PROBLEMS IN CATALAN MIXING AND MATCHINGS IN REGULAR HYPERGRAPHS

A Thesis
Presented to
The Academic Faculty

by

Emma Cohen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Algorithms, Combinatorics, and Optimization

School of Mathematics
Georgia Institute of Technology
December 2016

PROBLEMS IN CATALAN MIXING AND MATCHINGS IN REGULAR HYPERGRAPHS

Approved by:

Prasad Tetali, Advisor
School of Mathematics and School of Computer Science
Georgia Institute of Technology

Dana Randall
School of Computer Science and School of Mathematics
Georgia Institute of Technology

Santanu Dey
School of Industrial and Systems Engineering
Georgia Institute of Technology

William T. Trotter
School of Mathematics
Georgia Institute of Technology

Milena Mihail
School of Computer Science
Georgia Institute of Technology

Date Approved: Not Yet Approved

To my mother

ACKNOWLEDGEMENTS

I want to thank my advisor Prasad Tetali; my committee other members Santanu Dey, Milena Mihail, Dana Randall, and Tom Trotter; the reader David Galvin; my coauthors Péter Csikvári, Will Perkins, and Damir Yeliussizov; my fiancé Christopher Berlind; my family and friends (who are too many to list).

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

This dissertation consists of two parts, falling under the closely related fields of counting and sampling.

In the first part, we explore the relationships between several natural notions of adjacency on Catalan structures and their associated random walks. We use a matroid interpretation of Dyck words of length $2n$ to give a new order n^2 bound on the mixing time for the transposition walk. We also give a general mixing bound for random walks on the Boolean cube when censored to remain within some large monotone subset.

In the second part, we extend several related extremal results about the number of matchings and independent sets in regular graphs. First we propose a method for tackling the Upper Matching Conjecture of Friedland, et al. for matchings of small fixed size. Next we prove a conjecture of Galvin regarding the extremal graph for number of Widom-Rowlinson configurations, a particular instance of graph homomorphisms. Finally, we make progress towards unifying the extremal bounds of Kahn, Galvin & Tetali, and Zhao for independent sets and of Davies, et al., for matchings by giving two general bounds for matchings in regular, uniform hypergraphs, improving on a similar bound due to Ordentlich & Roth.

CHAPTER I

INTRODUCTION

1.1 Overview

This dissertation consists of two parts, falling under the closely related fields of counting and sampling.

In the first part, we start in Chapter 2 by exploring the relationships between several natural notions of adjacency on Catalan structures and their associated random walks. In Chapter 3 we derive a new order n^2 bound on the mixing time for the transposition walk by combining a matroid interpretation of Dyck paths of length $2n$ due to Ardila (2003) with a general bound on the mixing time for balanced matroid basis exchange walks given by Feder & Mihail (1992). Finally, in Chapter 4 we give a general mixing bound for random walks on the Boolean cube when censored to remain within some large monotone subset, improving on a similar bound due to Ding & Mossel (2014).

In the second part, we extend several related extremal results about the number of matchings and independent sets in regular graphs. In Chapter 5 we propose a method for tackling the Upper Matching Conjecture of Friedland, et al. (2008)—that a union of complete bipartite graphs has the most matchings of size t among d-regular graphs on n vertices—as it applies to matchings of small fixed size. Next, in Chapter 6 we discuss an analogous question for graph homomorphisms to a fixed target graph H, a broad class of problems which includes independent sets and colorings. We prove a conjecture of Galvin regarding the extremal graph for the special case of counting Widom-Rowlinson configurations in regular graphs. Finally, in Chapter 7 we make progress towards unifying the extremal bounds of Kahn (2001), Galvin & Tetali

(2004), and Zhao (2009) for independent sets and of Davies et al. for matchings by giving two general bounds for matchings in regular, uniform hypergraphs, improving on a similar bound due to Ordentlich & Roth (2004).

1.2 Notation

A *graph* $G = (V, E)$ is a set of vertices V along with a collection of edges E consisting of unordered pairs of vertices. Either E or its elements may in fact be multisets; an edge of the form $\{v, v\}$ is called a *loop* at $v \in V$. We may refer to the vertex set and edge set of G by $V(G)$ and $E(G)$, respectively, and write $n(G) := |V(G)|$ and $e(G) := |E(G)|$, respectively. If the graph G is clear from context we may omit it from the notation.

The degree of a vertex $v \in V(G)$ is $d(v) = d_G(v) := |\{e \in E(G) \mid v \in e\}|$. Note that a loop at v only counts once towards its degree. A graph is *d-regular* if $d(v) = d$ for all $v \in V(G)$.

For a graph G and $A \subseteq V(G)$, the subgrpah of G *induced* on A has vertex set A and edges $\{e \in E(G) \mid e \subseteq A\}$, and is denoted $G[A]$.

For integers $a \leq b$ will write $[a, b]$ for the closed discrete interval $\{a, a + 1, \ldots, b\}$ and $[a] := [1, a]$. (This notation overlaps with the common notation for real intervals, but it should be clear from context which meaning is intended.) For sets A, B we also adopt the standard notations $A^B := \{f : B \to A\}$ and $\binom{A}{a} := \{S \subseteq A \mid |S| = a\}$.

PART I

CHAPTER II

CATALAN MIXING

2.1 Introduction

There are many natural classes of combinatorial structures which are enumerated by the Catalan sequence: $1, 2, 5, 14, 42, \ldots$ (OEIS A000108), the nth term of which is the Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$. Some examples include the set of all triangulations of a regular polygon of $n+2$ sides, the set of non-crossing partitions of an n-set (the lattice on which is of much interest to researchers in free probability), and the set of strings consisting of n balanced pairs of parentheses—or equivalently, described as *Catalan strings* $x \in \{\pm 1\}^{2n}$ of n 1's and n -1's with non-negative partial sums, $\sum_{i=1}^{j} x_i \geq 0$, for all j. This last structure (of Catalan strings) is also known as the set of Dyck paths, and is visualized as lattice paths of n up-steps and n down-steps, representing the 1's and the -1's respectively. (See Figure 1.)

For any given $n \geq 1$, generating a Catalan structure (such as triangulation or a Dyck path) uniformly at random from the set of C_n many, is a straightforward task, and can be done in time linear in n, not too unlike generating uniformly at

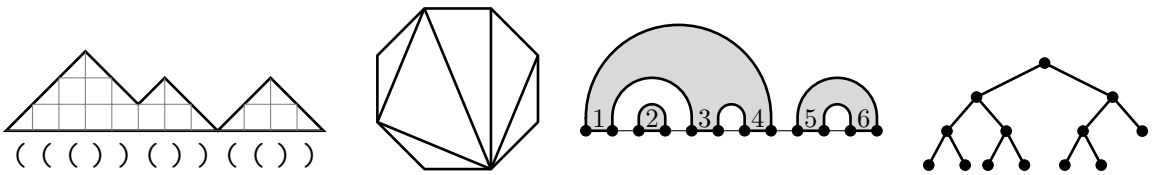


Figure 1: Some Catalan structures for $n = 6$: (a) a Dyck path of length 12 and its corresponding parenthesization; (b) a triangulation of an octagon; (c) the noncrossing partition $\{\{1, 3, 4\}, \{2\}, \{5, 6\}\}$ of $\{1, ..., 6\}$, represented as a noncrossing matching on 12 points; (d) a binary plane tree with 6 internal nodes.

random one of the $n!$ permutations of $[n]$. Furthermore, natural bijections between Catalan realizations ensure that the ability to sample from any one Catalan realization translates easily to sampling from the others. However, the study (of convergence to equilibrium) of particular Markov chain Monte Carlo algorithms which yield a random Catalan structure is seemingly much more interesting. The inspiration for this endeavor in part stems from an open problem of Aldous [4], who conjectured that the random walk on triangulations of a polygon on n sides, performed using uniform diagonal flips, ought to take time roughly $n^{3/2}$ (up to factors logarithmic in n). Despite much effort by various researchers in this topic over several years, the best known bounds for the mixing time of the chain on triangulations remain those of McShine & Tetali [55], who proved an upper bound of $O(n^4)$ on the relaxation time (also known as the inverse spectral gap), and Molloy, Reed & Steiger [56] who showed $\Omega(n^{3/2})$ as a lower bound.

In general, tight analysis of natural Markov chains on Catalan structures seems to be a challenging problem. To date, the only known sharp upper bounds for Catalan walks are those of Wilson [71] for adjacent transpositions on Catalan strings (which follows from a more general analysis of adjacent transpositions in lattice paths) and the usual bounds on mixing times for the (almost trivial) Catalan realization consisting of cyclic strings of n zeros and $n+1$ ones (as discussed in [17]).

The remainder of this chapter gives an overview of some of the most natural and interesting Markov chains on Catalan structures, and discusses their similarities and differences. In Chapter 3 we use a different method to give a new bound for the mixing time of the Markov chain on Catalan strings whose moves consist of transposing any two elements (subject to maintaining the non-negativity of partial sums). Our analysis generalizes to a larger class of random walks on constrained lattice paths. Finally, in Chapter 4 we discuss a more general approach to bounding mixing times for Markov chains on the Boolean cube subject to a monotone constraint.

2.1.1 Markov chain preliminaries

A Markov chain is a sequence of random variables $(X_i)_{i\geq 0}$ from a (finite) state space Ω such that the conditional distribution of X_{t+1} given X_t is independent of the history $(X_i)_{0\leq i<t}$ (i.e., the sequence is *memoryless*). The Markov chain is *homogeneous* if this conditional distribution is independent of the time t, in which case we can describe it by its *transition matrix* P, the stochastic matrix with entries $P(x,y) = \Pr(X_i = y \mid X_{i-1} = x)$. If the state at a given time has probability distribution vector f, then the distribution of the next step is Pf and after t steps the distribution is $P^t f$.

In this dissertation we will only be concerned with homogeneous Markov chains with symmetric transition matrices. The *transition graph* of a symmetric Markov chain is the graph with vertex set Ω and $x, y \in \Omega$ adjacent if $P(x,y) > 0$, representing the possible transitions. One can think of a symmetric Markov chain as a random walk on its transition graph. A symmetric Markov chain is *ergodic* if its transition graph is connected and non-bipartite (taking loops into account), and it is well-known that in this case the limiting distribution of $P^t(x,\cdot)$ as $t \to \infty$ is uniform (for every initial state x). This *stationary distribution* π satisfies $P\pi = \pi$.

Definition 2.1. The *mixing time* $t_{\text{mix}} = t_{\text{mix}}(\epsilon)$ of a Markov chain is the smallest time t such that

$$\frac{1}{2}\sum_{y\in\Omega} |P^{t'}(x,y) - \pi(y)| < \epsilon$$

for all initial states x and at every time $t' \geq t$.

In other words, the mixing time is the number of steps necessary to ensure that the state is close to uniform random (regardless of the initial state).

