \neq \emptyset \text{ and } X \setminus S \neq \emptyset\}$. Let $a = \min\{|T \setminus X| : X \in \Psi\}$, and $b = \min\{|X \setminus S| : X \in \Psi \text{ and } |T \setminus X| = a\}$. Choose $R \in \Psi$ with $|T \setminus R| = a$ and $|R \setminus S| = b$.

Now before proceeding, we use the sets S, T and R to partition $[k]$.

$$A = \{i \in [k] \mid i \in S, i \in T, i \notin R\},$$
$$B = \{i \in [k] \mid i \in S, i \in T, i \in R\},$$
$$C = \{i \in [k] \mid i \in S, i \notin T, i \notin R\},$$
$$D = \{i \in [k] \mid i \in S, i \notin T, i \in R\},$$
$$E = \{i \in [k] \mid i \notin S, i \notin T, i \notin R\},$$
$$F = \{i \in [k] \mid i \notin S, i \notin T, i \in R\}.$$

It is clear from the definitions that the sets A, B, C, D, E and F are disjoint. Also, since $T \subseteq S$, they partition $[k]$. From the definitions, $A = T \setminus R$ and $F = R \setminus S$ so, by the choice of R, A and F are non-empty. Let $g = f_{C\cup E\to 0, B\cup D\to 1}$.

By definition, every element of Ω_g is a subset of $A \cup F$. Also, for $Y \subseteq A \cup F$, "$Y \in \Omega_g$" means the same thing as "$Y \cup B \cup D \in \Omega_f$". We establish some facts before dividing the analysis into sub-cases.

Fact 1: $A \in \Omega_g$**.** We have $\Omega_h = \{U \mid T \subseteq U \subseteq S\}$ and $T = A \cup B$ so $A \cup B \cup D \in \Omega_h$. Since $A \cup B \cup D \subseteq S$, this means $A \cup B \cup D \in \Omega_f$. Equivalently, $A \in \Omega_g$.

Fact 2: $F \in \Omega_g$**.** From the definition of R, $R \in \Omega_f$. Also, $R = B \cup D \cup F$ so $F \cup B \cup D \in \Omega_f$. Equivalently, $F \in \Omega_g$.

Fact 3: If $Y \in \Omega_g$ **then either** $Y \cap A \in \{\emptyset, A\}$ **or** $Y \cap F = \emptyset$ **(or both).** Suppose for contradiction that $\emptyset \subset Y \cap A \subset A$ and $Y \cap F$ is non-empty. Note that $R = B \cup D \cup F$. Let $R' = B \cup D \cup Y$. Note that $T \setminus R = A$ and $T \setminus R' = A \setminus Y \subset A$ so $|T \setminus R'| < |T \setminus R|$. We will show a contradiction to the choice of R by showing that $R' \in \Psi$. First, since $Y \in \Omega_g$, $R' \in \Omega_f$. Also, $T \setminus R' = A \setminus Y$ is non-empty and $R' \setminus S = Y \cap F$ is non-empty.

Fact 4: If $Y \in \Omega_g$ **and** $Y \cap A = \emptyset$ **then** $Y \in \{\emptyset, F\}$**.** Suppose for contradiction that $\emptyset \subset Y \subset F$. As in the proof of Fact 3, let $R' = B \cup D \cup Y$. Note that $T \setminus R = T \setminus R' = A$. Also, $R \setminus S = F$ and $R' \setminus S = Y$ so $|R \setminus S| > |R' \setminus S|$. Once again, we will show a contradiction to the choice of R by showing that $R' \in \Psi$. As in the proof of Fact 3, since $Y \in \Omega_g$, $R' \in \Omega_f$. Also, $T \setminus R'$ is non-empty since $T \setminus R$ is. Finally, $R' \setminus S = Y$, which is non-empty.

Fact 5: If $Y \in \Omega_g$ **and** $Y \cap F = \emptyset$ **then** $Y = A$**.** Since $Y \in \Omega_g$, we have $Y \cup B \cup D \in \Omega_f$. But since $Y \subseteq A$, we have $Y \cup B \cup D \subseteq S$, so $Y \cup B \cup D \in \Omega_h$. Since $\Omega_h = \{U \mid T \subseteq U \subseteq S\}$ we have $T \subseteq Y \cup B \cup D$ so $A \subseteq Y$.

Given Facts 1–5, we have only the following sub-cases.

Case 2a: $\Omega_g = \{A, F\}$**.** In this case, we will show that f supports perfect equality so it satisfies proposition 3 in the statement of the lemma. Using Lemma 6.4.4, we conclude that either f supports perfect equality (in which case we are finished) or f supports perfect pinning-to-0 and also perfect pinning-to-1, which we now assume. Let H_0 be a k-tuple hypergraph, with a vertex u_0 such that $\mu_{f;H_0}(\sigma_{u_0} = 0) = 1$. Let H_1 be a k-tuple hypergraph, with a vertex u_1 such that $\mu_{f;H_1}(\sigma_{u_1} = 0) = 1$. We have already noted that A is non-empty. Suppose, without loss of generality, that $1 \in A$ (otherwise, we simply re-order the arguments of $[k]$). Now let H' be the k-tuple hypergraph with vertices $v_0, v_1, \ldots, v_k$ and hyperarcs $(v_0, v_2, \ldots, v_k)$ and $(v_1, v_2, \ldots, v_k)$. Construct H from H' by doing the following:

- For every $i \in C \cup E$, take a new copy of H_0 and identify vertex u_0 with v_i.
- For every $i \in B \cup D$, take a new copy of H_1 and identify vertex u_1 with v_i.

Now since $\Omega_g = \{A, F\}$, $\mu_{f;H}(\sigma(v_0) = \sigma(v_1) = 0) = \mu_{f;H}(\sigma(v_0) = \sigma(v_1) = 1) = 1/2$. Thus, f supports perfect equality, so we have finished Case 2a.

Case 2b: $\exists Y \in \Omega_g$ such that $Y \cap A = A$ and $Y \cap F$ is non-empty. We will show that f satisfies proposition 2 in the statement of the lemma. Specifically, consider some $t \in A$. We will show that $f_{t\to 1}$ is not affine.

Let $Y' = Y \cap F$ so that $Y = A \cup Y'$. Let

$$S_1 := B \cup D \cup Y = B \cup D \cup A \cup Y', \quad S_2 := A \cup B = T, \quad S_3 := A \cup B \cup C.$$

We claim that $S_1\backslash\{t\}, S_2\backslash\{t\}, S_3\backslash\{t\} \in \Omega_{f_{t\to 1}}$. Since $t \in S_1, S_2, S_3$ (from $t \in A$), the claim will follow by showing that $S_1, S_2, S_3 \in \Omega_f$. Indeed, since $Y \in \Omega_g$, we have that $S_1 \in \Omega_f$. Also, since $S_2 = T$, we have that $S_2 \in \Omega_h$ so $S_2 \in \Omega_f$. Finally, since $T = A \cup B \subseteq S_3 \subseteq A \cup B \cup C \cup D = S$, we have that $S_3 \in \Omega_h$ so $S_3 \in \Omega_f$.

Let $S' := A \cup B \cup C \cup D \cup Y' = S \cup Y'$ and note that $\chi_{S'} = \chi_{S_1} \oplus \chi_{S_2} \oplus \chi_{S_3}$ (see Section 6.4.2 for the relevant notation) by the disjointness of A, B, C, D, E, F. Since Y' is non-empty by assumption, we obtain that S' is not in Ω_f by maximality of S. Note that $t \in S'$, so we have that $S'\backslash\{t\} \notin \Omega_{f_{t\to 1}}$.

To sum up, we have shown that

$$S_1\backslash\{t\}, S_2\backslash\{t\}, S_3\backslash\{t\} \in \Omega_{f_{t\to 1}} \text{ but } S'\backslash\{t\} \notin \Omega_{f_{t\to 1}}$$

Since $\chi_{S_1\backslash\{t\}} \oplus \chi_{S_2\backslash\{t\}} \oplus \chi_{S_3\backslash\{t\}} = \chi_{S'\backslash\{t\}}$, by Item 1 of Lemma 6.3.1, $f_{t\to 1}$ is not affine.

This concludes the proof of Lemma 6.4.15. □

Now we can prove Theorem 6.2.9, which we restate here for convenience.

Theorem 6.2.9. *Let $k \geq 2$ and let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function. Then at least one of three following propositions is true:*

1. *f is affine;*
2. *f supports perfect equality;*
3. *f simulates a hard function.*

Proof. We prove this Theorem by induction on the arity of f.

- $k = 2$. So $R_f \subseteq \{00, 01, 10, 11\}$. If f is not affine, then $|R_f| = 3$.

  If $00 \notin R_f$ or $11 \notin R_f$, let G be a graph with two vertices u and v and an edge (u, v). Then either $\mu_{f;G}(\sigma_u = 1, \sigma_v = 1) = \mu_{f;G}(\sigma_u = 0, \sigma_v = 1) = \mu_{f;G}(\sigma_u = 1, \sigma_v = 0) = \frac{1}{3}$ or $\mu_{f;G}(\sigma_u = 0, \sigma_v = 0) = \mu_{f;G}(\sigma_u = 0, \sigma_v = 1) = \mu_{f;G}(\sigma_u = 1, \sigma_v = 0) = \frac{1}{3}$. So f simulates a hard function.

  If $01 \notin R_f$ or $10 \notin R_f$, f^* will be f_{EQ} and thus f supports perfect equality by Lemma 6.4.5.

- $k \geq 3$. Suppose that for all $2 \leq k' < k$, all k'-ary functions f' satisfy at least one of the three propositions in the statement. We now prove that an arbitrary $f : \{0, 1\}^k \to \{0, 1\}$ satisfies at least one of the propositions as well. If f is affine, then it satisfies proposition 1 in the statement of the lemma, so we assume that f is not affine, so $f \notin \mathsf{EASY}(k)$. We have the following case analysis.

  Case 1. $f^* \notin \mathsf{EASY}(k)$. By Lemma 6.3.6, f^* either simulates a hard function in which case f simulates the same hard function as well or f^* supports perfect equality in which case f supports perfect equality as well.

  Case 2. $f^* \in \mathsf{EASY}(k)$. There are 6 sub-cases to consider:

  Case $2a$. $f^* = f_{\mathsf{EQ}}$. By Lemma 6.4.5, f supports perfect equality.

  Case $2b$. $f^* = f_{\mathsf{odd}}$ or $f^* = f_{\mathsf{even}}$. By Lemma 6.4.6, f supports perfect equality.

  Case $2c$. $f^* = f_{\mathsf{allone}}$. By Lemma 6.4.2, f supports perfect pinning-to-1. By Lemma 6.4.12, f is semi-trivial (and thus f is affine), or f supports perfect equality or simulates a hard function, or there exists $t \in [k]$ such that $f_{t\to 1}$ is not affine. If $f_{t\to 1}$ is not affine for some $t \in [k]$, $f_{t\to 1}$ must support perfect equality or simulate a hard function by the induction hypothesis. So f supports or simulates the same function by Lemma 6.4.8.

  Case $2d$. $f^* = f_{\mathsf{allzero}}$. The proof for this case is completely analogous to the case $3.c$ by switching the spins 0 and 1 (cf. Lemma 6.4.13).

  Case $2e$. $f^* = f_{\mathsf{one}}$. $f^* = f_{\mathsf{one}}$ means $f(\mathbf{x}) = 1$ for all $\mathbf{x} \in \{0, 1\}^k$, so f is affine.

  Case $2f$. $f^* = f_{\mathsf{zero}}$. By Lemma 6.4.4, we have that either f supports perfect equality or f supports both perfect pinning-to-0 and perfect pinning-to-1. We assume that the latter is the case (otherwise we are done). By

Lemma 6.4.15, f is affine, or f supports perfect equality or simulates a hard function, or there exists some $S, T \subseteq [k]$ such that $f_{S\to 0, T\to 1}$ is not affine. The only case where we aren't immediately finished is the final one. In this case, the arity of $f_{S\to 0,T\to 1}$ must be at least 2 since every unary function is affine. Thus, since $f_{S\to 0,T\to 1}$ is not affine, it must support perfect equality or simulate a hard function g by the induction hypothesis. Then, f either supports perfect equality or simulates the hard function g by Lemma 6.4.8.

This concludes the case analysis and the proof of Theorem 6.2.9. □

6.5 The case where f supports perfect equality

In this section, we assume that f is not affine but that it supports perfect equality. In this case, due to the presence of perfect equality, we will be able to employ results and techniques from [26] to show Theorem 6.2.10.

6.5.1 Constraint Satisfaction Problems and implementations

In the introduction to this paper, we illustrated how Boolean relations can implement more complicated interactions by considering the "not-all-equal" relation of arity 3 and using it to "implement" ferromagnetic Ising interactions. At this point, it is useful to make the notion of "implement" more precise. There are various notions in the literature of *implementations*. We use (a generalisation of) the notion from [26], which is essentially the "faithful, perfect" variant of "implementation" from [18].

Definition 6.5.1. Let Γ be a Boolean constraint language. The language Γ *implements* a t-ary function $g : \{0,1\}^t \to \mathbb{R}_{\geq 0}$, if for some $t' \geq t$ there is a CSP instance I with variables $x_1, \ldots, x_{t'}$ and constraint language Γ such that for every tuple $(s_1, \ldots, s_t) \in \{0,1\}^t$, there are precisely $g(s_1, s_2, \ldots, s_t)$ satisfying assignments σ of I with $\sigma(x_1) = s_1, \ldots, \sigma(x_t) = s_t$.

When f supports perfect equality, we will use the following "transitivity" lemma, which will allow us to use some known implementations.

Lemma 6.5.2. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function which supports perfect equality. Let Γ be a Boolean constraint language and let $g : \{0,1\}^t \to \mathbb{R}_{\geq 0}$ be a t-ary*

function such that g is not $f_{\mathsf{zero}}^{(t)}$ and Γ implements g. Then, if f perfectly simulates Γ, f also perfectly simulates the function g.