Most of the literature studying Markov chains is devoted to bounding mixing times, and a wide variety of methods has been developed to this end (see [49]). One of the most common methods is a *coupling argument* [3, 16], which shows that after a certain amount of time two dependent but individually faithful copies of a Markov

chain are almost surely in the same state; hence if one copy begins at stationarity then at that time the other must be stationary as well, regardless of its initial state. As coupling arguments do not seem to work well for the applications we are interested in, we will instead mostly consider several other Markov chain parameters—conductance, spectral gap, and log-Sobolev constant—which are closely related to the mixing time and are often easier to bound.

Definition 2.2. The *spectral gap* γ and log-Sobolev constant α of a symmetric Markov chain are

$$\gamma := \inf_f \frac{\mathcal{E}(f)}{\mathrm{Var}(f)} \qquad \text{and} \qquad \alpha := \inf_f \frac{\mathcal{E}(f)}{\mathcal{L}(f)},$$

where

$$\begin{aligned}
\mathcal{E}(f) &:= \frac{1}{2} \sum_{x,y \in \Omega} (f(x) - f(y))^2 P(x,y)\pi(x), \\
\mathrm{Var}(f) &:= \frac{1}{2} \sum_{x,y \in \Omega} (f(x) - f(y))^2 \pi(y)\pi(x), \qquad \text{and} \\
\mathcal{L} &:= \sum_{x \in \Omega} f(x)^2 (\log f(x)^2 - \log \mathrm{E}(f^2))\pi(x),
\end{aligned}$$

and the infima are taken over non-constant functions $f : \Omega \to \mathbb{R}$.

The term *spectral gap* originates from the equivalent definition of γ as the difference between the largest and second-largest eigenvalues of the transition matrix.

The inverse of the spectral gap, $1/\gamma$, is called the *relaxation time*, and is well-known to be closely related to the mixing time:

Theorem 2.3. *For a symmetric Markov chain,*

$$\frac{1}{\gamma} \leq t_{mix}(1/e) \leq \frac{2 + \log|\Omega|}{2\gamma}.$$

There is a tighter relationship betwen the log-Sobolev constant and the mixing time, due to Diaconis [26]

Theorem 2.4. *For a symmetric Markov chain,*

$$\frac{1}{2\alpha} \leq t_{mix}(1/e) \leq \frac{4 + \log\log|\Omega|}{4\alpha}.$$

Thus in order to bound the mixing time it suffices to bound γ or α (although such a bound may not give a tight bound on the mixing time.) In more combinatorial situations it is sometimes easier to bound the conductance.

Definition 2.5. The *conductance* (or *bottleneck ratio*) of a symmetric Markov chain is

$$\Phi := \min_{\substack{S \subset \Omega \\ \pi(S) \geq 1/2}} \frac{\sum_{x,y \in S} P(x,y)}{|S|}.$$

The conductance is closely related to the spectral gap, and in fact can be formulated (up to a constant factor) in the same way but with the infimum only over functions $f : \Omega \to \{0, 1\}$. Indeed, the conductance cannot be too far away from the spectral gap:

Theorem 2.6 ([66, 48])**.** *The conductance Φ and spectral gap γ of a reversible Markov chain satisfy*

$$\frac{\Phi^2}{2} \leq \gamma \leq 2\Phi.$$

There are several Markov chain comparison theorems which allow us to bound these parameters of a random walk given a known bound for a similar walk on the same state space, such as the following bounds described by Randall & Tetali (among others [25, 57]). The theorems extend to general reversible Markov chains, but we state them here in the simpler case of symmetric Markov chains.

Theorem 2.7 ([63])**.** *Let P and $\tilde{P}$ be symmetric transition matrices on the same state space Ω with transition graphs $G = (\Omega, E)$ and $\tilde{G} = (\Omega, \tilde{E})$, respectively. For each $x, y \in \tilde{E}$ let $\Gamma_{x,y}$ be an x-y path in G. Define the* congestion ratio *of the set of*

canonical paths Γ *to be*

$$B := \max_{(z,w)\in E} \sum_{\substack{(x,y)\in\tilde{E} \\ (z,w)\in\Gamma_{x,y}}} \Big(\frac{\tilde{P}(x,y)}{P(z,w)}|\Gamma_{x,y}|\Big).$$

Then

$$\frac{1}{\gamma} \le B\,\frac{1}{\tilde{\gamma}}, \qquad \frac{1}{\alpha} \le B\,\frac{1}{\tilde{\alpha}}, \qquad \textit{and}$$

$$t_{mix} \le O\big(\log|\Omega| B\,\tilde{t}_{mix}\big)$$

That is, if it is possible to simulate the transitions in $\tilde{P}$ by paths in P without relying too heavily on any given transition, then the mixing time of P cannot be that much larger than the mixing time of $\tilde{P}$.

One would think that such a theorem might be applicable to this situation, where we have several walks on the the space of Catalan structures of order n, which we can identify across realizations by specifying a bijection. However, although there tend to be simple—often almost trivial—bijections between different realizations of the Catalan sequence, the most natural Markov chains for different realizations seem to translate rather unnaturally along these bijections and can indeed be very different chains.

2.2 Some Catalan walks

Here we introduce in more depth the realizations of the Catalan sequence which we will use later. Note that this list of Catalan structures is far from comprehensive; see [69] for a more complete (but still not exhaustive) list of over 200 different realizations of the Catalan numbers. For each realization here we give one or two natural notions of adjacency and discuss the corresponding Markov chains. These Markov chains are all symmetric and ergodic; in fact, each of them can be described as a random walk on its transition graph G, which moves from state x along any edge incident to x, each with transition probability $P(x,y) = p < 1/2\Delta(G)$ for every $y \sim x$, where $\Delta(G)$ is the maximum degree in G (the walk stays at x with the remaining probability).

2.2.1 Catalan strings and Dyck paths

We have already mentioned Catalan strings, but we state the definition again for reference and completeness:

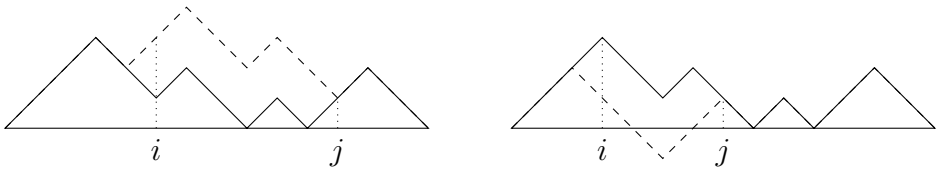
Definition 2.8. A *Catalan string* of order n is a sequence $x \in \{\pm 1\}^{2n}$ of n 1's and n -1's with non-negative partial sums $h_x(i) := \sum_{i=1}^{j} x_i \geq 0$ for all $0 \leq j \leq 2n$.

If the sequence is thought of as a string then the symbols used become rather immaterial, and in many cases it is convenient to use $\{+, -\}$, $\{1, 0\}$, $\{(,)\}$, or $\{\mathsf{U}, \mathsf{D}\}$ (for *up* and *down*) in place of $\{1, -1\}$. This last comes from the common visualization of a Catalan string as a *Dyck path* by plotting the *height* function h_x on the 2-dimensional integer lattice. That is, the path starts at $(0, 0)$ and takes $2n$ steps northeast (up) or southeast (down), returning to $(2n, 0)$ and never passing below the x-axis. (See Figure 1(a).)

Since the numbers of up- and down-steps are constant among Catalan strings of length $2n$, a natural way to design a random walk on the space of C_n such strings is by permuting the string via natural actions of generators of the symmetric group S_{2n}. (For x a string of length k and $\sigma \in S_k$, we will denote the usual group action by $\sigma(x)_i = x_{\sigma(i)}$.) Unfortunately, generating S_{2n} also means that such moves will generate *invalid* strings (which violate the non-negative partial sum constraint), so we must disallow moves which produce invalid strings. This approach yields some natural random Markov chains on the space of Catalan strings.

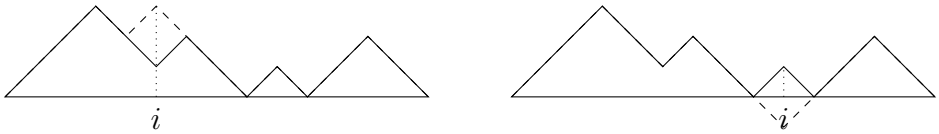
Definition 2.9. From an initial Catalan string x, a step of the (lazy) *random transposition walk* on Catalan strings is as follows:

- Pick two uniform random indices $i, j \in [2n]$. Let $\tau_{i,j} = (i\ j) \in S_{2n}$ be the permutation which transposes indices i and j.
- If $x' = \tau_{i,j}(x)$ is a valid Catalan string, the next state is x'. Otherwise, if x' is invalid, the next state is x.



(a) Valid (b) Invalid

Figure 2: A valid move and an invalid (censored) move of the random transposition walk on Catalan strings.



(a) Valid (b) Invalid

Figure 3: A valid move and an invalid (censored) move of the adjacent transposition walk on Catalan strings.

We can describe a similar walk using only transpositions at adjacent indices.

Definition 2.10. From an initial Catalan string x, a step of the (lazy) *adjacent transposition walk* on Catalan strings is as follows:

- Pick a uniform random index $i \in [2n-1]$.
- If $x' = \tau_{i,i+1}(x)$ is a valid Catalan string, the next state is either x or x' with equal probability. Otherwise, if x' is invalid, the next state is x.

In either case, if the proposed string x' is not a valid Catalan string, the move is said to have been *invalid*, or *censored*. Note that in both cases it is possible to have $x = x'$, but for random transpositions this happens with probability $1/2$ (whenever the chosen indices are either both up-steps or both down-steps), so both walks are lazy. (Laziness is useful to ensure ergodicity of the random walk.)

Of course, similar walks can be made from any symmetric set of generators of S_{2n}. However, one should note that the censoring mechanic for staying within the space of valid Catalan strings favors moves which are more "local," in the sense that the

more the string is changed by a single move the more likely the resulting string is to be invalid and the move censored. For this reason, allowing *more* generators does not necessarily decrease the mixing time, since adding moves which are often censored will decrease the probability of moving at all. For example, one Markov chain which is well-studied on permutations is the *riffle shuffle*, which simulates the standard method of shuffling cards by splitting the deck into two halves (with binomially distributed sizes) and randomly interleaving them, maintaining the order within each half. This mixes very rapidly ($O(\log n)$) for permutations [3], but the censoring step makes direct application to Catalan strings unlikely to mix well. In particular, in a Catalan string one is more likely to find up-steps near the beginning of the string and down-steps near the end, but with its aggressive mixing strategy the riffle shuffle actively tries to fix this. As a result, almost every move will be censored and the Markov chain will go very slowly. It is interesting, however, to ask whether the riffle shuffle can be modified to work well for Catalan strings, yielding a natural, nontrivial Markov chain with polylogarithmic mixing time.

The transition graph for the adjacent transposition walk is actually the Hasse diagram of the Stanley lattice, the partially ordered set (poset) on Dyck paths with comparison $x \leq y$ if the path x is never above the path y. Indeed, we will see that many of the Catalan walks we will study are in fact walks on the Hasse diagrams of related poset lattices.

Without the censoring step we would have a random (or adjacent) transposition walk on the set of $\binom{2n}{n}$ binary strings with an equal number of 1s and -1s. It is well-known (see e.g., [27])—sometimes under the name *Bernoulli-Laplace model*—that $O(n \log n)$ such random transpositions are necessary and sufficient to reach (close to) equilibrium. It is also well-known that the so-called random transposition shuffle of n distinct cards mixes in $\Theta(n \log n)$ time. However, the constraint of having non-negative height seems to thwart any type of straightforward (or otherwise) analyses,

despite the best efforts of several experts in the field.

The uncensored adjacent transposition walk on binary strings (and on permutations) has also been studied. An analysis due to Wilson [71] gives $\Theta(n^3)$ relaxation time and $\Theta(n^3 \log n)$ mixing time for uncensored adjacent transpositions, thought of as lattice paths in $\Theta(n) \times \Theta(n)$ box, and similarly for adjacent transpositions on S_n. In fact, Wilson's lattice path upper bound also generalizes to the case of lattice paths constrained to any interval in the Stanley lattice, i.e., the adjacent transposition walk on paths in the $n \times n$ box, censored to remain between two (comparable) boundary paths. His lower bound relies on the lack of censoring and does not immediately apply to the case of Dyck paths.

We can apply Theorem 2.7 to give a simple polynomial mixing bound for random transpositions, based on Wilson's bound for adjacent transpositions.

Proposition 2.11. *The relaxation time of the random transposition walk on Catalan strings is at most* $O(n^4)$.

Proof. To bound the relaxation time of the random transposition walk by comparison to adjacent transpositions, we must simulate each allowable adjacent transposition $x \to y$ using a path $\Gamma_{x,y}$ consisting of random transpositions. However, each adjacent transposition is itself a random transposition, and so we can use paths of length 1. The congestion ratio is then just the ratio of the transition probabilities, which in this case is $O(n)$, and so the result folows from Theorem 2.7 and Wilson's $\Theta(n^3)$ bound on the relaxation time for adjacent transpositions. □

Since the Catalan numbers satisfy $\log C_n = \Theta(n)$, this gives a mixing time upper bound of $O(n^5)$. Unfortunately, Wilson's analysis does not bound the log-Sobolev constant (which, if it were also $O(n^3 \log n)$, would yield a better bound in this comparison).

This bound, while polynomial, is rather large. In the case of adjacent transpositions the censoring has little effect on the mixing time, and we might expect that the same should be true for random transpositions.

Conjecture 2.12. *The mixing time of the Catalan random transposition walk is* $\Theta(n \log n)$.

There is a nearly trivial $O(n)$ lower bound on the mixing time due to the diameter of the transition graph. While a tight analysis of the mixing time for Catalan random transpositions has remained elusive, we will use more sophisticated methods to give an $O(n^2 \log n)$ upper bound in Chapter 3.

2.2.2 Cyclic strings

Perhaps simultaneously the simplest and the strangest of Catalan realizations is that of cyclic strings.

Definition 2.13. A *cyclic string* of order n is an equivalence class $\overline{x}$ of strings $x \in \{\pm 1\}^{2n+1}$ consisting of $n+1$ ones and n minus ones, modulo the equivalence relation of cyclic rotation (i.e., strings x and y are considered equivalent if $x = \rho^k(y)$ for some k, where $\rho = (1\ 2\ \ldots\ 2n{+}1) \in S_{2n+1}$ is the long cycle).

It is not hard to see that the $2n+1$ cyclic rotations of such a string are all distinct, making it clear that the number of cyclic strings of order n is indeed $C_n = \frac{1}{2n+1}\binom{2n+1}{n} = \frac{1}{n+1}\binom{2n}{n}$. In fact, one of the best-known proofs of the formula (due to Narayana [59]) for enumerating Dyck paths is via a bijection between cyclic strings and Catalan strings: each cyclic string equivalence class has a unique representative of the form $1x$, where x is a valid Catalan string.

Cyclic strings are a very simple Catalan realization, in that they are easy to construct and easy to count. However, they are somewhat anomalous as a Catalan realization in that they do not exhibit in any obvious way the same recursive structure that is common to almost every other Catalan realization (see Section 2.2.6).

The class of cyclic strings is very similar to that of Catalan strings (as evidenced by the simple bijection), but is somewhat more forgiving in terms of making local changes, since there are no invalid strings. Indeed, Caputo [17] observed that it is easy to construct any number of efficient Markov chains on cyclic strings.

Definition 2.14. Let S be a symmetric set of generators of S_{2n+1} which is closed under conjugation by $\rho = (1\ 2\ \ldots\ 2n+1)$. From an initial cyclic string $\overline{x}$ with representative $x \in \overline{x}$, a step of the (lazy) *S-permutation walk* on cyclic strings is as follows:

- Pick a uniform random element $\sigma \in X$, and let $x' = \sigma(x)$ (where $\sigma(x)_i = x_{\sigma(i)}$ is the natural group action).
- The next state is either $\overline{x}$ or $\overline{x}'$ with equal probability.

Since X is closed under conjugation by r the transition probability of moving from $\overline{x}$ to $\overline{x}'$ is independent of the representative x, and so the Markov chain on equivalence classes is well-defined. Furthermore, since there are no invalid strings we do not have to worry about censoring moves.

Since S is symmetric and generates S_{2n+1} the resulting Markov chain is ergodic, and indeed the mixing time of such a Markov chain is at most the mixing time of the corresponding Markov chain on S_{2n+1}. In particular, the aforementioned results for uncensored transpositions on S_n immediately give $O(n^3 \log n)$ and $O(n \log n)$ upper bounds for the mixing times of the adjacent and random transposition walks, respectively, on cyclic strings. (It is not obvious how to show that the projection onto equivalence classes does not speed up mixing.)

Of course, given a bijection between two Catalan realizations (such as the one mentioned above between cyclic strings and Catalan strings) one can think of any Markov chain on one realization as a Markov chain on the other. Indeed, the adjacent (random) transposition chain on cyclic strings includes all of the transitions

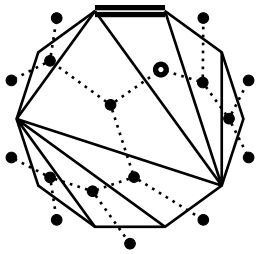


Figure 4: A triangulation and its dual binary plane tree. The double line indicates the base of the polygon and the unfilled vertex indicates the root of the tree.

which are available in the adjacent (random) transposition chain on Catalan strings. Unfortunately, the chains on cyclic strings also include many more transitions which are very unnatural for Catalan strings; these would be censored by the Catalan string walks but are instead "fixed" by a rotation of sorts which may dramatically rearrange the string. Despite the marked similarities between these Markov chains, the inclusion of these extra moves makes comparison arguments seem infeasible for obtaining meaningful bounds on the mixing times of Markov chains on Catalan strings.

2.2.3 Triangulations and binary plane trees

Definition 2.15. A *triangulation* of order n is a collection of $n-1$ pairwise non-crossing diagonals of a convex $(n+2)$-gon.

We will think of the vertices of the $(n+2)$-gon as being labeled by $[0, n+1]$, with the edge $(0, n+1)$ as the "base" of the polygon. Any triangulation of order n gives a partition of the $(n+2)$-gon into exactly n triangles.

Definition 2.16. A *rooted binary plane tree* is a rooted tree in which each non-leaf vertex has two children: a left child and a right child. The tree is said to have order n if it has n internal (i.e., non-leaf) nodes.

There is a natural bijection between triangulations and binary plane trees: the dual tree of a given triangulation is a binary plane tree (taking the root of the tree to be the base of the polygon). See Figure 4.

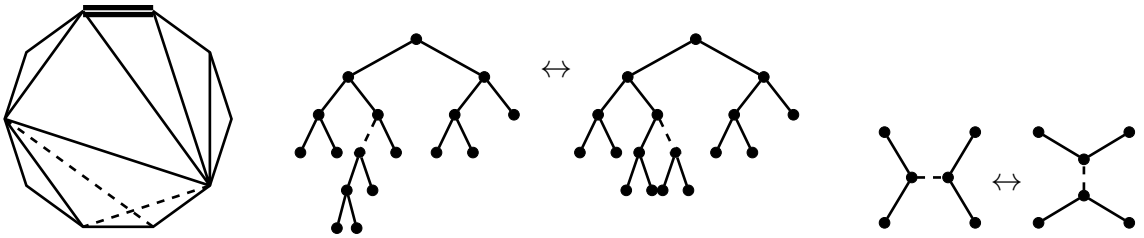


Figure 5: (a) a diagonal flip and (b) the corresponding node rotation. (c) If we forget about the rootedness of the tree, a node rotation simply reconnects a small subsection of the tree.

Definition 2.17. From an initial triangulation T, a step of the (lazy) *diagonal flip walk* is as follows:

- Pick a uniform random diagonal and remove it, leaving a quadrilateral.
- Split the quadrilateral by adding back one of its two diagonals, with equal probability, and let this new triangulation be the next state.

Note that half the time the diagonal that is added back in is the original one, so the walk is indeed lazy. Unlike the transposition walks on Catalan strings, this walk does not require any censoring: all $n - 1$ moves are always available and valid. If translated to binary plane trees along the bijection above, a transition in the diagonal flip walk corresponds to a *node rotation*, a local rearrangement of a binary tree which is well-attested in the computer science literature (for example in the context of self-balancing trees). Furthermore, an analog of the diagonal flip walk yields a possible tool for sampling and studying triangulations of a *nonconvex* polygon, which are of much interest in computational geometry [10, 12] (although the nonconvex case is even more complicated since it does require censoring; it is not even known when the nonconvex walk mixes in polynomial time). The transition graph for the diagonal flip walk is the Hasse diagram of the Tamari poset lattice (a sublattice of the Stanley lattice), or equivalently the 1-skeleton of the associahedron.

A standard bijection between binary plane trees and Catalan strings constructs a

string from a given tree by visiting each node of the tree in depth-first search order and recording whether it is a left child (U) or a right child (D). Using this bijection, the lazy diagonal flip or node rotation walk corresponds to the following Markov chain on Catalan strings:

- From initial state x, pick a uniform random non-initial U entry of x (say, at index i). Let $\sigma_{i,j} = (i\ i{+}1\ \ldots\ j) \in S_{2n}$ denote the "insertion" move which removes element $j > i$ and reinserts it before element i.