Proof. Since Γ implements g, there exists some $t' \geq t$ and a CSP instance I with variables $X := \{x_1, \ldots, x_{t'}\}$ and constraints in Γ such that for every tuple $(s_1, \ldots, s_t) \in \{0,1\}^t$, there are precisely $g(s_1, s_2, \ldots, s_t)$ satisfying assignments σ of I with $\sigma(x_1) = s_1, \ldots, \sigma(x_t) = s_t$. Since $g \neq f_{\mathsf{zero}}^{(t)}$, we conclude that, for all $s_1, \ldots, s_t \in \{0,1\}$,

$$\mu_I(\sigma(x_1) = s_1, \ldots, \sigma(x_t) = s_t) = \frac{g(s_1, s_2, \ldots, s_t)}{\sum_{(s'_1, s'_2, \ldots, s'_t) \in \{0,1\}^t} g(s'_1, s'_2, \ldots, s'_t)}. \tag{6.4}$$

Since f supports perfect equality, there exists a k-tuple hypergraph $H_{\mathsf{eq}} = (V_{\mathsf{eq}}, \mathcal{F}_{\mathsf{eq}})$ and vertices $y, z \in V_{\mathsf{eq}}$ such that

$$\mu_{f;H_{\mathsf{eq}}}(\sigma(y) = \sigma(z) = 0) = \mu_{f;H_{\mathsf{eq}}}(\sigma(y) = \sigma(z) = 1) = 1/2. \tag{6.5}$$

We will use the CSP instance I and the hypergraph H_{eq} to construct a k-tuple hypergraph $H = (V, \mathcal{F})$ with vertices $v_1, \ldots, v_{t'}$ in V satisfying

$$\mu_{f;H}(\sigma(v_1) = s_1, \ldots, \sigma(v_{t'}) = s_{t'}) = \mu_I(\sigma(x_1) = s_1, \ldots, \sigma(x_{t'}) = s_{t'}) \tag{6.6}$$

for all $s_1, \ldots, s_{t'} \in \{0,1\}$. From this, the lemma follows since we can sum over the values of $s_{t+1}, \ldots, s_{t'} \in \{0,1\}$ to obtain that

$$\mu_{f;H}(\sigma(v_1) = s_1, \ldots, \sigma(v_t) = s_t) = \mu_I(\sigma(x_1) = s_1, \ldots, \sigma(x_t) = s_t)$$

for all $s_1, \ldots, s_t \in \{0,1\}$, which in conjunction with (6.4) yields that f perfectly simulates g.

To formally construct the k-tuple hypergraph H, we will need some notation. Suppose that I has m constraints and for $j \in [m]$ write the j'th constraint as $f_j(x_{j,1}, \ldots, x_{j,w(j)})$, where $w(j)$ is the arity of $f_j \in \Gamma$ and, for all $i \in [w(j)]$, $x_{j,i} \in \{x_1, \ldots, x_{t'}\}$. Since f perfectly simulates Γ and every f_j is in Γ, for every constraint $C_j = f_j(x_{j,1}, \ldots, x_{j,w(j)})$, there is a k-tuple hypergraph $H_j = (V_j, \mathcal{F}_j)$ and vertices $v_{j,1}, \ldots, v_{j,w(j)}$ of H_j such that for all $s_1, \ldots, s_{w(j)} \in \{0,1\}$, it holds that

$$\mu_{f;H_j}(\sigma(v_{j,1}) = s_1, \ldots, \sigma(v_{j,w(j)}) = s_{w(j)}) = \frac{f_j(s_1, \ldots, s_{w(j)})}{|R_{f_j}|}. \tag{6.7}$$

Note that the expression $|R_{f_j}|$ in the denominator in (6.7) is not zero because the constraint C_j has a satisfying assignment, since I does (which follows from the fact that $g \neq f_{\mathsf{zero}}^{(t)}$ and from the definition of I).

Consider now the k-tuple hypergraph $H' = (V', \mathcal{F}')$ which is simply the disjoint union of $H_1, \ldots, H_m$ (i.e., $V' = \cup_{j=1}^{m} V_j$ and $\mathcal{F}' = \cup_{j=1}^{m} \mathcal{F}_j$). Note that, for every subset $S \subseteq V'$ and every assignment $\tau : S \to \{0,1\}$, it holds that

$$\mu_{f;H'}(\sigma_S = \tau) = \prod_{j=1}^{m} \mu_{f;H_j}(\sigma_{S \cap V_j} = \tau_{S \cap V_j}). \tag{6.8}$$

To complete the construction of the desired H, we need some further notation. For a variable $x_i \in \{1, \ldots, x_{t'}\}$ of the CSP instance I, let $U_i \subseteq V'$ denote the subset of vertices of H' which correspond to occurrences of the variable x_i in the CSP instance I. More precisely, assume that the variable x_i has d occurrences in I for some integer $d \geq 1$, and let $C_{j_1}, \ldots, C_{j_d}$ be the constraints in which x_i appears (note that the indices $j_1, \ldots, j_d$ are not necessarily distinct). Further, let $t_1, \ldots, t_d$ denote the indices of the positions where x_i appears in $C_{j_1}, \ldots, C_{j_d}$ respectively. Then $U_i := \{v_{j_1,t_1}, \ldots, v_{j_d,t_d}\}$ is precisely the subset of vertices of H' which correspond to occurrences of the variable x_i in the CSP instance I. Let $U := \cup_{i \in [t']} U_i$ (note that in general $U \neq V'$ since H' may contain vertices that do not correspond to occurrences of variables of I).

We are now ready to complete the construction of the desired k-tuple hypergraph $H = (V, \mathcal{F})$. Start by setting H equal to H'. Then, for each $i \in [t']$ and each pair of vertices $u, u' \in U_i$, add to H a distinct copy of the k-tuple hypergraph H_{eq}, identifying the vertices y and z of H_{eq} with the vertices u and u'. Having defined H, we next choose the specified vertices $v_1, \ldots, v_{t'}$. In fact, it suffices, for each $i \in [t]$, to let v_i be an arbitrary vertex in U_i.

It remains to prove that (6.6) holds. We call an assignment $\tau : U \to \{0,1\}$ *relevant* if for every $i \in [t']$ there exists $s_i \in \{0,1\}$ such that for every vertex $v \in U_i$, it holds that $\tau(v) = s_i$. For relevant assignments τ, we will refer to the tuple $(s_1, \ldots, s_{t'})$ as the CSP assignment corresponding to τ. For non-relevant τ, the copies of H_{eq} on top of the sets $U_1, \ldots, U_{t'}$ ensure that $\mu_{f;H}(\sigma_U = \tau) = 0$. For all relevant $\tau : U \to \{0,1\}$, we have from (6.5) that

$$\mu_{f;H}(\sigma_U = \tau) = \mu_{f;H'}(\sigma_U = \tau)$$

and, hence, using (6.8), we have that

$$\mu_{f;H}(\sigma_U = \tau) = \prod_{j=1}^{m} \mu_{f;H_j}(\sigma_{U\cap V_j} = \tau_{U\cap V_j}). \tag{6.9}$$

Note that for every $j \in [m]$ we have $U \cap V_j = \{v_{j,1}, \ldots, v_{j,w(j)}\}$ and, hence, (6.7) gives

$$\mu_{f;H_j}(\sigma_{U\cap V_j} = \tau_{U\cap V_j}) = \frac{f_j(\tau(v_{j,1}), \cdots, \tau(v_{j,w(j)}))}{|R_{f_j}|}. \tag{6.10}$$

Let $\propto$ be the notation that means "in proportion to". Then (6.10) follows from (6.9) and (6.10) that

$$\mu_{f;H}(\sigma_U = \tau) \propto \prod_{j=1}^{m} f_j(\tau(v_{j,1}), \cdots, \tau(v_{j,w(j)})) \text{ for all relevant } \tau. \tag{6.11}$$

For a relevant $\tau : U \to \{0,1\}$, let $(s_1, \ldots, s_{t'})$ be the CSP assignment corresponding to τ. Then, the product in the r.h.s. of (6.11) is 1 iff $(s_1, \ldots, s_{t'})$ encodes a satisfying assignment of the CSP instance I. Since the relevant $\tau : U \to \{0,1\}$ and assignments to the CSP instance I are in 1-1 correspondence, we obtain (6.6), as wanted. This concludes the proof of Lemma 6.5.2. □

Lemma 6.5.3. *Let $f : \{0,1\}^k \to \{0,1\}$ and $g : \{0,1\}^t \to \{0,1\}$ be Boolean functions such that f simulates g, and $g \neq f^{(t)}_{\mathsf{zero}}$. Suppose that g supports pinning-to-s for some $s \in \{0,1\}$. Then f supports pinning-to-s as well.*

Proof. Without loss of generality, we assume that $s = 0$. Suppose that the function g supports pinning-to-0. Our goal is to show that f supports pinning-to-0 as well.
First, let $Z_g := \sum_{(s_1,s_2,\ldots,s_t)\in\{0,1\}^t} g(s_1, s_2, \ldots, s_t)$. Since $g \neq f^{(t)}_{\mathsf{zero}}$, $Z_g > 0$. Since g supports pinning-to-0, by Definition 6.2.3, there exists a t-tuple hypergraph $H_0 = (V_0, \mathcal{F}_0)$ and a vertex $v_0 \in V_0$ such that

$$\mu_{g;H_0}(\sigma_{v_0} = 0) \geq 9/10. \tag{6.12}$$

(The choice of the constant $9/10$ is arbitrary, any constant greater than $1/2$ would work. Also, $Z_{g;H_0} > 0$.) For all $\eta : V_0 \to \{0,1\}$ define $A_\eta := \frac{w_{g;H_0}(\eta)}{(Z_g)^{|\mathcal{F}_0|}}$ and define $M := \sum_{\eta:V_0\to\{0,1\}} A_\eta = \frac{Z_{g;H_0}}{(Z_g)^{|\mathcal{F}_0|}}$. Since $Z_{g;H_0}$ and Z_g are positive, $M > 0$. Also,

$$\mu_{g;H_0}(\eta) = \frac{A_\eta}{M}. \tag{6.13}$$

Now let $\varepsilon := \min\{M/8, 1/8\}$, $\varepsilon_1 := \varepsilon/(|\mathcal{F}_0|\, 2^{2|V_0|})$ and $\varepsilon_2 := \varepsilon/(2^{|V_0|})$. Since f simulates the function g, by Definition 6.2.7 and Lemma 6.3.5, there exists a k-tuple hypergraph $H_g = (V_g, \mathcal{F}_g)$ and t vertices $v_1, v_2, \ldots, v_t$ of H_g such that, for all $(s_1, s_2, \ldots, s_t) \in \{0,1\}^t$,

$$\left|\mu_{f;H_g}(\sigma(v_1) = s_1, \sigma(v_2) = s_2, \ldots, \sigma(v_t) = s_t) - \frac{g(s_1, s_2, \ldots, s_t)}{Z_g}\right| \leq \varepsilon_1. \tag{6.14}$$

Construct the k-tuple hypergraph $H = (V, \mathcal{F})$ as follows. For every hyperarc e of H_0, say $e = (u_1, \ldots, u_t) \in \mathcal{F}_0$, take a distinct copy of H_g, which we will denote by $H_g^{(e)}$, and identify the vertices $u_1, \ldots, u_t \in V_0$ with the vertices $v_1, \ldots, v_t$ of H_g. Note that H is a union of copies of H_g which intersect only at vertices in V_0. Now for all $\eta : V_0 \to \{0,1\}$ define $A'_\eta := \prod_{e\in\mathcal{F}_0} \mu_{f;H_g^{(e)}}(\sigma_e = \eta_e)$ and $M' := \sum_{\eta:V_0\to\{0,1\}} A'_\eta$. Then

$$\mu_{f;H}(\sigma_{V_0} = \eta) = \frac{A'_\eta}{M'}. \tag{6.15}$$

By (6.14), for every $e = (u_1, \ldots, u_t) \in \mathcal{F}_0$, it holds that

$$\left|\mu_{f;H_g^{(e)}}(\sigma_e = \eta_e) - \frac{g\big(\eta(u_1), \ldots, \eta(u_t)\big)}{Z_g}\right| \leq \varepsilon_1. \tag{6.16}$$

Recall that for real numbers $a_1, \ldots, a_n \in [0,1]$ and $b_1, \ldots, b_n \in [0,1]$, it holds that $|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i| \leq \sum_{i=1}^n |a_i - b_i|$. Thus, using (6.16), we obtain that, for every $\eta : V_0 \to \{0,1\}$, it holds that

$$|A'_\eta - A_\eta| = \left|\prod_{e\in\mathcal{F}_0} \mu_{f;H_g^{(e)}}(\sigma_e = \eta_e) - \frac{w_{g;H_0}(\eta)}{(Z_g)^{|\mathcal{F}_0|}}\right| \leq \varepsilon_1 |\mathcal{F}_0|. \tag{6.17}$$

Summing this over all $\eta : V_0 \to \{0,1\}$, we obtain that

$$|M' - M| = \left|\sum_{\eta:V_0\to\{0,1\}} A'_\eta - \sum_{\eta:V_0\to\{0,1\}} A_\eta\right| \leq \varepsilon_1 |\mathcal{F}_0| 2^{|V_0|} = \varepsilon_2. \tag{6.18}$$

Note that the expression $\varepsilon_1|\mathcal{F}_0|$ in (6.17) is at most ε_2. Also, for all $\eta : V_0 \to \{0,1\}$, $A_\eta \leq M$ and $A'_\eta \leq M'$. The bounds in (6.17) and (6.18) yield that $A_\eta - \varepsilon_2 \leq A'_\eta \leq A_\eta + \varepsilon_2$ and $M - \varepsilon_2 \leq M' \leq M + \varepsilon_2$. By the choice of ε, we have $M - \varepsilon > M/2$ and hence $M - \varepsilon_2 > M/2$ as well. Further, we have the bound

$$\left|\frac{A_\eta}{M} - \frac{A'_\eta}{M'}\right| \leq \max\left\{\frac{A_\eta}{M} - \frac{A_\eta - \varepsilon_2}{M + \varepsilon_2}, \frac{A_\eta + \varepsilon_2}{M - \varepsilon_2} - \frac{A_\eta}{M}\right\} \leq \frac{\varepsilon_2(A_\eta + M)}{M(M - \varepsilon_2)} \leq \frac{2\varepsilon_2}{M - \varepsilon_2} \leq \frac{4\varepsilon_2}{M} \leq \frac{1}{2^{|V_0|+1}}. \tag{6.19}$$

From (6.13) and (6.15) and (6.19), we thus obtain that for every $\eta : V_0 \to \{0,1\}$, it holds that

$$|\mu_{f;H}(\sigma_{V_0} = \eta) - \mu_{g;H_0}(\eta)| \leq 1/2^{|V_0|+1}.$$

Summing this over the $2^{|V_0|-1}$ possible values of $\eta_{V_0 \setminus \{v_0\}}$, we obtain that for $s \in \{0,1\}$ it holds that

$$|\mu_{f;H}(\sigma_{v_0} = s) - \mu_{g;H_0}(\sigma_{v_0} = s)| \leq 1/4.$$

Combining this with (6.12), we obtain that

$$\mu_{f;H}(\sigma_{v_0} = 0) > 1/2 > \mu_{f;H}(\sigma_{v_0} = 1).$$

Thus, by Lemma 6.3.4, we obtain that f supports pinning-to-0. This concludes the proof of Lemma 6.5.3. □

The following lemma is similar to Lemma 6.5.2 except that, instead of assuming that f perfectly simulates Γ, we only assume that f simulates Γ so instead of concluding that f perfectly simulates g, we only conclude that f simulates g.