- If $x_{i-1} = \mathsf{D}$, let j be the first index of the shortest balanced substring ending with step $i-1$, and let $x' = \sigma_{j,i}(x)$ (that is, remove the U at index i and reinsert it before the substring).

- If $x_{i-1} = \mathsf{U}$, let j be the last index of the shortest balanced substring starting with step i. Let $x' = \sigma_{i-1,j}^{-1}(x)$ (that is, remove the U at index $i-1$ and reinsert it after the substring).

- The next state is either x or x' with equal probability.

A slightly different description of this correspondence is given in [11].

To see that the index j is always well-defined, it is useful to think of the elements as matched parentheses, in which case j is just the index of the parenthesis matching either x_{i-1} or x_i. Furthermore, since we can think of the move as removing and reinserting a balanced substring, the string x' is always balanced and so there is no need for censoring. Note that while the transitions of the diagonal flip walk (as realized on Catalan strings) can be expressed as permutations of the string, this walk is *not* in the class we have mentioned before of walks given by generators of S_{2n}, since in this case the permutations (insertions) which are available depend on the current state. In particular, we can only pick an index containing a U to remove, and it can only be reinserted in one or two particular locations which depend on the current string.

In 1988 Sleator et al. [67] showed that the diameter of the transition graph of the diagonal flip walk on triangulations of order n is $2n - 6$ (for large enough n), which immediately gives an $\Omega(n)$ lower bound on the mixing time for this walk. The best known lower bound on the mixing time is $\Omega(n^{3/2})$, due to Molloy et al. [56]. In the same work they also gave a very large polynomial upper bound, which was quickly improved by McShine and Tetali [55] to an $\Omega(1/n^4)$ bound on the spectral gap and $O(n^5)$ for the mixing time, using comparison to adjacent transpositions via a canonical paths argument and, implicitly, the equivalent walk bijected to Catalan strings. Molloy et al. also mention empirical evidence for a conjecture of Aldous that their lower bound is close to the truth:

Conjecture 2.18 ([2])**.** *The mixing time of the diagonal flip walk is* $O(n^{3/2} \log n)$.

One would think that it would be yield better results to compare to a Markov chain with more transitions and a better mixing time, as opposed to the adjacent transposition walk which has few moves and a long mixing time. However, to date our efforts to improve this bound by using a comparison with random transpositions instead have failed, whether using the known bounds for random transpositions described in Chapter 3 or even the conjectured $O(n \log n)$.

2.2.4 Noncrossing matchings and noncrossing partitions

Definition 2.19. A *noncrossing matching* (or *noncrossing chord diagram*) of order n is a perfect matching M on the points $[2n]$ with no *crossings*, i.e., pairs (chords) $\{i, j\}, \{i', j'\} \in M$ such that $i < i' < j < j'$.

Note that each chord $\{i, j\}$ in a noncrossing matching has endpoints of opposite parities, since the points in the interval $[i + 1, j - 1]$ must have a perfect matching (they cannot be matched to anything outside of the interval).

Definition 2.20. A *noncrossing partition* of order n is a partition $\mathcal{P}$ of $[n]$ with no

crossings, i.e., $i < i' < j < j'$ such that $i, j \in P \in \mathcal{P}$ are both in one block of $\mathcal{P}$ and $i', j' \in P' \neq P$ are both in a different block.

There is a natural bijection between noncrossing matchings and noncrossing partitions given by treating the pairs of the matching as boundaries of the parts (see Figure 1(c).)

A natural Markov chain on noncrossing matchings is as follows:

Definition 2.21. From an initial noncrossing matching M, a step of the (lazy) *chord swap walk* on noncrossing matchings is as follows:

- Pick uniform random *even* indices $i, i' \in [2n]$ and let j, j' be their (odd) matches (i.e., $\{i, j\}, \{i', j'\} \in M$).
- Let $M' = M \cup \{\{i, j'\}, \{i', j\}\} \setminus \{\{i, j\}, \{i', j'\}\}$ be M with the endpoints of the two chords rematched (respecting parity).
- If M' is noncrossing, the next state is either M or M' with equal probability. Otherwise, if M' has crossings, the next state is M.

Once again, the Markov chain is censored to maintain the noncrossing property. There is a similar Markov chain on noncrossing partitions:

Definition 2.22. From an initial noncrossing partition $\mathcal{P}$, a step of the (lazy) *merge/split walk* on noncrossing partitions is as follows:

- Pick two uniform random elements $i, i' \in \{1, \ldots, n\}$, and let $P, P' \in \mathcal{P}$ be the blocks containing them. (Without loss of generality, assume $i \leq i'$.)
- If $P = P'$, let $Q_1 = P \setminus [i+1, i']$ and let $Q_2 = P \cap [i+1, i']$. Let $\mathcal{P}' = \mathcal{P} \cup \{Q_1, Q_2\} \setminus \{P\}$ be the noncrossing partition obtained by *splitting* P into two parts determined by i and i'.

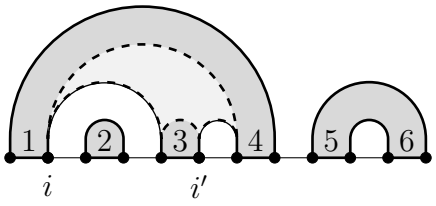


Figure 6: A move of the chord swap and merge/split chains on noncrossing matchings and noncrossing partitions.

- If $P \neq P'$, let $\mathcal{P}' = \mathcal{P} \cup \{P \cup P'\} \setminus \{P, P'\}$ be the partition obtained by *merging* parts P and P'.

- If $\mathcal{P}'$ is noncrossing, the next state is $\mathcal{P}$ or $\mathcal{P}'$ with equal probability. Otherwise, if $\mathcal{P}'$ has crossings, the next state is $\mathcal{P}$.

Identifying noncrossing matchings and partitions via the bijection above, these two Markov chains have the same transition graph, but with slightly different transition probabilities. Indeed, if we restrict the *merge* step to only occur in the case where i is the largest element in P which is less than every element of P' and i' is the largest element of P', then the transition in the merge/split chain obtained by picking elements i, i' corresponds precisely to the transition in the chord swap chain obtained by picking indices $2i, 2i'$.

Possible applications of this walk lie, for example, in the simulation of DNA folding, where such matchings can represent bonds between complementary base pairs [40]. Analysis of the chord-swap chain could also shed light on enumeration of meanders [41].

Like several of the walks already mentioned, the transition graph of the merge/split walk is also the Hasse diagram of a lattice, this time the Kreweras lattice of noncrossing partitions ordered by refinement (which is a sublattice of the Tamari lattice [47]).

One correspondence between chord diagrams and Catalan strings considers each

left endpoint of a chord as an opening bracket and each right endpoint as a closing bracket. In this way the chord swap walk corresponds to the following walk on Catalan strings (from initial state x):

- Pick a uniform random even index i and a uniform random odd index $j \in [2n]$, and let $x' = \tau_{i,j}(x)$.
- If the substring between indices i and j is balanced and x' is valid, the next state is x'.

Note that this walk is not explicitly lazy, but its transition graph is the same as that of the chord swap walk, and each (allowable) x' in this walk occurs with probability $1/n^2$, whereas each allowable M' in the chord swap walk occurs with probability $2/n^2$ (since the two chords can be chosen in either order).

The possible transitions in the chord swap walk are a subset of the transitions in the random transposition walk, and contain all of the adjacent transpositions. Therefore the argument in Proposition 2.11 applies to this case as well, and so it is easy to see that the same $O(n^4 \log n)$ upper bound on the relaxation time holds for chord swaps, and similarly that the relaxation time for chord swaps is at least that of random transpositions. Weconjecture that the actual mixing time is somewhere in between these two bounds.

Conjecture 2.23. *The mixing time of the chord swap walk is* $\Theta(n^2 \log n)$.

Bernardi & Bonichon [11] give a different bijection between noncrossing partitions and Catalan strings (that is, a bijection which is not the composition of the two bijections above): one can also construct a noncrossing partition from a Catalan string by labeling the pairs of brackets in the order they are opened and partitioning $[n]$ into sets corresponding to sequences of consecutive closing brackets. By this bijection, a move of the merge/split walk corresponds instead to swapping a sequence of consecutive down-steps with a Catalan substring directly either preceding or following

it.

2.2.5 Pattern-avoiding permutations

The final Catalan realization we will touch upon is that of pattern-avoiding permutations.

Definition 2.24. A *312-avoiding permutation* (PAP) of order n is a permutation $\sigma \in S_n$ with no $i < j < k$ such that $\sigma(i) > \sigma(k) > \sigma(j)$ (i.e., avoiding the pattern 312).

A natural bijection between Catalan strings and 312-avoiding permutations is as follows: Given a Catalan string (thought of as matched pairs of brackets), label the pairs of brackets in the order that they are opened. A corresponding 312-avoiding permutation is given by the order in which the brackets are closed. While we have been writing elements of the permutation group S_n in cycle notation, we will usually think of a 312-avoiding permutation x as the string $(x(1), x(2), \ldots, x(n))$.

As with the other string-like Catalan realizations, we can devise natural Markov chains from generators of the permutation group. Here we will consider random and adjacent transpositions, but other generators (such as insertions) may also be interesting.

If we use arbitrary transpositions on 312-avoiding permutations, the resulting moves correspond precisely to the Catalan string transpositions $\tau_{i,j}$ with $i < j$ such that the substring strictly between indices i and j is of the form $\mathsf{U}^k\mathsf{D}^k$ for some k. That is, the allowable transitions of the PAP random transposition walk are a subset of those in the chord swap walk, and are a superset of those in the adjacent transposition walk on Catalan strings. Since there are only $O(n)$ available moves from any given state and once again the adjacent transpositions are all included, the relaxation time of the PAP random transposition walk is polynomial and at most $O(n^3 \log n)$.

Each adjacent transposition on a 312-avoiding permutation corresponds to an adjacent transposition on Catalan strings, but adjacent transpositions on Catalan strings may correspond to nonadjacent transpositions on 312-avoiding permutations.

Despite the heavily restricted moves, this walk still has the potential to be interesting. For now we will just show that it is ergodic.

Proposition 2.25. *Adjacent transpositions connect the space of 312-avoiding permutations.*

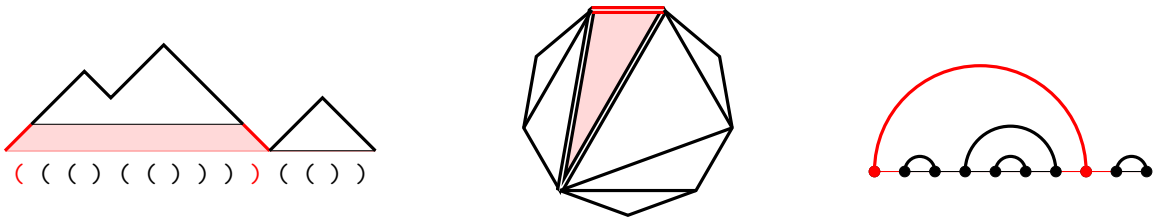
Proof. We will give a path from each PAP x to the identity permutation, by induction on n. This is trivial for $n \leq 2$. For larger n, let i be the index containing entry 1 and let $a = (x_1, \ldots, x_{i-1})$ and $b = (x_{i+1}, \ldots, x_n)$. Every element of a is less than every element of b, and a and b are themselves 312-avoiding, so a is a 312-avoiding permutation of $[2, k]$ and b is a 312-avoiding permutation of $[k+1, n]$. Inductively use adjacent transpositions to sort a and b, yielding $(2, 3, \ldots, k, 1, k+1, \ldots, n)$. Then use adjacent transpositions to move the element 1 to the left into its proper position. □

2.2.6 Recursive decomposition of Catalan structures

We should mention one aesthetically appealing approach to bounding mixing times which could be generally applicable to nearly any Catalan structure. It is well-known that for $n \geq 1$ the Catalan numbers satisfy the recurrence

$$C_n = \sum_{i=1}^{n} C_{i-1} C_{n-i},$$

which is evidenced in almost every Catalan realization $\mathcal{C}_n$ by a natural bijection $\mathcal{C}_n \to \bigcup_{i=1}^{n} (\mathcal{C}_{i-1} \times \mathcal{C}_{n-i})$ (for $n \geq 1$) or in other words a canonical decomposition of each structure of order n into two structures of total order $n-1$ (with an extra element used as "glue"). For example, a nontrivial binary plane tree consists of a left subtree and a right subtree (either or both of which may be trivial), held together by a single extra internal vertex (the root). Similarly, any Catalan string x can be



(a) Catalan string and Dyck path (b) Triangulation (c) Chord diagram

Figure 7: Recursive decompositions of some Catalan structures. The "glue" is colored red.

written uniquely as the concatenation $\mathsf{U}a\mathsf{D}b$, where a and b are again (possibly empty) Catalan strings; a triangulation can be decomposed into triangulations of two smaller polygons by removing the triangle containing its base.

Natural Markov chains on Catalan structures tend to respect this recursive decomposition, in that a valid transition from $(x, y) \in \mathcal{C}_{i-1} \times C_{n-i}$ consists of either (a) making any valid move on either x or y or (b) moving to some element of $\mathcal{C}_{j-1} \times C_{n-j}$ for $j \neq i$.

It is common to exploit recursive structure in bounding mixing times, and indeed both the triangulation bound of [56] and the balanced matroid bounds of [29, 43, 44] (which we will use in Chapter 3) exploit different recursive structures of those objects. A more general Markov chain decomposition method is given in [52].

			t_{mix} bounds		
Walk	Max deg.	Diameter	Lower	Upper	Conjecture
Catalan adj. trans.	$2n-3$	$\frac{n(n-1)}{2}$	$\frac{2n^3 \log n}{\pi^2}$	$\frac{n^3 \log n}{\pi^2}$	$\frac{n^3 \log n}{\pi^2}$
Catalan rand. trans.	$(n-1)^2$	$n-1$	$\Omega(n)$	$O(n^2 \log n)$	$\Theta(n \log n)$
Diagonal flip	$n-1$	$2n-6$	$\Omega(n^{3/2})$	$O(n^5)$	$\Theta(n^{3/2} \log n)$
Chord swap	$\frac{n(n-1)}{2}$	$n-1$	$\Omega(n)$	$O(n^4 \log n)$	$\Theta(n^2 \log n)$
PAP adj. trans.	$n-1$	$(n-1)^2$	$\Omega(n^2)$	?	?
PAP rand. trans.	$3(n-2)$	$\lceil \frac{3(n-2)}{2} \rceil$	$\Omega(n)$	$O(n^3 \log n)$	?

Table 1: Summary of Catalan walks with known and conjectured bounds.

CHAPTER III

MIXING FOR CATALAN RANDOM TRANSPOSITIONS

In this chapter, we describe progress towards pinning down the mixing time of the random transposition walk on Catalan strings. Our main result is an $O(n^2 \log n)$ upper bound on the mixing time of the random transposition walk on Catalan strings of order n. The proof relies on a rephrasing of the walk as a basis exchange walk on a balanced matroid. In Section 3.1 we describe the Catalan matroid (due to Ardila [6]) and the prior work [29, 44, 43] regarding mixing times for the basis exchange walk on balanced matroids. All that remains to bound the mixing time of our Dyck path walk is the proof in Section 3.2 that the Catalan matroid is balanced.

Although our focus is the transposition walk on Dyck paths, our analysis will actually extend to a broader class of walks. A *lattice path* of length m is a string $P \in \{1, -1\}^m$ (as before, we will frequently write U and D for up and down in place of 1 and -1). The *height* of P at index i is $h_P(i) = \sum_{j=1}^{i} x_i$, and we can draw P on the grid as the graph of h_P. Lattice paths include Dyck paths, and of course a Dyck path of order n is just a lattice path P of length $2n$ with $h_P \geq 0$ and $h_P(2n) = 0$. We will refer to a lattice path P of length m as a lattice path *from* $(0, 0)$ *to* $(m, h_P(m))$.

From this we can define a partial order on the set of lattice paths from $(0, 0)$ to $(m, 2r - m)$ by letting $P \leq Q$ whenever $h_P \leq h_Q$, pointwise. Note that $P \leq Q$ if and only if $q_i \leq p_i$ for all $i \in [r]$, where p_i (resp. q_i) is the index of the ith up-step in P (resp. Q). This partial order is in fact a lattice[1] with maximal element $\mathsf{U}^r\mathsf{D}^{m-r}$ and minimal element $\mathsf{D}^{m-r}\mathsf{U}^r$. This also gives another characterization of Dyck paths as lattice paths P from $(0, 0)$ to $(2n, 0)$ with $P \geq (\mathsf{UD})^n$.

[1] Apologies for the overlap in terminology.

We have already mentioned the random transposition walk on lattice paths from $(0,0)$ to $(2n,0)$, and that the generalization of this walk to lattice paths from $(0,0)$ to (m,h) is equivalent to the Bernoulli-Laplace model on $\binom{[m]}{r}$, where $r=(h+m)/2$ is the number of up-steps in every such path, which has been studied extensively [27]. The adjacent transposition walk on all lattice paths has been thoroughly analyzed in [71]. The analysis of the upper bound for adjacent transpositions applies just as well to the case where the walk is censored to remain in the poset interval $[P,Q]$ for some pair of lattice paths $P<Q$. Our goal in this chapter is to give corresponding bounds for the random transposition walk censored to $[P,Q]$.

3.1 Lattice path matroids

3.1.1 Matroids and the basis exchange walk

Recall that a nonempty set $\mathcal{B}\subseteq 2^U$ is the set of *bases* of a *matroid* $M=(U,\mathcal{B})$ if the following *basis exchange axiom* holds:

> **Matroid Basis Exchange Axiom.** For any bases $A,B\in\mathcal{B}$ and every $e\in A\setminus B$ there exists $f\in B\setminus A$ such that $A\setminus\{e\}\cup\{f\}\in\mathcal{B}$ is a basis.

Among other things, this axiom guarantees that all bases have the same cardinality, which is the *rank* of M.

We will also make use of two dual operations on matroids: contraction and deletion.

Definition 3.1. For a matroid $M=(U,\mathcal{B})$ and an element $e\in M$, the matroid M *contract* e is $M_e=(U,\mathcal{B}_e)$, where $\mathcal{B}_e=\{B\in\mathcal{B}:e\in B\}$. Similarly, M *delete* e is $M^e=(U,\mathcal{B}^e)$, where $B^e=\{B\in\mathcal{B}:e\notin B\}$.[2]

[2]Note that this definition differs slightly from the usual one, in which the element e being contracted or deleted is removed from the ground set. In our case it will be convenient to leave e in place, so as to more easily identify $\mathcal{B}_e\subseteq\mathcal{B}$ and preserve identities such as $\mathcal{B}=\mathcal{B}_e\cup\mathcal{B}^e$ (with $\mathcal{B}_e\cap\mathcal{B}^e=\emptyset$).

The results of contraction and deletion are again matroids, and a *minor* of M is any matroid which can be obtained from M through a series of contractions and deletions.

The order in which contractions and deletions are performed does not matter, so we will write M_I^J for the matroid obtained from M by contracting the elements in I and deleting the elements in J, and $\mathcal{B}_I^J$ will denote the set of bases of M_I^J.

Given a matroid M, Feder and Mihail [29] study the following *basis exchange walk* on the state space $\mathcal{B}$ of bases:

From state $B \in \mathcal{B}$,

- Pick uniform and independent random elements $a \in U$ and $b \in B$, and let $B' = B \cup \{a\} \setminus \{b\}$.
- If $B' \in \mathcal{B}$ then the next state is either B or B' with equal probability, otherwise the next state is B.

The matroid basis exchange axiom guarantees that the transitions connect the state space, and since the walk is symmetric and lazy its stationary distribution is uniform. It is tentatively conjectured that the basis exchange walk is fast for any matroid (i.e., the mixing time is bounded by some polynomial in $m = |U|$), but there is little evidence in favor of this.

On the other hand, [29] introduced the notion of *balanced* matroids, which capture the notion that for a randomly chosen basis (of the matroid or any of its minors), conditioning on the occurrence of one element (in the basis) can only make the occurrence of any other less probable. They show that for the case of balanced matroids the walk is indeed rapidly mixing, by using decomposition techniques to bound its spectral gap.

Definition 3.2. A matroid $M = (U, \mathcal{B})$ is *negatively correlated* if for every pair of

distinct elements $e, f \in U$

$$\frac{|\mathcal{B}_e|}{|\mathcal{B}|} \geq \frac{|\mathcal{B}_{ef}|}{|\mathcal{B}_f|}. \tag{1}$$

The matroid M is *balanced* if M and all of its minors are negatively correlated.

Negative correlation is equivalent to the very natural condition that for a uniform random basis $B \in \mathcal{B}$, $\Pr[e \in B] \geq \Pr[e \in B | f \in B]$.

Indeed, many common classes of matroids are balanced, including uniform matroids (whose bases are all size-r subsets of the ground set), graphic matroids (with the ground set being the edges of a connected graph and the bases being spanning trees of the graph), matroids of rank ≤ 3, and regular matroids (matroids which can be represented over every field) [29, 19]. We will rely on the following extension of Feder and Mihail's result for balanced matroids, due to Jerrum and Son.

Theorem 3.3 ([43])**.** *The spectral gap γ and log-Sobolev constant α for the basis exchange walk on a balanced matroid M of rank r on a ground set of size m are lower bounded by*

$$\gamma \geq \frac{2}{mr} \qquad \textit{and} \qquad \alpha \geq \frac{1}{2mr}.$$

In particular, by Theorem 2.4 the bound on the log-Sobolev constant implies that the mixing time of the basis exchange walk on a balanced matroid is at most $O(mr \log\log|\mathcal{B}|)$.