Lemma 6.5.4. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function which supports equality. Let Γ be a Boolean constraint language and let $g : \{0,1\}^t \to \mathbb{R}_{\geq 0}$ be a t-ary function such that g is not $f_{\mathsf{zero}}^{(t)}$ and Γ implements g. Then, if f simulates Γ, f also simulates the function g.*

Proof. The proof is similar to the proof of Lemma 6.5.2, but the imperfect nature of the simulation adds technical details. Since Γ implements g we can follow the proof of Lemma 6.5.2 to define the CSP instance I with variables $\{x_1, \ldots, x_{t'}\}$ and constraints in Γ satisfying (6.4).
We will use the CSP instance I and the fact that f supports equality to construct a k-tuple hypergraph $H = (V, \mathcal{F})$ with an admissible set $\mathcal{V}^*$ for H with respect to f and vertices $v_1, \ldots, v_{t'}$ in V satisfying

$$\mu_{f;H}^{\mathrm{cond}(\mathcal{V}^*)}(\sigma(v_1) = s_1, \ldots, \sigma(v_{t'}) = s_{t'}) = \mu_I(\sigma(x_1) = s_1, \ldots, \sigma(x_{t'}) = s_{t'}) \tag{6.20}$$

for all $s_1, \ldots, s_{t'} \in \{0,1\}$. From this, the lemma follows since we can sum over the values of $s_{t+1}, \ldots, s_{t'} \in \{0,1\}$ to obtain that

$$\mu_{f;H}^{\mathrm{cond}(\mathcal{V}^*)}(\sigma(v_1) = s_1, \ldots, \sigma(v_t) = s_t) = \mu_I(\sigma(x_1) = s_1, \ldots, \sigma(x_t) = s_t)$$

for all $s_1, \ldots, s_t \in \{0,1\}$, which in conjunction with (6.4) yields that f simulates g.

To formally construct the k-tuple hypergraph H, we will need some notation. As in the proof of Lemma 6.5.2, suppose that I has m constraints and for $j \in [m]$ write the j'th constraint as $f_j(x_{j,1}, \ldots, x_{j,w(j)})$, where $w(j)$ is the arity of $f_j \in \Gamma$ and, for all $i \in [w(j)]$, $x_{j,i} \in \{x_1, \ldots, x_{t'}\}$. Since f simulates Γ and every f_j is in Γ, for every constraint $C_j = f_j(x_{j,1}, \ldots, x_{j,w(j)})$, there is a k-tuple hypergraph $H_j = (V_j, \mathcal{F}_j)$, an admissible collection $\mathcal{V}^j = (V^j_{\mathsf{pin0}}, V^j_{\mathsf{pin1}}, \mathcal{V}^j_{\mathsf{eq}})$ for H_j with respect to f_j and vertices $v_{j,1}, \ldots, v_{j,w(j)}$ of H_j such that for all $s_1, \ldots, s_{w(j)} \in \{0,1\}$, it holds that

$$\mu^{\mathrm{cond}(\mathcal{V}^j)}_{f;H_j}(\sigma(v_{j,1}) = s_1, \ldots, \sigma(v_{j,w(j)}) = s_{w(j)}) = \frac{f_j(s_1, \ldots, s_{w(j)})}{|R_{f_j}|}. \tag{6.21}$$

Consider now the k-tuple hypergraph $H = (V, \mathcal{F})$ which is simply the disjoint union of $H_1, \ldots, H_m$ (i.e., $V = \cup_{j=1}^m V_j$ and $\mathcal{F} = \cup_{j=1}^m \mathcal{F}_j$). Further, let $\mathcal{V} = (V_{\mathsf{pin0}}, V_{\mathsf{pin1}}, \mathcal{V}_{\mathsf{eq}})$, where $V_{\mathsf{pin0}} = \cup_{j=1}^m V^j_{\mathsf{pin0}}$, $V_{\mathsf{pin1}} = \cup_{j=1}^m V^j_{\mathsf{pin1}}$ and $V_{\mathsf{eq}} = \cup_{j=1}^m \mathcal{V}^j_{\mathsf{eq}}$. We wish to argue that $\mathcal{V}$ is admissible for H with respect to f. The various disjointness constraints in Definition 6.2.6 are satisfied since $H_1, \ldots, H_m$ are disjoint (using the fact that each $\mathcal{V}^j$ is admissible for H_j with respect to f_j). We have assumed, in the statement of the lemma, that f supports equality. To show that $\mathcal{V}$ is admissible for H with respect to f, we need to show that if some f_j supports pinning-to-s for some $s \in \{0,1\}$ then so does f. This follows from Lemma 6.5.3 since, by assumption, f simulates f_j. Thus, $\mathcal{V}$ is admissible for H with respect to f.

Note that, for every subset $S \subseteq V$ and every assignment $\tau : S \to \{0,1\}$, it holds that

$$\mu^{\mathrm{cond}(\mathcal{V})}_{f;H}(\sigma_S = \tau) = \prod_{j=1}^{m} \mu^{\mathrm{cond}(\mathcal{V}_j)}_{f;H_j}(\sigma_{S \cap V_j} = \tau_{S \cap V_j}). \tag{6.22}$$

Having completed the construction of the desired H, to recover (6.6), it remains to specify $\mathcal{V}^*$ and the vertices $v_1, \ldots, v_{t'}$. For each $i \in [t']$, define U_i as in the proof of Lemma 6.5.2. Also, let $U := \cup_{i \in [t']} U_i$, The main idea is that $\mathcal{V}^*$ is the same as $\mathcal{V}$ except that the sets $U_1, \ldots, U_{t'}$ are added to V_{eq} because we want to condition on the fact that the variables in each of these sets are equal. In order to formally specify $\mathcal{V}^* = (V^*_{\mathsf{pin0}}, V^*_{\mathsf{pin1}}, V^*_{\mathsf{eq}})$ there is a slight technical difficulty because V^*_{pin0} and V^*_{pin1} have to be disjoint from each other and from all sets in V^*_{eq}. In order to deal with this (rather unimportant, but technical) detail, we give an algorithm for defining $\mathcal{V}^*$. Let $\mathcal{V}^0 = \mathcal{V}$. Then, for $i = 1, \ldots, t'$ define $\mathcal{V}^i = (V^i_{\mathsf{pin0}}, V^i_{\mathsf{pin1}}, \mathcal{V}^i_{\mathsf{eq}})$ as follows.

- Let $\mathcal{V}^i = \mathcal{V}^{i-1}$.

- Note that no vertex in U_i is in $V_{\mathsf{pin0}} \cap V_{\mathsf{pin1}}$. This follows since I is satisfiable (since g is not the always-zero function f_{zero}).
- If any vertex in U_i is in V_{pin0} then replace V^i_{pin0} with $V^{i-1}_{\mathsf{pin0}} \cup U_i$.
- Otherwise, if any vertex in U_i is in V_{pin1} then replace V^i_{pin1} with $V^{i-1}_{\mathsf{pin1}} \cup U_i$.
- Otherwise, if U_i does not intersect any sets in $\mathcal{V}^{i-1}_{\mathsf{eq}}$ then replace $\mathcal{V}^i_{\mathsf{eq}}$ with $\mathcal{V}^{i-1}_{\mathsf{eq}} \cup \{U_i\}$.
- Otherwise, let $W_1, \ldots, W_z$ be the sets in $\mathcal{V}^{i-1}_{\mathsf{eq}}$ that intersect U_i and replace $\mathcal{V}^i_{\mathsf{eq}}$ with $(\mathcal{V}^{i-1}_{\mathsf{eq}} \setminus \{W_1, \ldots, W_z\}) \cup \{U_i \cup W_1 \cup \cdots \cup W_z\}$.

Finally, let $\mathcal{V}^* = \mathcal{V}^{t'}$. Then choose the vertices $v_1, \ldots, v_{t'}$ to be arbitrary vertices in $U_1, \ldots, U_{t'}$, respectively.

It remains to prove that (6.20) holds. As in the proof of Lemma 6.5.2, we call an assignment $\tau : U \to \{0,1\}$ *relevant* if for every $i \in [t']$ there exists $s_i \in \{0,1\}$ such that for every vertex $v \in U_i$, it holds that $\tau(v) = s_i$. For relevant assignments τ, we will refer to the tuple $(s_1, \ldots, s_{t'})$ as the CSP assignment corresponding to τ. Clearly, for non-relevant τ, we have that $\mu^{\mathrm{cond}(\mathcal{V}^*)}_{f;H}(\sigma_U = \tau) = 0$ since $\mathcal{V}^*$ forces equality on each of the sets $U_1, \ldots, U_{t'}$. For all relevant $\tau : U \to \{0,1\}$, we have that

$$\mu^{\mathrm{cond}(\mathcal{V}^*)}_{f;H}(\sigma_U = \tau) = \frac{\mu^{\mathrm{cond}(\mathcal{V})}_{f;H}(\sigma_U = \tau)}{\mu^{\mathrm{cond}(\mathcal{V})}_{f;H}(\sigma^{\mathsf{eq}}_{U_1}, \ldots, \sigma^{\mathsf{eq}}_{U_t})}$$

and hence

$$\mu^{\mathrm{cond}(\mathcal{V}^*)}_{f;H}(\sigma_U = \tau) \propto \mu^{\mathrm{cond}(\mathcal{V})}_{f;H}(\sigma_U = \tau) \text{ for all relevant } \tau. \tag{6.23}$$

Using (6.22), we have that

$$\mu^{\mathrm{cond}(\mathcal{V})}_{f;H}(\sigma_U = \tau) = \prod_{j=1}^{m} \mu^{\mathrm{cond}(\mathcal{V}^j)}_{f;H_j}(\sigma_{U \cap V_j} = \tau_{U \cap V_j}). \tag{6.24}$$

Note that for every $j \in [m]$ we have $U \cap V_j = \{v_{j,1}, \ldots, v_{j,w(j)}\}$ and, hence, (6.21) gives

$$\mu^{\mathrm{cond}(\mathcal{V}^j)}_{f;H_j}(\sigma_{U \cap V_j} = \tau_{U \cap V_j}) = \frac{f_j(\tau(v_{j,1}), \cdots, \tau(v_{j,w(j)}))}{|R_{f_j}|}. \tag{6.25}$$

It follows from (6.23), (6.24) and (6.25) that

$$\mu^{\mathrm{cond}(\mathcal{V}^*)}_{f;H}(\sigma_U = \tau) \propto \prod_{j=1}^{m} f_j(\tau(v_{j,1}), \cdots, \tau(v_{j,w(j)})) \text{ for all relevant } \tau. \tag{6.26}$$

For a relevant $\tau : U \to \{0,1\}$, let $(s_1, \ldots, s_{t'})$ be the CSP assignment corresponding to τ. Then, the product in the r.h.s. of (6.26) is 1 iff $(s_1, \ldots, s_{t'})$ encodes a satisfying assignment of the CSP instance I. Since the relevant $\tau : U \to \{0,1\}$ and assignments to the CSP instance I are in 1-1 correspondence, we obtain (6.20), as wanted.
This concludes the proof of Lemma 6.5.4. □

The following Boolean functions, which were considered in [26], will be important in what follows: δ_0 and δ_1 (defined in Definition 6.4.7), and XOR, Implies, NAND, OR. For convenience, we state the corresponding relations here.

- $R_{\delta_0} = \{(0)\}$ and $R_{\delta_1} = \{(1)\}$ (these correspond to satisfying assignments of $\neg x$ and x, respectively).
- $R_{\mathsf{XOR}} = \{(0,1),(1,0)\}$ (corresponds to satisfying assignments of $x \neq y$).
- $R_{\mathsf{Implies}} = \{(0,0),(0,1),(1,1)\}$ (corresponds to satisfying assignments of $x \Rightarrow y$).
- $R_{\mathsf{NAND}} = \{(0,0),(0,1),(1,0)\}$ (corresponds to satisfying assignments of $\neg x \vee \neg y$).
- $R_{\mathsf{OR}} = \{(0,1),(1,0),(1,1)\}$ (corresponds to satisfying assignments of $x \vee y$).

6.5.2 The case of self-dual functions

A Boolean function f is said to be *self-dual* if, for all $\mathbf{x}$, $f(\mathbf{x}) = f(\overline{\mathbf{x}})$. In this section, we show (Theorem 6.5.6, below) that if f is a self-dual Boolean function which is not affine, and f supports perfect equality, then f simulates a hard function. First, we establish a useful lemma.

Lemma 6.5.5. *Let $f : \{0,1\}^k \to \{0,1\}$ be a self-dual function Boolean with $f \neq f^{(k)}_{\mathsf{zero}}$ and $f(\mathbf{0}) = 0$. Further, suppose that f supports perfect equality. Then, f perfectly simulates* XOR.

Proof. From $f(\mathbf{0}) = 0$ and self-duality, we have that $f(\mathbf{1}) = 0$. Since $f \neq f^{(k)}_{\mathsf{zero}}$, there must be some $\mathbf{x} \notin \{\mathbf{0}, \mathbf{1}\}$ such that $f(\mathbf{x}) = 1$. By self-duality, we have that $f(\overline{\mathbf{x}}) = 1$ as well. Let $U_0 = \{i \in [k] \mid x_i = 0\}$ and $U_1 = \{i \in [k] \mid x_i = 1\}$.
Since f supports perfect equality, there exists a k-tuple hypergraph $H_{\mathsf{eq}} = (V_{\mathsf{eq}}, \mathcal{F}_{\mathsf{eq}})$ and vertices $y, z \in V_{\mathsf{eq}}$ such that

$$\mu_{f;H_{\mathsf{eq}}}(\sigma(y) = \sigma(z) = 0) = \mu_{f;H_{\mathsf{eq}}}(\sigma(y) = \sigma(z) = 1) = 1/2.$$

Construct the k-tuple hypergraph H as follows. First, take a single hyperarc $(v_1, \ldots, v_k)$. Then, for every $s \in \{0,1\}$ and every i, j such that i and j are both in U_s, add a new copy of H_{eq}, identifying y with v_i and z with v_j. Finally, choose $v_1 \in U_0$ and $v_2 \in U_1$. Then

$$\mu_{f;H}(\sigma(v_1) = 0, \sigma(v_2) = 1) = \mu_{f;H}(\sigma(v_1) = 1, \sigma(v_2) = 0) = 1/2,$$

so f perfectly simulates XOR. □

Theorem 6.5.6. *Suppose that f is a self-dual Boolean function which is not affine and supports perfect equality. Then f simulates a hard function.*

Proof. Let k be the arity of f. The proof has two cases depending on whether $f(\mathbf{0}) = 0$ or $f(\mathbf{0}) = 1$. We begin with the case where $f(\mathbf{0}) = 1$ (we will reduce the proof for the other case to this one).
So, assume first that $f(\mathbf{0}) = 1$. Since f is not affine, by applying Item 2 of Lemma 6.3.1 to $\mathbf{a} = \mathbf{0}$, we obtain that there exist $\mathbf{b}, \mathbf{c} \in \{0,1\}^k$ such that $f(\mathbf{b}) = f(\mathbf{c}) = 1$ but $f(\mathbf{b} \oplus \mathbf{c}) = 0$. By self-duality, we also have that $f(\overline{\mathbf{b}}) = f(\overline{\mathbf{c}}) = 1$. Note that

$$\mathbf{b} \neq \mathbf{c}, \qquad \mathbf{b} \neq \mathbf{0}, \mathbf{1}, \qquad \mathbf{c} \neq \mathbf{0}, \mathbf{1}.$$

Indeed, it cannot be the case that $\mathbf{b} = \mathbf{c}$ since then $f(\mathbf{b} \oplus \mathbf{c}) = f(\mathbf{0}) = 1$. Analogously, $\mathbf{b} = \mathbf{0}$ would give that $f(\mathbf{b} \oplus \mathbf{c}) = f(\mathbf{c}) = 1$. Similarly, $\mathbf{b} = \mathbf{1}$ would give that $f(\mathbf{b} \oplus \mathbf{c}) = f(\overline{\mathbf{c}}) = 1$. By symmetry between $\mathbf{b}$ and $\mathbf{c}$, we have that $\mathbf{c} \neq \mathbf{0}, \mathbf{1}$.
Let w, x, y, z be Boolean variables. For $i \in [k]$, let

$$r_i = \begin{cases} w, & \text{if } b_i = 0, c_i = 0, \\ x, & \text{if } b_i = 0, c_i = 1, \\ y, & \text{if } b_i = 1, c_i = 0, \\ z, & \text{if } b_i = 1, c_i = 1. \end{cases}$$

Let $V := \{r_1, \ldots, r_k\}$ (note that V has at most 4 elements). Also, consider the Boolean function $h : \{0,1\}^{|V|} \to \{0,1\}$ defined by $h = f(r_1, \ldots, r_k)$.
We next study in more detail the function h. Observe that

- V must contain at least one of x, y since $\mathbf{b} \neq \mathbf{c}$.
- V must contain at least one of w, x since $\mathbf{b} \neq \mathbf{1}$.
- V must contain at least one of w, y since $\mathbf{c} \neq \mathbf{1}$.