3.1.2 The Catalan matroid and other lattice path matroids

To aid our setting of the Catalan transposition walk in terms of matroids, we refer to the observation (made independently by Ardila [6] and Bonin and De Mier [14]) that the set of Dyck paths can be thought of as bases of a matroid.

Definition 3.4. The *Catalan matroid* of order n is $\mathcal{C}(n) = ([2n], \mathcal{B}(n))$, where the elements of $\mathcal{B}(n)$ are the index-sets of up-steps in Dyck paths of length $2n$.

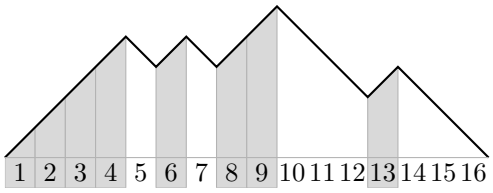


Figure 8: The Dyck path (UUUUDUDUUDDDUDDD) above corresponds to the basis $\{1, 2, 3, 4, 6, 8, 9, 13\}$ of the Catalan matroid of order 8.

The Catalan matroid is actually the transversal matroid for the set system $\mathcal{S} = \{[1], [3], \ldots, [2n-1]\}$ (that is, its bases are the systems of distinct representatives of $\mathcal{S}$). In short, the representative of the set $[2i-1]$ will be the index of the ith up-step in the corresponding Dyck path (although this assignment of distinct representatives need not be unique). Indeed,

Observation 3.5. *The basis exchange walk on the Catalan matroid is exactly the random transposition walk on Catalan strings.*

In other words, given Theorem 3.3, to obtain a mixing time bound for the Dyck transposition walk it suffices to show that the Catalan matroids are balanced.

To this end, it would be convenient if the Catalan matroids belonged to some known class of balanced matroids. The most general class of matroids known to be balanced is *regular matroids*, but Ardila shows that $\mathcal{C}(n)$ is not representable over any $\mathbb{F}_q$ for $q \leq n-2$ and thus is not regular. Transversal matroids also need not be balanced in general: Choe and Wagner [19] give a transversal matroid of rank 4 which is not balanced.

Thus the bulk of our work here will be to show that the Catalan matroid is balanced, and from this the main mixing result will follow immediately. For our later discussion it will help to work with a minor-closed class of matroids (which the Catalan matroids certainly are not). Bonin and de Mier ([14]) discuss the following class of *lattice path matroids*, which generalize the Catalan matroids by allowing any pair of bounding paths.

Definition 3.6. For two lattice paths $A \leq B$ from $(0,0)$ to $(m, 2r-m)$, consider the set $[A,B]$ of lattice paths P from $(0,0)$ to $(m,2r-m)$ with $A \leq P \leq B$. The *lattice path matroid* $\mathcal{L}[A,B]$ (which is of rank r on ground set $[m]$) has as its bases the index sets of up-steps of paths in $[A,B]$.

Although it is not immediately obvious that $\mathcal{L}[A,B]$ is a matroid, [14] observes that in fact the lattice path matroid $\mathcal{L}[A,B]$ is the transversal matroid of the set system $\{[a_1,b_1],[a_2,b_2],\ldots,[a_r,b_r]\}$, where a_i (resp. b_i) is the index of the ith up-step in A (resp. B). Indeed, it is shown in that paper that the transversal matroids of set systems $\{[a_1,b_1],[a_2,b_2],\ldots,[a_r,b_r]\}$ with $a_1 \leq a_2 \leq \cdots \leq a_r$ and $b_1 \leq b_2 \leq \cdots \leq b_r$ are precisely the lattice path matroids[3]. The class of basis exchange walks on lattice path matroids also includes the (unconstrained) Bernoulli-Laplace model as the basis exchange walk on a uniform matroid.

In addition to showing this correspondence, Bonin and de Mier show that the class of lattice path matroids is closed under taking minors and duals. They also define a smaller (in fact, minimal) minor-closed class of matroids containing $\mathcal{C}_n$, which they call *generalized Catalan matroids*, consisting of the lattice path matroids $\mathcal{L}[A,B]$ where $A = (\mathsf{U})^r(\mathsf{D})^{m-r}$ is the maximal path among all paths from $(0,0)$ to $(m,2r-m)$. Indeed, Sohoni [68] has actually already shown the balanced property of generalized Catalan matroids, although he calls them *Schubert matroids* and does not link his result to rapid mixing of the Catalan random transposition chain (although he does mention the Catalan adjacent transposition chain). Our proof applies to the more general class of lattice path matroids.

Since the class of lattice path matroids is minor-closed, for our main result it suffices to show that lattice path matroids are negatively correlated. However, it is worth noting that the same method can also be used to show negative correlation

[3]The set system of this form corresponding to a given lattice path matroid may not be unique; for example, the Catalan matroid is the transversal matroid for both $\{[1],[3],\ldots,[2n-1]\}$ and $\{[1],[2,3],\ldots,[n,2n-1]\}$.

directly for any minor of a lattice path matroid, by restricting certain indices to always have up- or down-steps as necessary.

3.2 Mixing bound for random transpositions

To show negative correlation for lattice path matroids, it will be convenient to use an equivalent formulation of (1), which is easily obtained by repeatedly applying the identity $|\mathcal{B}| = |\mathcal{B}_e| + |\mathcal{B}^e|$:

$$|\mathcal{B}_{ef}||\mathcal{B}^{ef}| \leq |\mathcal{B}_e^f||\mathcal{B}_f^e|\,. \tag{2}$$

Now we are ready to prove

Theorem 3.7. *For every pair of lattice paths $A \leq B$ from $(0,0)$ to $(m, 2r-m)$ the lattice path matroid $\mathcal{L}[A,B]$ is negatively correlated.*

Proof. Let $\mathcal{L}[A,B] = ([m], \mathcal{B})$ be any lattice path matroid.

To prove inequality (2) for every pair $e < f \in [m]$, we will construct an injective map

$$\varphi_{ef} : \mathcal{B}_{ef} \times \mathcal{B}^{ef} \to \mathcal{B}_e^f \times \mathcal{B}_f^e.$$

Note that we can associate members of these sets with lattice paths in $\mathcal{L}[A,B]$ by

- $\mathcal{B}_{ef}$: paths with up-steps at indices e, f,
- $\mathcal{B}^{ef}$: paths with down-steps at indices e, f,
- $\mathcal{B}_e^f$: paths with an up-step at index e and a down-step at index f, and
- $\mathcal{B}_f^e$: paths with a down-step at index e and an up-step at index f.

Let $P \in \mathcal{B}_{ef}$ and $Q \in \mathcal{B}^{ef}$ be lattice paths and consider the following cases. (The figures shown are for the Catalan matroid.)

Case 1. Suppose the paths P, Q intersect (without necessarily crossing) in the region (I) between e and f. Note that this includes all cases where the path P is below Q at e and above Q at f or vice versa. Take the first such intersection point x in (I) and switch the paths P, Q after x to obtain new paths $P' \in \mathcal{B}_e^f, Q' \in \mathcal{B}_f^e$ as shown in Figure 9. Set $\varphi_{ef}(P, Q) = (P', Q')$.

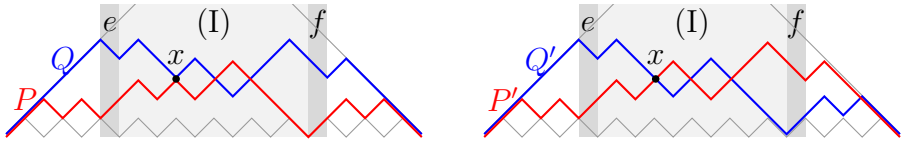


Figure 9: The injection for Case 1.

Case 2. Suppose the paths do not meet in region (I) and consider the paths P, Q in the region (II) after position f. Imagine translating the fragment of path Q after f to P so that their initial points in f coincide. If the imaginary fragment intersects P, let x be the first such point of intersection. Construct new paths P', Q' by swapping the segments of P and Q between f and x (including f itself). See Figure 10. If the new paths P', Q' are both between A and B (so that $P' \in \mathcal{B}_e^f$ and $Q' \in B_f^e$), then we set $\varphi_{ef}(P, Q) = (P', Q')$.

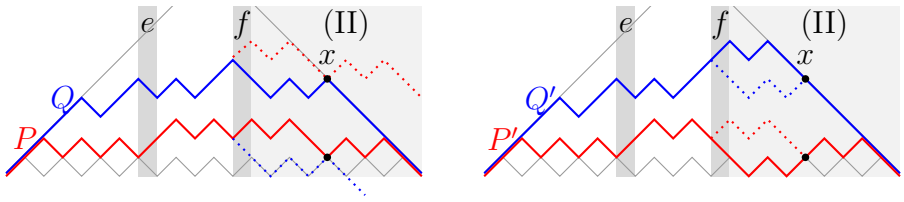


Figure 10: The injection for Case 2.

Case 3. Otherwise, perform the mirror image of the operation from Case 2 in the part (III) up to position e, as shown in Figure 11 and again set $\varphi_{ef}(P, Q) = (Q', P')$.

Case 1 covers all pairs of paths P, Q such that P is above Q at e and below it at f or vice versa. Case 2 covers all remaining pairs of paths with P above Q at f, since in this case the fragment of Q must intersect P and both resulting paths (P', Q') are

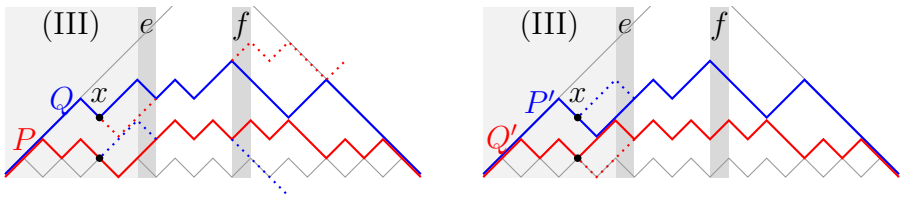


Figure 11: The injection for Case 3. Case 2 fails because the translated segment of Q leaves the valid region before intersecting Q.

between P and Q. Finally, Case 3 covers all still remaining pairs of paths with P below Q at e. Hence every pair of paths P, Q is covered by one of these cases.

Finally we must argue that φ_{ef} is injective. Given P', Q' we must be able to determine which case above was applied to produce them, and from this it is simple to recover P and Q. Since we know e, f it is easy to identify the regions (I), (II) and (III) in the three cases. If P', Q' intersect in region (I) then they must have come from Case 1, as neither of the later cases can produce such an intersection. If they do not, first try to apply the inverse operation for Case 2, which is the same as the forwards operation; if this was not possible for the starting paths P, Q then it is not possible for P', Q' either (as performing the transformation of Case 3 cannot cause the transformation of Case 2 to become valid if it was not already), so if we have a pair of paths P', Q' for which the second move is possible and does result in lattice paths between A and B then we know we must have arrived at it through Case 2. Otherwise, we were in Case 3. See Figure 12.