- V must contain at least one or y, z since $\mathbf{b} \neq \mathbf{0}$.
- V must contain at least one of x, z since $\mathbf{c} \neq \mathbf{0}$.

Thus, the cases to consider are $V = \{w, x, y, z\}$, $|V| = 3$, or $V = \{x, y\}$. However, $V = \{x, y\}$ is not possible since then $\mathbf{b} \oplus \mathbf{c} = \mathbf{1}$ and $f(\mathbf{1}) = 1$ (contradicting that $f(\mathbf{b} \oplus \mathbf{c}) = 0$). We now consider the function h (and the corresponding relation R_h) in each of the possible cases.

- Case 1. $V = \{x, y, z\}$.

  Note that $(x, y, z) = (0, 0, 0) \in R_h$ since $f(\mathbf{0}) = 1$. Also, $(0, 1, 1) \in R_h$ since $f(\mathbf{b}) = 1$. Also, $(1, 0, 1) \in R_h$ since $f(\mathbf{c}) = 1$. By self-duality $(1, 1, 1)$, $(1, 0, 0)$, $(0, 1, 0)$ are also in R_h. Then $(x, y, z) = (1, 1, 0)$ is not in R_h since $f(\mathbf{b} \oplus \mathbf{c}) = 0$ and by self-duality neither is $(0, 0, 1)$. So $h(x, y, z)$ is completely determined. Then, for the function $g(x, y) := \sum_z h(x, y, z)$, we have that

  $$g(0, 0) = g(1, 1) = 1 \quad \text{and} \quad g(0, 1) = g(1, 0) = 2,$$

  which is a hard function.

- Case 2. $V = \{w, x, y\}$. This case is similar to Case 1 by switching the spins 0 and 1.

- Case 3. $V = \{w, x, z\}$. $(w, x, z) = (0, 0, 0)$ is in R_h since $f(\mathbf{0}) = 1$. $(0, 0, 1)$ is in R_h since $f(\mathbf{b}) = 1$. $(0, 1, 1)$ is in R_h since $f(\mathbf{c}) = 1$. By self-duality, $(1, 1, 1)$, $(1, 1, 0)$ and $(1, 0, 0)$ are also in R_h. $(0, 1, 0)$ is not in R_h since $f(\mathbf{b} \oplus \mathbf{c}) = 0$. By self-duality, $(1, 0, 1)$ is not in R_h. Then, for the function $g(w, z) = \sum_x h(w, x, z)$, we have that

  $$g(0, 0) = g(1, 1) = 1 \quad \text{and} \quad g(0, 1) = g(1, 0) = 2,$$

  which is a hard function.

- Case 4. $V = \{w, y, z\}$. This case follows from Case 3 by switching $\mathbf{b}$ and $\mathbf{c}$.

- Case 5. $V = \{w, x, y, z\}$

  Similarly to the other cases, we have the following tuples in R_h: $(w, x, y, z) = (0, 0, 0, 0)$, $(0, 0, 1, 1)$, $(0, 1, 0, 1)$, and their complements and we know that $(w, x, y, z) = (0, 1, 1, 0)$ and its complement are not in R_h. Let $h_0 = h(0, x, y, z)$. Let

  $$C = \{(0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 1)\}.$$

For every possible subset S of C, we have to consider the possibility that $R_{h_0} = S \cup \{(0,0,0),(0,1,1),(1,0,1)\}$. This is a lot of cases, but fortunately, some of them can be combined.

- Case 5a. $(x,y,z) = (0,1,0)$ is in S but $(1,1,1)$ is not. Then, for the function $g(x,y) := \sum_w h(w,x,y,x)$, we have

  $$g(0,0) = g(1,1) = 1 \quad \text{and} \quad g(0,1) = g(1,0) = 2,$$

  which is a hard function.

- Case 5b. $(x,y,z) = (1,1,1)$ is in S but $(0,1,0)$ is not. Then, for the function $g(w,x) := \sum_y h(w,x,y,x)$, we have

  $$g(0,0) = g(1,1) = 1 \quad \text{and} \quad g(0,1) = g(1,0) = 2,$$

  which is a hard function.

- Case 5c. $(x,y,z) = (1,0,0)$ is in S but $(1,1,1)$ is not. This case is symmetric to Case 5a.

- Case 5d. $(x,y,z) = (1,1,1)$ is in S but $(1,0,0)$ is not. This case is symmetric to Case 5b.

- Case 5e. $(x,y,z) = (0,1,0)$ and $(1,0,0)$ are both in S.
  Then, for the function $g(x,y) := \sum_w h(w,x,y,w)$, we have

  $$g(0,0) = g(1,1) = 1 \quad \text{and} \quad g(0,1) = g(1,0) = 2,$$

  which is a hard function.

- Case 5f. $S = \emptyset$. Then, for the function $g(x,y) := \sum_{w,z} h(w,x,y,z)$, we have

  $$g(0,0) = g(1,1) = 1 \quad \text{and} \quad g(0,1) = g(1,0) = 2,$$

  which is a hard function.

- Case 5g. $S = \{(0,0,1)\}$. Then, for the function $g(w,z) := \sum_{x,y} R_h(w,x,y,z)$, we have

  $$g(0,0) = g(1,1) = 1 \quad \text{and} \quad g(0,1) = g(1,0) = 3,$$

  which is a hard function.

It remains to argue that each of the functions g used in Cases 1—5 can be simulated using the function f. This is direct. $\{f\}$ implements the function h and $\{h\}$ implements the function g, so $\{f\}$ implements g. Since f supports perfect equality, we can apply Lemma 6.5.2 taking $\Gamma = \{f\}$. Since (trivially) f perfectly simulates Γ, we find that f perfectly simulates g. This completes the proof for the case where $f(\mathbf{0}) = 1$.
We next argue for the case where $f(\mathbf{0}) = 0$. Since f is not affine, we have that $f \neq f^{(k)}_{\mathsf{zero}}$, so there exists $\mathbf{t} \neq \mathbf{0}$ such that $f(\mathbf{t}) = 1$. Let $S := \{i \in [k] \mid t_i = 1\}$ and note that $S \neq \emptyset$. Consider the function f' defined by $f'(\mathbf{x}) := f(\mathbf{x} \oplus \mathbf{t})$ for all $\mathbf{x} \in \{0,1\}^k$. Note that

- $f'(\mathbf{0}) = 1$, since $f(\mathbf{t}) = 1$.
- f' is self-dual. Indeed, for $\mathbf{x} \in \{0,1\}^k$ we have

  $$f'(\overline{\mathbf{x}}) = f(\mathbf{x} \oplus \mathbf{1} \oplus \mathbf{t}) = f(\mathbf{x} \oplus \mathbf{t}) = f'(\mathbf{x}),$$

  where the middle equality follows from the self-duality of f.
- f' is not affine. Since f is not affine, we know from Lemma 6.3.1(1) that there are $\mathbf{a}, \mathbf{b}, \mathbf{c}$ such that $f(\mathbf{a}) = f(\mathbf{b}) = f(\mathbf{c}) = 1$ and $f(\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c}) = 0$. Let $\mathbf{a}' = \mathbf{a} \oplus \mathbf{t}$, $\mathbf{b}' = \mathbf{b} \oplus \mathbf{t}$ and $\mathbf{c}' = \mathbf{c} \oplus \mathbf{t}$. Then by the definition of f', $f'(\mathbf{a}') = f'(\mathbf{b}') = f'(\mathbf{c}') = 1$. But $f'(\mathbf{a}' \oplus \mathbf{b}' \oplus \mathbf{c}') = f'(\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c} \oplus \mathbf{t}) = f(\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c}) = 0$, so f' is not affine.

By the previous argument, we thus have $\{f'\}$ implements a hard function g. We will show that f simulates g. Indeed, observe that the constraint language $\{f, \mathsf{XOR}\}$ implements f' (just apply XOR to the bits of f which correspond to non-zero entries of the vector $\mathbf{t}$). Since $f(\mathbf{0}) = 0$ and f supports perfect equality, by Lemma 6.5.5 we have that f perfectly simulates $\{f, \mathsf{XOR}\}$. Applying Lemma 6.5.2 with $\Gamma = \{f, \mathsf{XOR}\}$ and the g of Lemma 6.5.2 as f', we find that f perfectly simulates f'. Then applying Lemma 6.5.2 again with $\Gamma = \{f'\}$, and the g of Lemma 6.5.2 as g, we obtain that f simulates the hard function g, as wanted. This concludes the proof of Theorem 6.5.6. □

6.5.3 #BIS-easiness

The goal of Section 6.5 is to prove Theorem 6.2.10. The required #BIS-easiness results follows directly from [26].

Lemma 6.5.7 ([26, Lemma 9]). *Let Γ be a constraint language such that every relation in Γ belongs to IM_2. Then, $\#\mathsf{CSP}(\Gamma)$ is #BIS-easy.* □

6.5.4 #BIS-hardness

We next prove the required #BIS-hardness results (cf. Lemma 6.5.11 below). We will use the following results from the literature.

Lemma 6.5.8 ([14, Corollary 3]). *Let $\Delta \geq 6$. It is #BIS-hard to count the number of independent sets in bipartite graphs of maximum degree Δ.* □

The following lemma is from Lemma 13 of [26]. We take the lemma from there since we use the notation of [26]. However, the proof is originally from Lemmas 5.24 and 5.25 of [18].

Lemma 6.5.9 ([18]). *If f is a Boolean function that is not self-dual, then $\{f\}$ implements either δ_0 or δ_1.*

Proof. We just need to explain the terminology in [26, Lemma 13]. It will be then apparent that Items (i)–(iv) in [26, Lemma 13] show that $\{f\}$ implements δ_0 or δ_1. "0-valid" in [26] means that $\mathbf{0} \in R_f$, "1-valid" means that $\mathbf{1} \in R_f$ and "complement-closed" means self-dual. □

Lemma 6.5.10 ([26, Lemma 15], see also [18]). *If f is a Boolean function that is not affine, then $\{f, \delta_0\}$ implements one of* OR, Implies, NAND. *The same is true for $\{f, \delta_1\}$.* □

We are now ready to show that, for every f which supports perfect equality and is not affine, it holds that, for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is #BIS-hard.

Lemma 6.5.11. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function which supports perfect equality. Suppose that f is not affine. Then, for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_{\Delta}(f)$ is #BIS-hard.*

Proof. Assume first that f is self-dual. Then, by Theorem 6.5.6 (note that f is not affine and supports perfect equality by assumption), f perfectly simulates a hard function. By Lemma 6.2.8, we obtain that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. Now, recall that every problem in #P admits an FPRAS using an NP-oracle [76]. Since $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard, we can use it as an oracle to obtain an FPRAS for #BIS.
Assume next that f is not self-dual. By Lemma 6.5.9 we have that $\{f\}$ implements either δ_0 or δ_1. We only need to consider the case where $\{f\}$ implements δ_0, the case of δ_1

follows by just switching the spins 0 and 1. First, by Lemma 6.5.2 with $\Gamma = \{f\}$ and $g = \delta_0$, f perfectly simulates δ_0, so f perfectly simulates $\{f, \delta_0\}$. Recall that f is not affine. By Lemma 6.5.10, it thus follows that $\{f, \delta_0\}$ implements one of OR, NAND, Implies. Using Lemma 6.5.2 again, it follows that f perfectly simulates one of OR, NAND, Implies. Note that OR and NAND correspond to hard functions, so when f perfectly simulates either OR or NAND, we obtain from Lemma 6.2.8 that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. Thus, as in the case of self-dual functions, we may conclude that for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is $\#$BIS-hard.

Thus, it remains to consider the case where f perfectly simulates Implies. By Definition 6.2.7, this means that there exists a k-tuple hypergraph $H' = (V', \mathcal{F}')$ and vertices x, y in H' such that

$$\frac{Z_{0,0}}{Z_{f;H'}} = \frac{1}{3}, \quad \frac{Z_{0,1}}{Z_{f;H'}} = \frac{1}{3}, \quad \frac{Z_{1,1}}{Z_{f;H'}} = \frac{1}{3}, \quad \frac{Z_{1,0}}{Z_{f;H'}} = 0, \tag{6.27}$$

where, for $s_1, s_2 \in \{0, 1\}$, we denote

$$Z_{s_1,s_2} := \sum_{\substack{\sigma:V'\to\{0,1\};\\ \sigma_x=s_1,\,\sigma_y=s_2}} w_{f;H'}(\sigma).$$

Let Δ' be the degree of H'. We will show that for all $\Delta \geq 6\Delta'$, $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is $\#$BIS-hard.

We will use Lemma 6.5.8. In particular, let $G = (V_1 \cup V_2, E)$ be a bipartite graph of maximum degree 6 where V_1, V_2 denote the parts of G in its partition. Let $H = (V, \mathcal{F})$ be the k-tuple hypergraph obtained from G as follows. Start by putting all of the vertices in $V_1 \cup V_2$ into V. Then add additional vertices and hyperarcs as follows. For every edge $(v_1, v_2) \in E$ such that $v_1 \in V_1$ and $v_2 \in V_2$, take a distinct copy of H' and identify vertex x in H' with v_1 and vertex y in H' with v_2. Note that $V_1 \cup V_2 \subseteq V$ and that the degree of H is $6\Delta'$.

Let $\mathcal{I}_G$ denote the set of independent sets of G. Then, we claim that

$$Z_{f;H} = |\mathcal{I}_G| \cdot (Z_{f;H'}/3)^{|E|}. \tag{6.28}$$

Before proving (6.28), note that an oracle call to $\#\mathsf{Hyper2Spin}_{\Delta}(f)$ for $\Delta \geq 6\Delta'$ with input H and relative error $\varepsilon > 0$ yields via (6.28) an estimate for the number of independent sets in bipartite graphs of maximum degree 6 which is within relative error ε

from the true value. Thus, using Lemma 6.5.8, we obtain an AP-reduction from #BIS to $\#\mathsf{Hyper2Spin}_{\Delta}(f)$ for all $\Delta \geq 6\Delta'$, as wanted.

To show (6.28), let $\sigma : V \to \{0,1\}$ be an assignment such that $w_{f;H}(\sigma) > 0$. The copies of H' ensure that for every edge $(v_1, v_2) \in E$ such that $v_1 \in V_1$ and $v_2 \in V_2$ it holds that either $\sigma(v_1) \neq 1$ or $\sigma(v_2) \neq 0$. Thus, the set $(\sigma^{-1}(1) \cap V_1) \cup (\sigma^{-1}(0) \cap V_2)$ is an independent set of G. Conversely, for every independent set I of G, consider

$$\Omega_I = \{\sigma : V \to \{0,1\} \mid \sigma_{I \cap V_1} = \mathbf{1},\ \sigma_{I \cap V_2} = \mathbf{0},\ \sigma_{V_1 \setminus I} = \mathbf{0},\ \sigma_{V_2 \setminus I} = \mathbf{1}\}.$$

Then, using (6.27), we have that the number of assignments $\sigma \in \Omega_I$ such that $w_{f;H}(\sigma) > 0$ is equal to

$$\prod_{(v_1,v_2) \in E} Z_{\sigma(v_1)\sigma(v_2)} = (Z_{f;H'}/3)^{|E|}.$$

Summing this over all $I \in \mathcal{I}_G$, we obtain (6.28), thus completing the proof of Lemma 6.5.11. □

6.5.5 NP-hardness

In this section, we show (Lemma 6.5.14 below) that if f supports perfect equality, and it is not affine, and is not in IM_2 then, for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. To do this, we need some preparation. The ideas behind the following lemma are essentially from [26].

Lemma 6.5.12. *If f perfectly simulates* Implies*, then f simulates $\{\delta_0, \delta_1\}$ (not necessarily perfectly).*

Proof. We first show that f supports both pinning-to-0 and pinning-to-1. Since f perfectly simulates Implies, there exists a k-tuple hypergraph H and vertices v_1, v_2 in H such that

$$\mu_{00} = 1/3, \quad \mu_{01} = 1/3, \quad \mu_{11} = 1/3, \quad \mu_{10} = 0, \tag{6.29}$$

where, for $s_1, s_2 \in \{0,1\}$, we denote $\mu_{s_1 s_2} := \mu_{f;H}(\sigma_{v_1} = s_1, \sigma_{v_2} = s_2)$.

Note that

$$\mu_{f;H}(\sigma_{v_1} = 0) = \mu_{00} + \mu_{01}, \quad \mu_{f;H}(\sigma_{v_1} = 1) = \mu_{10} + \mu_{11},$$
$$\mu_{f;H}(\sigma_{v_2} = 0) = \mu_{00} + \mu_{10}, \quad \mu_{f;H}(\sigma_{v_2} = 1) = \mu_{01} + \mu_{11}.$$

It follows that

$$\begin{aligned} \mu_{f;H}(\sigma_{v_1} = 0) = 2/3, \quad \mu_{f;H}(\sigma_{v_1} = 1) = 1/3, \\ \mu_{f;H}(\sigma_{v_2} = 0) = 1/3, \quad \mu_{f;H}(\sigma_{v_2} = 1) = 2/3. \end{aligned} \tag{6.30}$$

Then, using (6.30), we obtain that H and its vertices v_1, v_2 satisfy the assumptions of Lemma 6.3.4 (v_1 satisfies Item 1 and v_2 Item 2) and hence we obtain that f supports pinning-to-0 and pinning-to-1.
To conclude that f simulates $\{\delta_0, \delta_1\}$, consider the k-tuple hypergraph H as above and consider the conditional distribution $\mu_{f;H}^{\mathcal{V}_0}$ where we pin the vertex v_2 to 0 (this is allowed since f supports pinning-to-0). Then,

$$\mu_{f;H}^{\mathcal{V}_0}(\sigma(v_1) = 0) = 1, \quad \mu_{f;H}^{\mathcal{V}_0}(\sigma(v_1) = 1) = 0$$

so f simulates δ_0. Analogously, by pinning the vertex v_1 to 1, we also obtain that f simulates δ_1, concluding the proof. □

Lemma 6.5.13 ([26, Proof of Lemma 19])**.** *If f is a Boolean function that is not in IM_2, then $\{f, \mathsf{Implies}, \delta_0, \delta_1\}$ implements either* OR *or* NAND. □

Lemma 6.5.14. *Let $f : \{0,1\}^k \to \{0,1\}$ be a function which supports perfect equality. Suppose that f is not affine and is not in IM_2. Then, for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is* NP*-hard.*

Proof. The proof is similar in structure to the proof of Lemma 6.5.11.
Assume first that f is self-dual. Then, by Theorem 6.5.6 (note that f is not affine and supports perfect equality by assumption), f perfectly simulates a hard function. By Lemma 6.2.8, we obtain that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.
Assume next that f is not self-dual. By Lemma 6.5.9 we have that $\{f\}$ implements either δ_0 or δ_1. We only need to consider the case where $\{f\}$ implements δ_0, the case of δ_1 follows by switching the spins 0 and 1. First, by Lemma 6.5.2 with $\Gamma = \{f\}$ and $g = \delta_0$, f perfectly simulates δ_0, so f perfectly simulates $\{f, \delta_0\}$. Recall that f is not affine. By Lemma 6.5.10, it thus follows that $\{f, \delta_0\}$ implements one of OR, NAND, Implies. Using Lemma 6.5.2 again, it follows that f perfectly simulates one of OR, NAND, Implies. OR and NAND correspond to hard functions, so when f perfectly simulates either OR or NAND, we obtain from Lemma 6.2.8 that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. Thus, it remains to consider the case where f perfectly simulates Implies.
Since f perfectly simulates Implies, by Lemma 6.5.12, we obtain that f simulates $\{\delta_0, \delta_1\}$. Thus, f simulates $\{f, \mathsf{Implies}, \delta_0, \delta_1\}$. By Lemma 6.5.13, using that f is not in IM_2, we

have that $\{f, \mathsf{Implies}, \delta_0, \delta_1\}$ implements either OR or NAND. By Lemma 6.5.4, we thus obtain that f simulates either OR or NAND. Hence, as above, we can use Lemma 6.2.8 to conclude that for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.
This concludes the proof. □

6.5.6 Proof of Theorem 6.2.10

We are ready to prove Theorem 6.2.10, which we restate here for convenience.

Theorem 6.2.10. *Let $f : \{0,1\}^k \to \{0,1\}$ be a Boolean function that is not affine. Suppose that f supports perfect equality.*

1. *If f is in IM_2, then for all sufficiently large Δ, $\#\mathsf{Hyper2Spin}_{\Delta}(f)$ is $\#\mathsf{BIS}$-equivalent.*
2. *If f is not in IM_2, then for all sufficiently large Δ, there exists a real number $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard.*

Proof. Item 1 is a consequence of Lemma 6.5.7 and Lemma 6.5.11. Item 2 is a consequence of Lemma 6.5.14. □

6.6 Proof of Theorem 6.1.7

In this section, we combine the pieces to prove Theorem 6.1.7.
We will need the following lemma.

Lemma 6.6.1. *Let $f_1 : \{0,1\}^{k_1} \to \{0,1\}$ and $f_2 : \{0,1\}^{k_2} \to \{0,1\}$ be Boolean functions such that f_1 is not affine and f_2 is not in IM_2. Then, the function f defined by $f(\mathbf{x}, \mathbf{y}) = f_1(\mathbf{x})f_2(\mathbf{y})$ is neither affine nor does it belong to IM_2.*

Proof. We first prove that f is not affine. Since f_1 is not affine, by Item 1 of Lemma 6.3.1, there exist $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)} \in R_{f_1}$ such that $\mathbf{x}^{(1)} \oplus \mathbf{x}^{(2)} \oplus \mathbf{x}^{(3)} \notin R_{f_1}$. Let $\mathbf{y}$ be such that $f_2(\mathbf{y}) = 1$ (such a $\mathbf{y}$ exists, otherwise f_2 would belong to IM_2).
For each $i = 1, 2, 3$, consider the vector $\mathbf{z}^{(i)}$ of length $k_1 + k_2$ obtained by concatenating the vectors $\mathbf{x}^{(i)}$ and $\mathbf{y}$. Since $\mathbf{x}^{(i)} \in R_{f_1}$ and $\mathbf{y} \in R_{f_2}$, we have that $f(\mathbf{z}^{(i)}) = f_1(\mathbf{x}^{(i)})f_2(\mathbf{y}) = 1$, so $\mathbf{z}^{(i)} \in R_f$ for $i = 1, 2, 3$. Observe that $f(\mathbf{z}^{(1)} \oplus \mathbf{z}^{(2)} \oplus \mathbf{z}^{(3)}) = f_1(\mathbf{x}^{(1)} \oplus \mathbf{x}^{(2)} \oplus \mathbf{x}^{(3)})f_2(\mathbf{y}) = 0$, so $\mathbf{z}^{(1)} \oplus \mathbf{z}^{(2)} \oplus \mathbf{z}^{(3)} \notin R_f$. Thus,

$$\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \mathbf{z}^{(3)} \in R_f \text{ but } \mathbf{z}^{(1)} \oplus \mathbf{z}^{(2)} \oplus \mathbf{z}^{(3)} \notin R_f,$$

so by Item 1 of Lemma 6.3.1, we have that f is not affine.
We next show that f does not belong to IM_2. Since $f_2 \notin \mathsf{IM}_2$, by Lemma 6.3.3, there exist $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in R_{f_2}$ such that either $\mathbf{y}^{(1)} \vee \mathbf{y}^{(2)} \notin R_{f_2}$ or $\mathbf{y}^{(1)} \wedge \mathbf{y}^{(2)} \notin R_{f_1}$. Assume that $\mathbf{y}^{(1)} \vee \mathbf{y}^{(2)} \notin R_{f_2}$, the other case is completely analogous and actually follows by duality (switching the spins 0 and 1). Let $\mathbf{x}$ be such that $f_1(\mathbf{x}) = 1$ (such an $\mathbf{x}$ exists, otherwise f_1 would be affine).
For each $i = 1, 2$, consider the vector $\mathbf{w}^{(i)}$ of length $k_1 + k_2$ obtained by concatenating the vectors $\mathbf{x}$ and $\mathbf{y}^{(i)}$. Since $\mathbf{x} \in R_{f_1}$ and $\mathbf{y}^{(i)} \in R_{f_2}$, we have that $f(\mathbf{w}^{(i)}) = f_1(\mathbf{x}) f_2(\mathbf{y}^{(i)}) = 1$, so $\mathbf{z}^{(i)} \in R_f$ for $i = 1, 2$. Observe that $f(\mathbf{w}^{(1)} \vee \mathbf{w}^{(2)}) = f_1(\mathbf{x}) f_2(\mathbf{y}^{(1)} \vee \mathbf{y}^{(2)}) = 0$, so $\mathbf{w}^{(1)} \vee \mathbf{w}^{(2)} \notin R_f$. Thus,

$$\mathbf{w}^{(1)}, \mathbf{w}^{(2)} \in R_f \text{ but } \mathbf{w}^{(1)} \vee \mathbf{w}^{(2)} \notin R_f,$$

so by Lemma 6.3.3, we have that f does not belong to IM_2.
This concludes the proof. □

Theorem 6.1.7. *Let Γ be a Boolean constraint language. Then,*

1. *If every function in Γ is affine then $\#\mathsf{CSP}(\Gamma)$ and $\#\mathsf{NoRepeatCSP}(\Gamma)$ are both in FP.*
2. *Otherwise, if $\Gamma \subseteq \mathsf{IM}_2$, then there exists an integer Δ_0 such that for all $\Delta \geq \Delta_0$, $\#\mathsf{CSP}_\Delta(\Gamma)$ and $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ are both $\#\mathsf{BIS}$-equivalent under AP-reductions, and*
3. *Otherwise, there exists an integer Δ_0 such that for all $\Delta \geq \Delta_0$, there exists a real number $c > 1$ such that $\#\mathsf{CSP}_{\Delta,c}(\Gamma)$ and $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ are both NP-hard.*

Proof. We consider each of the three cases.

1. If every function in Γ is affine (cf. Definition 6.1.3), then Z_I can be computed exactly in polynomial time using Gaussian elimination. This was already noted in the exact-counting dichotomy of Creignou and Hermann [17].
2. Suppose that $\Gamma \subseteq \mathsf{IM}_2$ and that Γ includes a function f which is not affine. By the unbounded-degree $\#\mathsf{BIS}$-easiness result of [26], which is stated here as Lemma 6.5.7, it follows that for all positive integers Δ, $\#\mathsf{CSP}_\Delta(\Gamma)$ is $\#\mathsf{BIS}$-easy. Clearly, every instance of $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ is an instance of $\#\mathsf{CSP}_\Delta(\Gamma)$, from which we obtain that $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ is $\#\mathsf{BIS}$-easy as well.

If f supports perfect equality, then by Theorem 6.2.10, for all sufficiently large Δ, the problem $\#\mathsf{Hyper2Spin}_\Delta(f)$ is $\#\mathsf{BIS}$-hard. As we noted in Section 6.1.1, the problem $\#\mathsf{Hyper2Spin}_\Delta(f)$ is equivalent to $\#\mathsf{NoRepeatCSP}_\Delta(\{f\})$ from which we obtain that $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ is $\#\mathsf{BIS}$-hard as well. Note that $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ is a restricted version of $\#\mathsf{CSP}_\Delta(\Gamma)$ (the restriction being that constraints may not repeat variables), so it follows immediately that $\#\mathsf{CSP}_\Delta(\Gamma)$ is $\#\mathsf{BIS}$-hard.

There is a final case that does not arise if $\#\mathsf{BIS}$ is not NP-hard to approximate, but we include it to make the proof complete. In particular, if f does not support perfect equality, then by Theorem 6.2.9, it simulates a hard function. So, by Lemma 6.2.8, for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. As observed in the proof of Lemma 6.5.11, this implies that $\#\mathsf{Hyper2Spin}_\Delta(f)$ is $\#\mathsf{BIS}$-hard. Then, as in the previous case, we obtain that $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ and $\#\mathsf{CSP}_\Delta(\Gamma)$ are $\#\mathsf{BIS}$-hard.

3. Suppose that there are functions $f_1, f_2 \in \Gamma$ such that f_1 is not affine and f_2 is not in IM_2 (it might be the case that $f_1 = f_2$). Then, consider the function $f(\mathbf{x}, \mathbf{y})$ defined by $f(\mathbf{x}, \mathbf{y}) = f_1(\mathbf{x})f_2(\mathbf{y})$. By Lemma 6.6.1, we have that f is neither affine nor does it belong to IM_2.