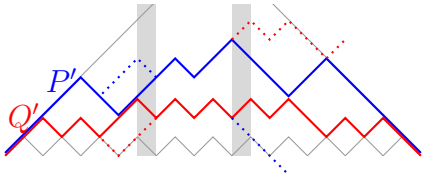


Figure 12: An example of recovering the case from the result. Since the paths do not intersect in region (I) we must have come from Case 2 or 3. Case 2 fails because the translated segment of P' leaves the valid region before rejoining Q', so the paths must have come from Case 3. Note that the transformation ϕ_{ef} is usually not onto.

□

Remark 3.8. Since Case 1 can be applied whenever P is above Q at e and below at f or vice versa, Case 2 can be applied whenever P is above Q at both e and f, and Case 3 can be applied whenever P is below Q at both e and f, it is tempting to try to divide the cases more naturally according to the relative heights of P and Q at e and f. However, note that in Cases 2 and 3 (and sometimes Case 1) the resulting paths P', Q' always have P' above Q' at both e and f, and this more natural division of cases does not allow us to recover uniquely which transformation was applied.

CHAPTER IV

CENSORED RANDOM WALKS

4.1 Monotone censoring

Most of the Markov chains we have mentioned on Catalan structures take transitions inspired by previously studied random walks and implement an extra *censoring* step to ensure that the walk remains within the Catalan space. That is, they are censored versions of Markov chains which tend to be more amenable to analysis. In this chapter, we give a general framework for censored Markov chains and discuss conditions under which mixing properties of a censored chain can be derived from the mixing properties of the original.

Definition 4.1. Let P be the transition matrix of a Markov chain on state space Ω and let $A \subset \Omega$. The Markov chain *censored to* A is the following Markov chain on state space A:

- From state $x \in A$, draw $x' \in \Omega$ from distribution $P(x, \cdot)$.
- If $x' \in A$, the next state is x'. Otherwise the next state is x.

The two processes coincide unless the original process tries to leave the set A, in which case the move is censored and the new chain remains where it is. This model is very general and also very common, encompassing the usual Markov chains for sampling independent sets and matchings inspired by statistical physics, as well as a broad array of other structures [39, 36, 58]. Relevant to this work, the Catalan random transposition chain is just the Bernoulli-Laplace process censored to Catalan strings. Similarly, the chord swap walk can be though of as a censoring of a walk on all perfect matchings in the complete bipartite graph $K_{n,n}$ (equivalent to random

transpositions in S_n). The merge/split walk can be thought of as a censoring of a merge/split walk on all partitions of $[n]$, and the walks we described on 312-avoiding permutations are censored versions of well-studied walks on the symmetric group S_n.

Of course, in order to get any fast mixing guarantees for a censored walk, there must be some sort of requirements on A, since *a priori* it is possible that the transition graph restricted to A is not even connected. The overarching question we explore in this chapter is:

Question 4.2. *What are sufficient conditions to be able to bound the mixing time of a censored Markov chain in terms of the original, uncensored chain?*

Ding and Mossel [28] give one example of such a bound for walks on the hypercube in the case that the chain is restricted to a *monotone*[1] subset of the state space.

Theorem 4.3 ([28])**.** *For any monotone set $A \subseteq \{\pm 1\}^n$, the Markov chain on Q_n censored to A satisfies*

$$\Phi \geq \frac{\pi(A)}{16n},$$

and so

$$t_{mix} \leq O\left(\left(\frac{n}{\pi(A)}\right)^2 \log|A|\right).$$

As we will show, their result extends to monotone subsets of any distributive lattice, provided that a certain *monotonicity testing* property of the uncensored chain holds. In short, the property says that it should be possible (with high probability) to test monotonicity of any subset of the state space by observing the endpoints of randomly chosen transitions of the Markov chain.

While we will also show that the result of [28] is asymptotically tight for the conductance of monotone subsets of the cube, it does not seem to give the correct mixing time bound, which we conjecture to be $O(n \log n/\mu(A))$.

[1]By convention, we will take *monotone* to mean *monotone increasing*, i.e., if $x < y$ then $x \in A \implies y \in A$.

To get a slightly improved bound, we next give a canonical paths argument which bounds the spectral gap for monotone subsets of the hypercube by $O(n^2/\mu(A))$, improving the mixing time result to $O(n^2 \log|A|/\mu(A))$. The $1/\mu(A)$ factor is tight (for both the spectral gap and the mixing time), but it seems likely that the dependence on n could be improved.

In a similar vein, Mathieu [53] gives an $\Omega(1/n)$ bound on the spectral gap and log-Sobolev constants of censored walks on the cube under the more stringent condition that the subset A is monotone and contains the middle layers: A must contain every x with at least $(\frac{1}{2} - \epsilon)n$ ones for some constant $\epsilon > 0$. The argument hinges on the fact that such a walk would spend most of its time among the middle layers of the cube even without any censoring. While this is at face value a much stronger assumption than the mere monotonicity of [28], the same argument seems to apply as well to sets of the form $A \cap B$ where A is monotone increasing and B is monotone decreasing, so long as there is a constant-sized band around the middle layers which remains uncensored.

In this chapter we will mostly discuss the case of censoring a walk on the Hasse diagram of a poset lattice to some monotone subset. The monotonicity requirement is enough to guarantee that the censored walk is connected, since it guarantees that there is a path from every point to the maximum of the lattice. Indeed, as mentioned in Chapter 2 many of the Catalan chains we are interested can be thought of as walks on the Hasse diagram of a lattice or as a censored version of such a walk (or both). It is worth noting that much of the analysis below for lattices in general also applies to walks on any subgraph of the comparability graph containing the Hasse diagram. For example, while the Catalan adjacent transposition walk is a monotone censoring of the random walk on the Hasse diagram of all lattice paths ordered by height (as in Chapter 3), the Catalan random transposition walk allows transitions also between other comparable pairs.

Unfortunately, these results are difficult to apply fruitfully to the Catalan walks we have mentioned. The first hurdle is that the relevant lattices are not distributive. Even if they were and we could prove monotonicity testing, the results would still be very weak, since in most cases (except Catalan strings) the Catalan class is only an exponentially small portion of the uncensored lattice.

4.2 Conductance bound from monotonicity testing

Let $G = (L, E)$ be a d-regular directed (acyclic) graph whose transitive closure is a distributive lattice L. We consider the random walk on *undirected* edges of G with transition probability $1/d$. For a subset $A \subseteq L$, the walk censored to A is just the random walk on the subgraph $G[A]$ induced by A with the same transition probability $1/d$. Note that the assumption of d-regularity and uniform edge probability is without much loss of generality: since we do not require the graph to be simple we can phrase any symmetric Markov chain on L with rational transition probabilities in this way by duplicating edges and adding loops.

We will give a lower bound on the *edge expansion*

$$\phi_A(S) := \frac{\partial_A(S)|A|}{d\,|S||A \setminus S|}$$

where $\partial_A(S)$ is the number of edges in $G[A]$ (or equivalently in G) with one endpoint in S and the other in $A \setminus S$. Note that since $\partial_A(S) = \partial_A(A \setminus S)$ the edge expansion is related to the conductance by

$$\Phi = \min_{\substack{S \subseteq A \\ |S| \leq |A|/2}} \frac{\partial_A(S)}{d\,|S|} \geq \min_{S \subseteq A} \phi_A(S),$$

and in fact the minimum of $\phi_A(S)$ differs from Φ by no more than a factor of 2.

Let $\oplus$ denote symmetric difference and write

$$\delta_{\mathrm{m}}(S) = \min_{S' \text{ monotone}} |S \oplus S'|$$

for the distance between S and the nearest monotone set and

$$\epsilon_{\mathrm{m}}(S) = |\{(x,y) \in E : x < y,\ x \in S,\ y \notin S\}|$$

for the number of edges of G on which S violates monotonicity, i.e., increasing edges leaving G.

Definition 4.4. The graph G is *β-testable* if for any $S \subseteq L$, the probability of detecting non-monotonicity of S by observing its values on a uniform random edge (x,y) of G satisfies

$$\Pr_{x<y\in E}[x \in S, y \notin S] = \frac{2\epsilon_{\mathrm{m}}(S)}{d|L|} \geq \beta\frac{\delta_{\mathrm{m}}(S)}{|L|}.$$

This is a fairly natural condition, and indeed Goldreich et al. show that the Boolean cube is $(2/n)$-testable [37]. We will show that this sort of testability property suffices to give the following extension of Ding and Mossel's result.

Theorem 4.5. *If G is β-testable and A is a monotone subset of L then*

$$\Phi(G_A) \geq \left(\frac{\beta}{8}\right)\frac{|A|}{|L|}.$$

Proof. Suppose that G is β-testable. That is, for every $S \subseteq L$ we have

$$\epsilon_{\mathrm{m}}(S) \geq \frac{\beta d}{2}\delta_{\mathrm{m}}(S).$$

Note that since A is monotone, for any $S \subseteq A$ we have $\epsilon_{\mathrm{m}}(S) \leq \partial_A(S)$ (as any increasing edge leaving S in G is also an edge of G_A leaving S.) In fact, the edges between S and $A \setminus S$ are precisely the increasing edges leaving S and the increasing edges leaving $A \setminus S$, so $\partial_A(S) = \epsilon_{\mathrm{m}}(S) + \epsilon_{\mathrm{m}}(A \setminus S)$.