   Thus, if f supports perfect equality, then by Theorem 6.2.10, for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard, which is equivalent to saying that $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\{f\})$ is NP-hard. Now note that there is an easy reduction from $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\{f\})$ to $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ — given an instance I of $\#\mathsf{CSP}_{\Delta,c}(\{f\})$, every constraint involving f is re-written as two constraints involving f_1 and f_2. Thus, $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ is also NP-hard. Since $\#\mathsf{NoRepeatCSP}_\Delta(\Gamma)$ is a restricted version of $\#\mathsf{CSP}_\Delta(\Gamma)$, we have that $\#\mathsf{CSP}_{\Delta,c}(\Gamma)$ is NP-hard as well.

   Otherwise, by Theorem 6.2.9, f simulates a hard function. So, by Lemma 6.2.8, for all sufficiently large Δ, there exists $c > 1$ such that $\#\mathsf{Hyper2Spin}_{\Delta,c}(f)$ is NP-hard. As in the previous paragraph, this implies that $\#\mathsf{NoRepeatCSP}_{\Delta,c}(\Gamma)$ and $\#\mathsf{CSP}_{\Delta,c}(\Gamma)$ are NP-hard. □

Chapter 7

Conclusion and Open Problems

Many works has been developed in the area of approximate counting, both in the algorithmic part and complexity part. One of the most exciting and challenging research directions is to establish the connection between the uniqueness phase transition of spin systems and the approximability phase transition of computing partition functions. This research has been done for some important models such as the hardcore model [78], the antiferromagentic Ising model [54] and the colouring model when $q = 4$ (as we have described in Chapter 4). For each of them, an FPTAS has already been designed in the uniqueness region, and these ranges of approximation were proved to be optimal [34, 72]. To date, no MCMC based FPRASes that achieve the conjectured optimal bound $q = \Delta + 1$ have been established for the colouring model. For some other cases, the tree uniqueness threshold is not the correct answer to the approximability threshold—we can even do better. For the ferromagnetic Ising model, for example, Jerrum and Sinclair established an FPRAS [48] which works for all graphs and all values of parameters on vertices and edges in the ferromagnetic region, even beyond the tree uniqueness threshold. However, perhaps surprisingly, the method of correlation decay does not work even in the uniqueness region. These facts support the conjecture that we can design polynomial-time approximation algorithms (FPTAS or FPRAS) for computing partition functions in the tree uniqueness region of spin systems, although the region that an FPTAS exists and the region that an FPRAS exists may be quite different.

To understand in which region a problem can be solved approximately in polynomial time, it is probably a good start to analyse the condition of uniqueness for the Gibbs distribution. Sometimes to prove uniqueness, we also need the decay of correlations in spin systems, which is related to the algorithmic techniques. The uniqueness threshold

has been well studied for some interesting models such as the antiferromagnetic Ising model, the colouring model and the ferromagnetic Potts model. But it is still open for the antiferromagnetic model. An up-to-date result was introduced in Chapter 3, which proved the uniqueness threshold for $q = 3, d \geq 2$ and $q = 4, d \in \{3, 4, 5, 6, 7, 8, 9\}$. However, our results is based on the discussion of cases under the constraints that for all $c' \in [q]$, either $p_{c'}^{(k)} = p_c^{(k)}$, or $p_{c'}^{(k)} = \alpha \cdot p_c^{(k)}$ and there exists $c' \in [q]$ such that $p_{c'}^{(k)} = \alpha \cdot p_c^{(k)}$. It is easy to see that for general q, the number of satisfying tuples will increase exponentially, which makes the task of analysing the function $h_{c_1,c_2,\beta}$ impossible. Therefore, proving uniqueness for the antiferromagnetic Potts model is still an interesting open problem. We conjecture that new tools and techniques are necessary for this problem.

Another part of open problems is the classification of complexity of approximate counting problems. Extensive research has been done on the classification of partition functions according to the complexity of computing their values exactly. However, much less is known in the approximate counting area. As we described in Chapter 6, we focused on whether there is a "barrier" value Δ_0 such that, for all $\Delta \geq \Delta_0$, approximation is intractable, because in general even some decision problems are still open. But our result cannot provide any assistance to the classification of Γ for some specific Δ. The best known result is on the conservative cases, which cannot apply to many natural problems such as the hypergraph hardcore model. In fact, our result also relied on the AP-reductions to the problem of counting independent sets in a bipartite graph (#BIS). However, #BIS itself is almost the most important open problem underlying the study of complexity of approximate counting. It is still unknown whether #BIS admits an FPTAS/FPRAS, or #BIS is NP-hard. It is conjectured to be of an intermediate complexity.

We conclude this chapter with introducing some new techniques to design FPTASes and the possible future works. Although correlation decay based FPTASes have played significant roles in the antiferromagnetic spin systems, it cannot work well in many other models. Recently, some other approaches have been developed. One is developed by Ankur Moitra, based on the Lovász Local Lemma and linear programming. It gives an algorithm to approximately compute the marginal probability of a variable in a SAT formula under Lovász Local Lemma condition [61, 41]. Another algorithmic technique is related to zeros of polynomials. It has been developed by Barvinok [3] and extended by Patel and Regts [66]. The idea is to view the partition function as a polynomial of given parameters and attempt to establish connections between the non-vanishing set and the

approximability region. As a possible future work, it will be promising to develop more powerful tools to design approximation algorithms and connect the approximability with other theory (such as the existing results on the uniqueness region, Lovász Local lemma and zero of polynomials) to identify the exact threshold of computational complexity and understand the essential hardness of approximation.

Chapter 8

Appendix: Mathematica Code

8.1 Verifying inequalities in Chapter 3

8.1.1 Lemma 3.3.5

Both of the queries in the following code give the output `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
bb = 1-b;
W = 1 - 3y + bb ((3y - 1) + (1 - x - 2y) + x(2x + y - 1) + y(x - y));
lhs = D[fu, b];
rhs = -fu^2 (d (1 - bb x)^(d/2) (1- bb(1-x-y))^(d/2) W /
((1- bb y)^(d+1) (1- bb x) (1- bb (1-x-y))));
FullSimplify[lhs == rhs]

fl = (1 - (1 - b) x)^d /
((1 - (1 - b) x)^d  + (1 - (1 - b) y)^d + (1 - (1 - b) (1 - x - y))^d);
lhs = D[fl, b];
rhs = fl^2 d ((2x+y-1) (1-bb(1-x-y))^(d-1) + (x-y) (1-bb y)^(d - 1)) /
((1-bb x)^(d+1));
FullSimplify[lhs == rhs]
```

8.1.2 Lemma 3.3.13

Both of the queries in the following code give the output `False`.

```
Resolve[Exists[{x, y, b}, 0 < y <= 1/3 && 1106/2500 <= x < 1 &&
0 < b <= 1 && (fu /. {d -> 2}) == x && (fl /. {d -> 2}) == y]]
Resolve[Exists[{x, y, b}, 0 < y <= 460/2000 && 1/3 <= x < 1 &&
0 < b <= 1 && (fu /. {d -> 2}) == x && (fl /. {d -> 2}) == y]]
```

8.1.3 Lemma 3.3.17

The following code gives the output `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
gu = fu - x /. {b -> 1 - 3/(d + 1), x -> m y};
A =  (1 - 3 y/(d + 1))^d;
B = (1- 3(1-y(m+1))/(d+1))^(d/2) (1- 3m y/(d+1))^(d/2);
W = A B/(A+2B)^2;
lhs = D[gu, y];
rhs = -m+(9 d W /(3y(m+1)+d-2)) (m(2m y+y-1)/(1+d-3m y) - (2m y+y+d-1)/(1+d-3y));
FullSimplify[lhs == rhs]
```

8.1.4 Lemma 3.3.18

The following code gives the output `True`.

```
fl = (1 - (1 - b) x)^d /
((1 - (1 - b) x)^d  + (1 - (1 - b) y)^d + (1 - (1 - b) (1 - x - y))^d);
gl = fl - y /. {b -> 1 - 3/(d + 1), x -> m y};
W = (m(2d-1)+d+1) ((3y(m+1)+d-2) / (d+1))^d /
((1 + d - 3m y) (d-2+ 3y(1 + m))) +
(m-1)(d+1)(1-(3 y)/(d+1))^d / ((1 + d - 3 y) (1 + d - 3m y));
lhs = D[gl, y];
rhs = -3 d (1 - (3m y)/(d+1))^d W / ( ((3y(m+1)+d-2)/(d+1))^d
+ (1 - (3m y)/(d+1))^d + (1-(3 y)/(d+1))^d )^2 - 1;
FullSimplify[lhs == rhs]
```

8.1.5 Lemma 3.3.20

The following code outputs `False` and `False`.

```
y2 = 7/(10 m + 12);
y1 = y2 + 3/500;
x2 = m y2;
x1 = m y1;
Resolve[Exists[m,  157/80 <= m < 32 && (0 >= y1 || y1 >= 1 - x1 - y1 ||
1 - x1 - y1 >= 1/3 ||  1/3 >= x1 || x1 >= 1 - y1 ) ]]
Resolve[Exists[m, 32 <= m && (0 >= y2 || y2 >= 1 - x2 - y2 ||
1 - x2 - y2 >= 1/3 ||  1/3 >= x2 || x2 >= 1 - y2 ) ]]
```

8.1.6 Lemma 3.3.21

Both queries in the following code for the differentiation in (3.30) output `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
bb = 1 - b;
R = (1 - bb x)^(d/2) (1 - bb (1 - x - y))^(d/2) / (1 - bb y)^d;
xlhs = D[fu, x];
xrhs = fu^2 R d bb^2 (2 x + y - 1) / ((1 - bb(1 - x - y))(1 - bb x));
ylhs = D[fu, y];
yrhs = -fu^2 R d bb (3 + bb(2 x + y - 2)) /((1 - bb(1 - x - y))(1 - bb y));
FullSimplify[xlhs == xrhs]
FullSimplify[ylhs == yrhs]
```

The proof of (3.33). We consider two cases — when $\mu < 32$ and when $\mu \geq 32$. The output is `False` in both cases.

```
lhs=(8-y)(2x+y-1)/( (8-x)(2x+y+22) );
x1 = 7 m/(10 m + 12) + 3 m/500;
y1 = 7/(10 m + 12) + 3/500;
Resolve[Exists[m, 1<m <32 && (lhs/.{x->x1, y->y1}) >= 1/24]]
x2 = 7 m/(10 m + 12);
y2 = 7/(10 m + 12);
Resolve[Exists[m, m>=32 && (lhs/.{x->x2, y->y2}) >= 1/24]]
```

Here is the code to show that Y is increasing in $\hat{\beta}_*$. The output is `False`.

```
Num = 3 + b (2 x + y - 2);
Den = (1 - b (1 - x - y)) (1 - b y);
Der = D[Num/Den  , b];
Resolve[Exists[{d, x, y, b},
d >= 0 && 0 < b <= 1/8 && 0 < y < 1 - x - y < 1/3 < x < 1 - y &&
Der < 0]]
```

Here is the code for Case 1. The output is `False`.

```
lhs = 3 (2 x + y + 22)/((8 - y) (x + y + 7)) /.
{x -> 7 m/(10 m + 12), y -> 7/(10 m + 12)};
rhs = 8/7;
Resolve[Exists[m, lhs >= rhs && m > 32]]
```

Here is the code for Case 2. The output is `False`.

```
lhs = 3 (2 x + y + 22)/((8 - y) (x + y + 7)) /.
{x -> 7 m/(10 m + 12) + 3 m/500, y -> 7/(10 m + 12) + 3/500};
rhs = 24 (25 m^2 + 60 m + 3536)/(25 m^2 + 60 m + 73536);
Resolve[Exists[m, lhs >= rhs && 1 < m]]
```

8.1.7 Lemma 3.3.23

The output of the following code, for the differentiation in (3.36), is `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
psi[d_,z_] := d/(d-3z+1) + Log[d +1-3z];
zeta[d_,x_,y_]:=2psi[d,y]-psi[d,x]-psi[d,1-x-y];
lhs = D[ (fu/.{b -> 1 - 3/(d + 1)}), d];
rhs = (1- 3x/(d+1))^(d/2) (1- 3y/(d+1))^d (1- 3(1-x-y)/(d+1))^(d/2)zeta[d,x,y]/
((1- 3y/(d+1))^d + 2(1- 3x/(d+1))^(d/2) (1- 3 (1-x-y)/(d+1))^(d/2))^2;
FullSimplify[lhs == rhs,d>=0]
```

The following code establishes Facts 1, 2, and 3 for $\mu = 157/80$. The output is `False`, 0, then `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
hu=fu-x /.{b -> 1 - 3/(d+1)};
psi[d_,z_] := d/(d-3z+1) + Log[d +1-3z];
zeta[d_,x_,y_]:=2psi[d,y]-psi[d,x]-psi[d,1-x-y];
Fd = D[zeta[d,x,y], d];
xm = 7 m/(10 m + 12) + 3 m/500 /. {m -> 157/80};
ym = 7/(10 m + 12) + 3/500 /. {m -> 157/80};
Resolve[Exists[d, d >= 23 && (Fd /. {x -> xm, y -> ym}) >= 0]]
Limit[zeta[d,xm,ym], d -> \[Infinity]]
Limit[hu /. {x -> xm, y -> ym}, d -> \[Infinity]]<0
```

The following code establishes Facts 1, 2, and 3 for $\mu = 32$. The output is `False`, 0, then `True`.

```
fu = (1 - (1 - b) y)^d /
((1 - (1 - b) y)^d + 2 (1 - (1 - b) x)^(d/2) (1 - (1 - b) (1 - x - y))^(d/2));
hu=fu-x /.{b -> 1 - 3/(d+1)};
psi[d_,z_] := d/(d-3z+1) + Log[d +1-3z];
zeta[d_,x_,y_]:=2psi[d,y]-psi[d,x]-psi[d,1-x-y];
Fd = D[zeta[d,x,y], d];
xm = 7 m/(10 m + 12) /. {m -> 32};
ym= 7/(10 m + 12) /. {m -> 32};
Resolve[Exists[d, d >= 23 && (Fd /. {x -> xm, y -> ym}) >= 0]]
Limit[zeta[d,xm,ym], d -> \[Infinity]]
Limit[hu /. {x -> xm, y -> ym}, d -> \[Infinity]]<0
```

8.1.8 Lemma 3.3.25

We first prove that $h_\ell(23, \mu) > 0$ for $\mu \geq 157/80$. The output to both queries is `False`.

```
fl = (1 - (1 - b) x)^d /
((1 - (1 - b) x)^d  + (1 - (1 - b) y)^d + (1 - (1 - b) (1 - x - y))^d);
hl = fl - y /. {b -> 1 - 3/(d+1), x -> m y};
x1 = 7 m/(10 m + 12) + 3 m/500;
```

```
y1 = 7/(10 m + 12) + 3/500;
h1 = hl /. {d -> 23, x -> x1, y -> y1};
Resolve[Exists[m, 157/80<=m <32 && h1 <= 0]]
x2 = 7 m/(10 m + 12);
y2 = 7/(10 m + 12);
h2 = hl /. {d -> 23, x -> x2, y -> y2};
Resolve[Exists[m, m >= 32 && h2 <= 0]]
```

The output of the following code, for the differentiation in (3.38), is `True`.

```
fl = (1 - (1 - b) x)^d/
((1 - (1 - b) x)^d + (1 - (1 - b) y)^d + (1 - (1 - b) (1 - x - y))^d);
psi[d_, z_] := d/(d - 3 z + 1) + Log[d + 1 - 3 z];
A = (1 - 3 x/(d + 1))^d;
B = (1 - 3 y/(d + 1))^d;
CC = (1 - 3 (1 - x - y)/(d + 1))^d;
lhs = D[(fl /. {b -> 1 - 3/(d + 1)}), d];
rhs = (A CC (psi[d, x] - psi[d, 1 - x - y]) + A B (psi[d, x] - psi[d, y]))/
(A + B + CC)^2;
FullSimplify[lhs == rhs, d >= 0]
```

8.1.9 Lemma 3.3.28

The code checks that all of the desired inequalities are satisfied. The output is `True`.

```
Flag = True;
u[0] = 1;
l[0] = 0;
For[dd = 3, dd <= 22, dd++, (
ffu = Ceiling[10000 fu /. {q -> 3, d -> dd, b -> 1 - 3/(dd + 1)}]/10000;
ffl = Floor[10000 fl /. {q -> 3, d -> dd, b -> 1 - 3/(dd + 1)}]/10000;
For[n = 0, n <= 60, n++, (
u[n + 1] = ffu /. {x -> u[n], y -> l[n]};
l[n + 1] = ffl /. {x -> u[n], y -> l[n]};
Flag = Flag && u[n] >= u[n + 1] && l[n] <= l[n + 1] &&
```

```
2 u[n] + l[n] >= 1 >= 2 l[n] + u[n])];
Flag = Flag && u[60]/l[60] <= 53/27;)];
Flag
```

8.1.10 Lemma 3.3.36

The code checks all relevant values of d and d_0. The output is `True`. The substitution of u for $\alpha^{1/d}$ is there to make the code run faster. Despite this, it takes more than 5 minutes to run on our machine.

```
b = 1 - 3/(d + 1);
x = 1 - (1 - b) (a - 1)/(b + 2 a);
y = 1 - (1 - b) (a - 1)/(b + a + 1);
phi = 1 + (3/(d + 1)) (1 - x^d0 y^(d - d0))/(b +
x^d0 (y^(d - d0) + 1)) /. {a -> u^d};
Flag = True;
For[dd = 2, dd <= 22, dd++, u0 = (53/27)^dd;
For[dd0 = 0, dd0 <= dd, dd0++,
Flag =  Flag && ! Resolve[Exists[u,
(phi /. {d -> dd, d0 -> dd0}) >= u && 1 < u <= u0]];];];
Flag
```

8.1.11 Lemma 3.3.37

The following code outputs `True`, therefore verifying the differentiation in (3.46).

```
b = 1 - 3/(d + 1);
x = 1 - (1 - b) (a - 1)/(b + 2 a);
y = 1 - (1 - b) (a - 1)/(b + a + 1);
phi = 1+(1/d)( x^(-2d0/(2+b)) y^(-(d-d0)/(2+b))
- x^(d0 b/(2+b)) y^((d-d0)(1+b)/(2+b)) )-a^(1/d);
gx=(2a+b)(a b+a+1);
gy=(a+b+1)(a b+2);
S=x^(2d0/(2+b)) y^((d-d0)/(2+b)) (a+b+1) (a b+2)/(1-b);
rhs=( (1-b)x^(-2d0/(2+b)) y^(-(d-d0)/(2+b)) / (d a gy) ) *
(2 d0 a (gy/gx) + (d-d0)a + x^d0 y^(d-d0) (d0 b a (gy/gx)+ (d-d0)(1+b)a)
```

```
-a^(1/d) S);
Resolve[Simplify[D[phi, a] - rhs] == 0]
```

The following code outputs `True True`, verifying the calculation for Step 3.

```
b = 1 - 3/(d + 1);
x = 1 - (1 - b) (a - 1)/(b + 2 a);
y = 1 - (1 - b) (a - 1)/(b + a + 1);
s0 = -(((2 d - 1) d (1 - k) a)/(d +
1) + ((d - 2) k ((d - 2) a + 2 (d + 1)))  /(3 (d + 1)));
W = 1 + (a - 1)/d - (d - 1) (a - 1)^2/(2 d^2);
t0 = W ((d + 1) a + 2 d - 1) ((d - 2) a + 2 (d + 1))/(3 (d + 1));
xi3 = s0 x^(d k) y^(d (1 - k)) +
t0 x^(2 k (d + 1)/3) y^((1 - k) (d + 1)/3);
FullSimplify[(xi3 /. {a -> 1}) == d + d k]
FullSimplify[(D[xi3, a] /. {a -> 1}) == (1/3) (- d k + 3 d - 4 k) ]
```

The final two code segments are for Step 4. The following code calculates t_2 and verifies that $\frac{\partial^2 t_2}{\partial \kappa^2} \geq 0$, $\left.\frac{\partial t_2}{\partial \kappa}\right|_{\kappa=0} \geq 0$ and $t_2|_{\kappa=0} \geq 0$. The output is `False`, `False`, and `False`.

```
b = 1 - 3/(d + 1);
x = 1 - (1 - b) (a - 1)/(b + 2 a);
y = 1 - (1 - b) (a - 1)/(b + a + 1);
s0 = -(((2 d - 1) d (1 - k) a)/(d +
1) + ((d - 2) k ((d - 2) a + 2 (d + 1)))/(3 (d + 1)));
W = 1 + (a - 1)/d - (d - 1) (a - 1)^2/(2 d^2);
t0 = W ((d + 1) a + 2 d - 1) ((d - 2) a + 2 (d + 1))/(3 (d + 1));

t2 = Simplify[ D[t0 x^(2 k (d + 1)/3) y^((1 - k) (d + 1)/3), {a,
2}]/(x^(2 k (d + 1)/3) y^((1 - k) (d + 1)/3))];

tk1 = D[t2, k];
tk2 = D[tk1, k];
Resolve[Exists[{d, a}, tk2 < 0  && d >= 23 && 1 <= a <= 53/27]]
Resolve[Exists[{d, a}, (tk1 /. {k -> 0}) < 0 && d >= 23 && 1 <= a <= 53/27]]
Resolve[Exists[{d, a}, (t2 /. {k -> 0}) < 0  && d >= 23 && 1 <= a <= 53/27]]
```

The following code calculates s_2 (as well as t_2) and verifies that $s_2 + t_2$ is positive. It takes about 10 minutes to run. The output is `False`. The reason for the transformation of α (as a function of r) is to speed up the calculation.

```
b = 1 - 3/(d + 1);
x = 1 - (1 - b) (a - 1)/(b + 2 a);
y = 1 - (1 - b) (a - 1)/(b + a + 1);
s0 = -(((2 d - 1) d (1 - k) a)/(d +
1) + ((d - 2) k ((d - 2) a + 2 (d + 1)))/(3 (d + 1)));
W = 1 + (a - 1)/d - (d - 1) (a - 1)^2/(2 d^2);
t0 = W ((d + 1) a + 2 d - 1) ((d - 2) a + 2 (d + 1))/(3 (d + 1));

t2 = Simplify[ D[t0 x^(2 k (d + 1)/3) y^((1 - k) (d + 1)/3), {a,
2}]/(x^(2 k (d + 1)/3) y^((1 - k) (d + 1)/3))];

s2 = Simplify[
D[s0 x^(d k) y^(d (1 - k)), {a, 2}]/(x^(d k) y^(d (1 - k)))];
p = Simplify[s2 + t2 /. {a -> 1 + 3 r}];
Resolve[Exists[{d, r, k},
p <= 0 && d >= 23 && 0 <= k <= 1 && 0 < r <= 26/81]]
```

8.1.12 Lemma 3.4.4

All of the queries in the following code give the output `True`.

```
w=qh^2 bh x(x-y)-qh(bh(x^2-3x y+x+(y-1)y)+y)+(y-1)(bh y-1);
lu=D[fu,b];
ru=-fu^2 d ((1-bh x)^(qh-1)(1-bh(1-(qh-1)x-y)))^(d/qh) w/
   ((1-bh y)^(d+1)(1-bh x)(1-bh(1-(qh-1)x-y)));
FullSimplify[lu==ru]
FullSimplify[w==1-q y+bh((q y-1)x+(qh x+y-1)(qh-1)(x-y))]
ll=D[fl,b];
rl=fl^2 ((2x+(qh-1)y-1)(1-bh(1-x-(qh-1)y))^(d-1)+(qh-1)(x-y)(1-bh y)^(d-1))d/
   (1-bh x)^(d+1);
```

```
FullSimplify[ll==rl]
```

8.1.13 Lemma 3.4.9

The following code checks that, for all d-tuples $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)})$ with $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)} \in \mathrm{Ex}_q(\alpha)$, it holds that $h_{c_1,c_2,\beta_*}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}) < \alpha^{1/d}$. The output is `True`, and the same is still true when the first line changes to $d = 5$ and $d = 6$.

```
q = 4; d = 4; b = 1 - q/(d + 1);
EX = Tuples[{1, alpha}, q-1];
G[vector_, colour_] := 1 - (1-b) (vector[[colour]]- 1)/
(b + Sum[vector[[cc]], {cc,1,q-1}]);
dTUPLES=Tuples[EX,d]; L=Length[dTUPLES];
UNIQ = True;
For[l = 1, l<= L && UNIQ, l++,
currenttuple=dTUPLES[[l]];
For[k=1, k <= d, k++,
vectorofchild[k]=currenttuple[[k]];
];
For[colour = 1, colour <= q-1, colour++,
g[colour] = Product[  G[vectorofchild[k], colour], {k, 1, d}];
];
h = 1 + (1 - b) (1 - g[1])/(b + Sum[g[c],{c,1,q-1}])/.{alpha->u^d};
CHECK = Resolve[Exists[u, h >= u && u > 1]];
If[CHECK == True, UNIQ = False];
];
Print[UNIQ]
```

8.1.14 Lemma 3.4.10

The code checks that all of the desired inequalities are satisfied. The output is `True`.

```
Flag = True;
u[0] = 1;
l[0] = 0;
For[dd = 7, dd <= 9, dd++,
```

```
(ffu = Ceiling[10000 fu /. {q -> 4, d -> dd, b -> 1 - 4/(dd + 1)}]/10000;
 ffl = Floor[10000 fl /. {q -> 4, d -> dd, b -> 1 - 4/(dd + 1)}]/10000;
 For[n = 0, n <= 60, n++,
 (u[n + 1] = ffu /. {x -> u[n], y -> l[n]};
  l[n + 1] = ffl /. {x -> u[n], y -> l[n]};
  Flag = Flag && u[n] >= u[n + 1] && l[n] <= l[n + 1] &&
   3 u[n] + l[n] >= 1 >= 3 l[n] + u[n])];
 Flag = Flag && u[60]/l[60] <= 41/10;)];
Flag
```

8.1.15 Lemma 3.4.13

The output is `true`, and the same is still true if we change d to 8 and 9.

```
x = 1 - (1 - b) (a - 1)/(b + 3 a);
y = 1 - (1 - b) (a - 1)/(b + 2 a + 1);
z = 1 - (1 - b) (a - 1)/(b + a + 2);
g1 = x^d0 y^(d1 + d2) z^d3;
g3 = x^d0 y^(d1 + d4) z^d5;
g4 = x^d0 y^(d2 + d4) z^d6;
h = 1 + (1 - b) (1 - g1)/(b + g1 + g3 + g4);
d = 7; b = 1 - 4/(d + 1);
UNIQ = True;
For[k0 = 0, k0 <= d, k0++,
For[k1 = 0, k0 + k1 <= d, k1++,
For[k2 = 0, k0 + k1 + k2 <= d, k2++,
For[k3 = 0, k0 + k1 + k2 + k3 <= d, k3++,
For[k4 = 0, k0 + k1 + k2 + k3 + k4 <= d, k4++,
For[k5 = 0, k0 + k1 + k2 + k3 + k4 + k5 <= d, k5++,
k6 = d - (k0 + k1 + k2 + k3 + k4 + k5);
hk = h/.{d0->k0, d1->k1, d2->k2, d3->k3, d4->k4, d5->k5, d6->k6};
CHECK = Resolve[Exists[u, 1 < u <= 41/10 && (hk /. {a -> u^d}) >= u]];
If[CHECK == True, UNIQ = False];
]]]]]]
```

```
Print[UNIQ]
```

8.2 Verifying inequalities in Chapter 4

8.2.1 Initialization

We use following code to initilize our computer assisted part of the proof.

```
Phi[x_] := 1/x/(1/2 - x);
M = (Maximize[{1/Phi[x]/(1-x), 0<=x<=1/2}, {x}])[[1]];
F[k_] := (1-f[k])(1-y[k])/Sum[(1-f[i])(1-y[i]), {i,1,4}];
```

8.2.2 Lemma 4.4.6

```
G[w_,f_] := (1-f)/Phi[1-w/(1-f)] + 4M w/(1-f);
Resolve[Exists[{w1,w2,f1,f2},
 (G[w1,f1]+G[w2,f2])>=10195/10000(2G[(w1+w2)/2,(f1+f2)/2]) &&
 1/13<=f1<=1/2 && 1/13<=f2<=1/2 && 1/2<=w1/(1-f1)<=1 &&
 1/2<=w2/(1-f2)<=1]]
Resolve[Exists[{w1,w2,f1,f2},
 (G[w1,f1]/3 +2G[w2,f2]/3)>=10181/10000(G[(w1+2w2)/3,(f1+2f2)/3]) &&
 1/13<=f1<=1/2 && 1/13<=f2<=1/2 && 1/2<=w1/(1-f1)<=1 && 1/2<=w2/(1-f2)<=1]]
```

8.2.3 Lemma 4.4.8

```
Asym = Phi[F[1]] F[1]
 ((1-F[1])/(1-y[1])/Phi[y[1]]+Sum[F[j]/(1-y[j])/Phi[y[j]], {j,2,4}] +
 4f[1]M Sum[F[k]/(1-f[k]),{k,2,4}]);
Alpha = Asym/.
 {f[2]->(1-f[1])/3,f[3]->(1-f[1])/3,f[4]->(1-f[1])/3,y[3]->y[2],y[4]->y[2]};
Resolve[Exists[{f[1],y[1],y[2]},
 Alpha>963/1000 && 1/13<=f[1]<=1/2 && 0<=y[1]<=1/2 && 0<=y[2]<=1/2]]
```

8.2.4 Lemma 4.4.10, case $f_1 = (1 - f_2)/6$

```
F1 = F[1]/.{f[4]->(1-f[1]-f[2])/2, f[3]->(1-f[1]-f[2])/2,y[4]->y[3]};
A = (1-f[1])(1-y[1])+(1-f[2])(1-y[2])+(1+f[1]+f[2])(1-y[3]);
```

```
P1 = A(1-F1)(1/(1-y[1])/Phi[y[1]]-4M f[2]/(1-1/13));
P2 = (1-f[2])/Phi[y[2]]+4f[1]M(1-y[2]);
P3 = (1+f[1]+f[2])/Phi[y[3]]+8M(f[1]+f[2])(1-y[3]);
Alpha1 = (1/A/(1/2-F1)*(P1+P2+P3))/.{f[1]->(1-f[2])/6};
Resolve[Exists[{f[2],y[1],y[2],y[3]}, Alpha1>9138/10000 &&
 0<=f[2]<=1/2 && 0<= y[1]<=1/2 && 0<=y[2]<=1/2 && 0<=y[3]<=1/2]]
```

8.2.5 Lemma 4.4.10, case $f_1 = (3/4 - f_2)/4$

```
F1 = F[1]/.{f[4]->(1-f[1]-f[2])/2, f[3]->(1-f[1]-f[2])/2, y[4]->y[3]};
A =  (1-f[1])(1-y[1])+(1-f[2])(1-y[2])+(1+f[1]+f[2])(1-y[3]);
P1 = A(1-F1)(1/(1-y[1])/Phi[y[1]]-4M f[2]/(1-1/13));
P2 = (1-f[2])/Phi[y[2]]+4f[1]M(1-y[2]);
P3 = (1+f[1]+f[2])/Phi[y[3]]+8M(f[1]+f[2])(1-y[3]);
Alpha2 = (1/A/(1/2-F1)*(P1+P2+P3))/.{f[1]->(3/4-f[2])/4};
Resolve[Exists[{f[2],y[1],y[2],y[3]}, Alpha2>9163/10000 &&
 0<=f[2]<=1/2 && 0<= y[1]<=1/2 && 0<=y[2]<=1/2 && 0<=y[3]<=1/2]]
```

8.2.6 Lemma 4.4.10, case $f_1 = 1/13$

```
F1 = F[1]/.{f[4]->(1-f[1]-f[2])/2, f[3]->(1-f[1]-f[2])/2, y[4]->y[3]};
A = (1-f[1])(1-y[1])+(1-f[2])(1-y[2])+(1+f[1]+f[2])(1-y[3]);
P1 = A(1-F1)(1/(1-y[1])/Phi[y[1]]-4M f[2]/(1-1/13));
P2 = (1-f[2])/Phi[y[2]]+4f[1]M(1-y[2]);
P3 = (1+f[1]+f[2])/Phi[y[3]]+8M(f[1]+f[2])(1-y[3]);
Alpha3 = (1/A/(1/2-F1)*(P1+P2+P3))/.{f[1]->1/13};
Resolve[Exists[{f[2],y[1],y[2],y[3]}, Alpha3>9102/10000 &&
 0<=f[2]<=1/2 && 0<= y[1]<=1/2 && 0<=y[2]<=1/2 && 0<=y[3]<=1/2]]
```

In order to speed up above three code snippets for Lemma 4.4.10, we can simplify $\texttt{Alpha}_i$ to the form $\frac{A}{B}$ where both $A \geq 0$ and $B \geq 0$ are polynomials. To verify $\frac{A}{B} \leq \alpha$, it is equivalent to verify $A - \alpha B \leq 0$, which can be done more efficiently by Mathematica.

8.2.7 Lemma 4.4.11

```
G[w_,f_] := (1-f)/Phi[1-w/(1-f)] + M w/(1-f);
```

```
Resolve[Exists[{w1,w2,f1,f2},
 (G[w1,f1]+G[w2,f2])>=1009/1000(2G[(w1+w2)/2, (f1+f2)/2]) &&
 0<=f1<=1/2 && 0<=f2<=1/2 && 1/2<=w1/(1-f1)<=1 && 1/2<=w2/(1-f2)<=1]]
Resolve[Exists[{w1,w2,f1,f2},
 (G[w1,f1]/3 + 2G[w2,f2]/3)>=1009/1000(G[(w1+2w2)/3, (f1+2f2)/3]) &&
 0<=f1<=1/2 && 0<=f2<=1/2 && 1/2<=w1/(1-f1)<= 1 && 1/2<=w2/(1-f2)<=1]]
```

8.2.8 Lemma 4.4.12

```
F1 = F[1]/.{f[4]->(1-f[1]-f[2])/2, f[3]->(1-f[1]-f[2])/2, y[4]->y[3]};
A = (1-f[1])(1-y[1])+(1-f[2])(1-y[2])+(1+f[1]+f[2])(1-y[3]);
P1 = A(1-F1)(1/(1-y[1])/Phi[y[1]]-2M f[2]/(1-f[1]));
P2 = (1-f[2])/Phi[y[2]]+ 2f[1] M(1-y[2]);
P3 = (1+f[1]+f[2])/Phi[y[3]]+4M(f[1]+f[2])(1-y[3]);
Alpha = (1/A/(1/2-F1)*(P1+P2+P3))/.{f[1]->0, y[1]->0};
Resolve[Exists[{f[2],y[2],y[3]},
 Alpha>9231/10000 && 1/13<=f[2]<=1/2 && 0<=y[2]<=6/13 && 0<=y[3]<=6/13]]
```

Bibliography

[1] David Aldous. Random walks on finite groups and rapidly mixing markov chains. In *Séminaire de Probabilités XVII 1981/82*, pages 243–297. Springer, 1983.

[2] K. S. Alexander. Mixing properties and exponential decay for lattice systems in finite volumes. *Ann. Probab.*, 32(1A):441–487, 2004.

[3] Alexander Barvinok. *Combinatorics and Complexity of Partition Functions*. Springer Publishing Company, Incorporated, 1st edition, 2017.

[4] V. Beffara and H. Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probability Theory and Related Fields*, 153(3):511–542, 2012.

[5] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

[6] Ivona Bezáková, Andreas Galanis, Leslie A. Goldberg, Heng Guo, and Daniel Štefankovič. Approximation via Correlation Decay when Strong Spatial Mixing Fails. *SIAM Journal on Computing*, 48(2):279–349, 2019.

[7] Antonio Blanca, Andreas Galanis, Leslie Ann Goldberg, Daniel Štefankovič, Eric Vigoda, and Kuan Yang. Sampling in Uniqueness from the Potts and Random-Cluster Models on Random Regular Graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[8] Elmar Böhler, Steffen Reith, Henning Schnoor, and Heribert Vollmer. Bases for boolean co-clones. *Inf. Process. Lett.*, 96(2):59–66, 2005.

[9] M. Bordewich, C. Greenhill, and V. Patel. Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Structures & Algorithms*, 48(1):21–52, 2016.

[10] G. R. Brightwell and P. Winkler. Random colorings of a Cayley tree. *Contemporary combinatorics*, 10:247–276, 2002.

[11] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS'97)*, pages 223–231. IEEE, 1997.

[12] Russ Bubley, Martin Dyer, Catherine Greenhill, and Mark Jerrum. On approximately counting colorings of small degree graphs. *SIAM Journal on Computing*, 29(2):387–400, 1999.

[13] Andrei A. Bulatov, Venkatesan Guruswami, Andrei Krokhin, and Dániel Marx. The Constraint Satisfaction Problem: Complexity and Approximability (Dagstuhl Seminar 15301). *Dagstuhl Reports*, 5(7):22–41, 2016.

[14] Jin-Yi Cai, Andreas Galanis, Leslie A. Goldberg, Heng Guo, Mark Jerrum, Daniel 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.

[15] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. The Complexity of Complex Weighted Boolean #CSP. *J. Comput. Syst. Sci.*, 80(1):217–236, 2014.

[16] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2216–2234. SIAM, 2019.

[17] Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Inf. Comput.*, 125(1):1–12, 1996.

[18] Nadia Creignou, Sanjeev Khanna, and Madhu Sudan. *Complexity Classifications of Boolean Constraint Satisfaction Problems*. Society for Industrial and Applied Mathematics, 2001.

[19] Nadia Creignou, Phokion G. Kolaitis, and Bruno Zanuttini. Preferred representations of boolean relations. *Electronic Colloquium on Computational Complexity (ECCC)*, (119), 2005.

[20] Víctor Dalmau and Daniel K. Ford. Generalized satisfability with limited occurrences per variable: A study through delta-matroid parity. In *Mathematical Foundations of Computer Science 2003, 28th International Symposium, MFCS 2003, Bratislava, Slovakia, August 25-29, 2003, Proceedings*, pages 358–367, 2003.

[21] A. Dembo, A. Montanari, A. Sly, and N. Sun. The replica symmetric solution for Potts models on d-regular graphs. *Communications in Mathematical Physics*, 327(2):551–575, 2014.

[22] Martin Dyer, Abraham D Flaxman, Alan M Frieze, and Eric Vigoda. Randomly coloring sparse random graphs with fewer colors than the maximum degree. *Random Structures & Algorithms*, 29(4):450–465, 2006.

[23] Martin Dyer and Alan Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures & Algorithms*, 23(2):167–179, 2003.

[24] Martin Dyer, Alan Frieze, Thomas P Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. *Random Structures & Algorithms*, 43(2):181–200, 2013.

[25] Martin Dyer, Leslie A. Goldberg, Catherine Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2003.

[26] Martin E. Dyer, Leslie A. Goldberg, and Mark Jerrum. An approximation trichotomy for Boolean #CSP. *J. Comput. Syst. Sci.*, 76(3-4):267–277, 2010.

[27] Martin E. Dyer, Leslie Ann Goldberg, Markus Jalsenius, and David Richerby. The Complexity of Approximating Bounded-Degree Boolean #CSP. In *The 27th International Symposium on Theoretical Aspects of Computer Science*, STACS '10, pages 323–334, 2010.

[28] Martin E. Dyer and Catherine S. Greenhill. On markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.

[29] C. Efthymiou. A simple algorithm for sampling colorings of $G(n, d/n)$ up to the Gibbs uniqueness threshold. *SIAM Journal on Computing*, 45(6):2087–2116, 2016.

[30] C. Efthymiou, T. P. Hayes, D. Štefankovič, and E. Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '18, pages 1759–1771, 2018.

[31] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2015.

[32] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6):50:1–50:60, 2015.

[33] A. Galanis, D. Štefankovič, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016.

[34] Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Improved inapproximability results for counting independent sets in the hard-core model. In *The 15th International Workshop on Randomization and Computation*, RANDOM '11, 2011.

[35] Andreas Galanis and Leslie A. Goldberg. The complexity of approximately counting in 2-spin systems on k-uniform bounded-degree hypergraphs. *Information and Computation*, 251:36–66, 2016.

[36] Andreas Galanis, Leslie Ann Goldberg, and Kuan Yang. Approximating Partition Functions of Bounded-Degree Boolean Counting Constraint Satisfaction Problems. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[37] Andreas Galanis, Leslie Ann Goldberg, and Kuan Yang. Uniqueness for the 3-state antiferromagnetic potts model on the tree. *Electron. J. Probab.*, 23:43 pp., 2018.

[38] 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(04):500–559, 2016.

[39] David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting colorings of a graph. *Journal of Discrete Algorithms*, 12:29–47, 2012.

[40] Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

[41] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. Counting hypergraph colourings in the local lemma regime. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 926–939, New York, NY, USA, 2018. ACM.

[42] Thomas P Hayes. Randomly coloring graphs of girth at least five. In *Proceedings of the 35th Annual ACM Symposium on Symposium on Theory of Computing (STOC'03)*, pages 269–278. ACM, 2003.

[43] Thomas P Hayes and Eric Vigoda. A non-markovian coupling for randomly sampling colorings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)*, pages 618–627. IEEE, 2003.

[44] Thomas P Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *The Annals of Applied Probability*, 16(3):1297–1318, 2006.

[45] Jonathan Hermon, Allan Sly, and Yumeng Zhang. Rapid mixing of hypergraph independent sets. *Random Structures & Algorithms*, 54(4):730–767, 2019.

[46] M R Jerrum, L G Valiant, and V V Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, July 1986.

[47] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–166, 1995.

[48] Mark Jerrum and Alistair Sinclair. Polynomial-Time Approximation Algorithms for the Ising Model. *SIAM Journal on Computing*, 22(5):1087–1116, 1993.

[49] J. Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002.

[50] Johan Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002.

[51] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[52] P. W. Kasteleyn. Graph theory and crystal physics. In *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[53] Liang Li, Pinyan Lu, and Yitong Yin. Approximate counting via correlation decay in spin systems. In *Proceedings of the 23th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*, pages 922–940. SIAM, 2012.

[54] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13)*, pages 67–84. SIAM, 2013.

[55] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. A deterministic algorithm for counting colorings with 2-δ colors. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1380–1404. IEEE Computer Society, 2019.

[56] Pinyan Lu, Kuan Yang, Chihao Zhang, and Minshen Zhu. An FPTAS for Counting Proper Four-Colorings on Cubic Graphs. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1798–1817. SIAM, 2017.

[57] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *Proceedings of APPROX-RANDOM*, pages 639–654. Springer, 2013.

[58] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.*, 161(3):447–486, 1994.

[59] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Communications in Mathematical Physics*, 161(3):487–514, 1994.

[60] F. Martinelli, E. Olivieri, and R. H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics*, 165(1):33–47, 1994.

[61] Ankur Moitra. Approximate counting, the Lovász Local Lemma and inference in graphical models. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 356–369, New York, NY, USA, 2017. ACM.

[62] Michael Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM Journal on Computing*, 33(3):721–737, 2004.

[63] E. Mossel and A. Sly. Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Structures & Algorithms*, 35(2):250–270, 2009.

[64] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *Ann. Probab.*, 41(1):294–328, 2013.

[65] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3):401–439, 2009.

[66] 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.

[67] F Peruggi, F di Liberto, and G Monroy. The Potts model on Bethe lattices. I. General results. *Journal of Physics A: Mathematical and General*, 16(4):811, 1983.

[68] Jesús Salas and Alan D Sokal. Absence of phase transition for antiferromagnetic potts models via the dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86(3-4):551–579, 1997.

[69] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: optimal bounds. *Probability Theory and Related Fields*, 168(1):153–197, 2017.

[70] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *Journal of Statistical Physics*, 155(4):666–686, 2014.

[71] Allan Sly. Uniqueness thresholds on trees versus graphs. *The Annals of Applied Probability*, 18(5):1897–1909, 2008.

[72] Allan Sly, Nike Sun, et al. Counting in two-spin models on d-regular graphs. *The Annals of Probability*, 42(6):2383–2416, 2014.

[73] L. E. Thomas. Bound on the mass gap for finite volume stochastic Ising models at low temperature. *Communications in Mathematical Physics*, 126(1):1–11, 1989.

[74] L. G. Valiant and V. V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47(1):85–93, 1986.

[75] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[76] Leslie G. Valiant and Vijay V. Vazirani. NP is as easy as detecting unique solutions. *Theor. Comput. Sci.*, 47(3):85–93, 1986.

[77] Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.

[78] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC'06)*, pages 140–149. ACM, 2006.

[79] Y. Yin and C. Zhang. Sampling in Potts model on sparse random graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016*, pages 47:1–47:22, 2016.