Consider nonempty $S \subset A$, and let $T = A \setminus S$. First we claim that S and T cannot both be close to monotone increasing. In particular, there cannot exist monotone sets $S', T' \subseteq L$ such that

$$|S \oplus S'| < \frac{|S||T|}{4|L|} \qquad \text{and} \qquad |T \oplus T'| < \frac{|S||T|}{4|L|}.$$

Indeed, if such S', T' existed, then we would have

$$|S| - |S'| \leq |S \oplus S'| < \frac{|S||T|}{4|L|} \leq \frac{1}{4}|S|$$

and so $|S'| > \frac{3}{4}|S|$. Similarly for T we would have $|T'| > \frac{3}{4}|T|$. Since S' and T' are monotone subsets of a distributive lattice the well-known FKG inequality [30] gives

$$|S' \cap T'| \geq \frac{|S'||T'|}{|L|} \geq \Big(\frac{9}{16}\Big)\frac{|S||T|}{|L|}.$$

But now we have a contradiction, since

$$0 = |S \cap T| \geq |S' \cap T'| - |S \oplus S'| - |T \oplus T'| \geq \Big(\frac{9}{16} - \frac{1}{2}\Big)\frac{|A||S|}{|L|} > 0.$$

Hence it is either the case that $|S \oplus S'| \geq \frac{|S||T|}{4|L|}$ for every monotone $S' \subseteq L$ or $|T \oplus T'| \geq \frac{|S||T|}{4|L|}$ for every monotone $T' \subseteq L$. Either way, we have for every nonempty $S \subset A$

$$\begin{aligned}
\frac{\partial_A(S)|A|}{d|S||T|} &= \frac{|A|}{d|S||T|}(\epsilon_{\mathrm{m}}(S) + \epsilon_{\mathrm{m}}(T)) \\
&\geq \frac{\beta|A|}{2|S||T|}(\delta_{\mathrm{m}}(S) + \delta_{\mathrm{m}}(T)) \\
&\geq \frac{\beta|A|}{2|S||T|}\left(\frac{|S||T|}{4|L|}\right) = \frac{\beta}{8}\left(\frac{|A|}{|L|}\right),
\end{aligned}$$

so $\Phi(G_A) \geq \frac{\beta}{8}\left(\frac{|A|}{|L|}\right)$ as desired. □

4.3 Monotone censoring and spectral gap

First of all, it is worth noting that while Ding & Mossel's bound on the mixing time may have room for improvement, the result about conductance is asymptotically tight both in its dependence on n and on $\mu(A)$. On the one hand, for $A = V$ ($\mu(A) = 1$) it is known that the conductance of the n-cube is $1/n$. To see the tightness in $\mu(A)$ consider the following example.

Example 4.6. Consider the case of a sum of two discrete cubes (say, k and ℓ cubes, $k \leq \ell$, $k + \ell = n$) intersecting at one point. This structure can be embedded into Q_n

as the induced subgraph on a monotone set $A = B \cup C$, where

$$B = \{(1, \ldots, 1, x) : \ x \in \{\pm 1\}^k\} \quad \text{and} \quad C = \{(y, 1, \ldots, 1) : x \in \{\pm 1\}^\ell\}.$$

Then $\mu(A) = (2^k + 2^\ell)/2^n \geq 2^k/2^n = 2^{-\ell}$ and $\partial_A(C) = k$, so

$$\Phi(Q_n[A]) \leq \frac{\partial_A(C)}{n|C|} = \frac{k}{n2^\ell} \leq \frac{1}{2^\ell} \leq \mu(A).$$

That is, if we wish to bound the mixing time $t_{\text{mix}}(A)$ of the random walk on the cube restricted to A, then we cannot use conductance to give a bound better than quadratic in $1/\mu(A)$ and in n.

Therefore we would prefer to bound either the spectral gap or log-Sobolev constant of the censored chain, which in turn would more tightly control the mixing time. It is in fact possible to modify the proof of the previous section to give some bound on the spectral gap, but the necessary testability property is much less natural than the one we have used here and the method does not seem to yield an improved result even for the Boolean cube. Instead, we give a canonical paths argument which is specific to monotone subsets of the Boolean cube and yields a slightly improved mixing time bound over that of Ding & Mossel.

The main lemma is as follows:

Lemma 4.7. *Let $A \subseteq \{\pm 1\}^n$ be a monotone set. Then there exists a set $\Gamma = \{\Gamma_{u,v}\}_{u \neq v \in A}$ of canonical paths between (ordered) pairs of points in A, each of length $\leq n$, such that each (directed) edge of Q_n is used at most 2^{n-1} paths.*

Indeed, Lemma 4.7 allows us to apply Theorem 2.7 to compare the censored random walk to the random walk on the complete (looped) graph with $P(x, y) = 1/|\Omega|$ for all $x, y \in \Omega$, which has spectral gap and mixing time 1. As a result,

Corollary 4.8. *The random walk on the Boolean cube Q_n censored to a monotone subset A has spectral gap and mixing time satisfying*

$$\frac{1}{\gamma} \leq B = \frac{n^2}{\pi(A)} \qquad \textit{and} \qquad t_{mix} \leq O\left(\frac{n^2}{\pi(A)} \log|A|\right).$$

This is an asymptotic improvement over Ding & Mossel's result in the case where $\pi(A) = o(1)$. It is also possible to use this same set of canonical paths to give an alternate proof of Ding & Mossel's conductance result. On the other hand, the log-Sobolev constant for the walk on the complete graph is order $\log|\Omega|$, so we actually get a slightly worse mixing result using the log-Sobolev constant here.

To prove Lemma 4.7 we first prove the following.

Lemma 4.9. *Let $A \subseteq \{\pm 1\}^n$ be a monotone subset. It is possible to construct paths in $Q_n[A]$ for each (ordered) pair of antipodal points in A in such a way that each path has length exactly n and each (directed) edge belongs to at most one path.*

Proof. We proceed by induction on n. For $n = 1$ the lemma is true. Suppose that $n > 1$. Take $A_+ = \{x \in A : \ x_n = 1\}$ and $A_- = \{x \in A : \ x_n = -1\}$. Clearly A_+ and A_- are monotone on $\{\pm 1\}^{n-1}$. From the inductive hypothesis we can construct canonical paths of length $n-1$ between antipodal points in A_+ and between antipodal points in A_- such that each edge is an element of at most one path. We consider simultaneously the four points

$$x = (x_1, \ldots, x_{n-1}, 1) \qquad -x = (-x_1, \ldots, -x_{n-1}, -1)$$
$$y = (x_1, \ldots, x_{n-1}, -1) \qquad -y = (-x_1, \ldots, -x_{n-1}, 1).$$

If neither antipodal pair is in A then we have no paths to produce, so suppose without loss of generality that $x \in A_+$ and $-x \in A_-$. Note that monotonicity of A implies $-y \in A_+$. Let us consider two cases.

(a) Suppose first that $y \notin A$, so that among these points we need only construct the two paths $\gamma_{x,-x}$ and $\gamma_{-x,x}$. Indeed, we can take the paths

$$\gamma_{x,-x} = (x) \to_{A_+} (-y) \to (-x) \qquad \gamma_{-x,x} = (-x) \to (-y) \to_{A_+} (x),$$

where we have used the known paths $\gamma_{x,-y}$ and $\gamma_{-y,x}$ between antipodal points in A_+.

(b) If $y \in A_-$ then we need to define the four paths

$$\gamma_{x,-x} = (x) \to_{A_+} (-y) \to (-x) \qquad \gamma_{-x,x} = (-x) \to_{A_-} (y) \to (x)$$
$$\gamma_{y,-y} = (y) \to_{A_-} (-x) \to (-y) \qquad \gamma_{y,-y} = (-y) \to_{A_+} (x) \to (y),$$

where we have now used known paths between antipodal points in both A_+ and A_-.

Clearly all of the paths constructed are of length n. In each case each directed edge between $\pm x$ and $\pm y$ is used in at most one path, as is each of the inductively assumed paths in $A_\pm$ (and hence each of the edges within $A_\pm$). Thus this procedure gives a set of paths joining antipodal points in A such that each directed edge belongs to at most one path. □

Now we can construct canonical paths for all of A.

Proof of Lemma 4.7. We define canonical paths for each pair $x, y \in A$ as follows. Let S be the set of coordinates (passive variables) where x, y agree and let $T = [n] \setminus S$ be the set of coordinates (active variables) where x, y differ. We use Lemma 4.9 to define canonical paths simultaneously for all pairs of points that have fixed bits on coordinates from S and are antipodal on coordinates from T. We use this construction for each $S \subset [n]$ and every setting of the fixed bits on S. This gives uniquely defined paths joining any two points in A.

Now for any fixed edge $x \sim y$ in A we would like to recover all of the canonical paths which might use the edge (x, y). There are 2^{n-1} possible choices of active and passive variables (the variable on which they differ must be active). If we are given x, y and the passive variables S, then we can recover the bits assigned to variables from this set. We also know that the path connects antipodal points inside a sub-cube generated by $T = [n] \setminus S$. Each edge from this sub-cube is used at most once in such a path. Therefore there are at most 2^{n-1} paths going through each fixed edge. □

4.4 Further Work

To conclude, we list some interesting remaining open questions.

1. Can the canonical paths argument of Section 4.3 be extended to a larger class of lattices? For example, it seems that the argument might adapt without too much difficulty for the lattice $[k]^n$ (or perhaps more generally any product of "nice" posets).

2. Can the $1/n^2$ factor in our bounds for the spectral gap or log-Sobolev constant of a censored chain be improved to $1/n$?

3. Are there conditions other than monotonicity which guarantee that censoring does not greatly impede the mixing time of a Markov chain?

4. Can such a monotonicity argument yield an improved bound for any of the Catalan chains in Chapter 2?

PART II

CHAPTER V

THE UPPER MATCHING CONJECTURE

5.1 Matchings in regular graphs

One of the classic #P-hard problems is calculating the permanent of a $\{0,1\}$-matrix or, equivalently, counting the number of perfect matchings in a (balanced) bipartite graph. Rather than tackling the hard problem of counting matchings in arbitrary graphs we focus here on finding the *extremal* graphs for the number of matchings, that is, the graphs which maximize the number of matchings subject to some constraint. We will focus on graphs with a fixed degree sequence, but there has also been interest in graphs with fixed numbers of vertices and edges [46].

We begin with some definitions.

Definition 5.1. A *matching* in a graph G is a set of pairwise disjoint edges. Write $\phi_k(G)$ for the number of size-k matchings in G, $\phi(G) = \sum_k \phi_k(G)$ for the total number of matchings (including the empty matching), and $\Phi(G) = \phi_{n/2}(G)$ for the number of *perfect* matchings (where n is the number of vertices of G).

For the case of perfect matchings, the celebrated Bregman-Minc theorem gives a tight bound on the permanent of a $\{0,1\}$ matrix with given row sums, showing that among d-regular bipartite graphs on $n = 2kd$ vertices the number of perfect matchings is maximized for a disjoint union of complete bipartite graphs $K_{d,d}$.

Theorem 5.2 ([15])**.** *Let G be a bipartite graph with bipartition $A \cup B$. Then*

$$\Phi(G) \leq \prod_{v \in A} (d(v)!)^{1/d(v)},$$

with equality when G is a disjoint union of complete bipartite graphs.

Following an unpublished result of Kahn and Lovász, many works have independently extended the Bregman-Minc theorem to graphs which are not necessarily bipartite [5, 22, 32], yielding a general bound on the number of perfect matchings in a graph with a given degree sequence which is tight precisely for unions of complete bipartite graphs.

Recently, Davies et al. have shown that a similar extremal result holds not only for the number $\Phi(G)$ of perfect matchings, but also for the total number of matchings $\phi(G)$.

Theorem 5.3 ([23])**.** *If G is a d-regular graph on n vertices, then*

$$\phi(G) \leq \phi(K_{d,d})^{n/2d}.$$

In fact, their result is rather stronger: they show that $kK_{d,d}$ maximizes the *matching polynomial*[1]

$$M_G(\lambda) := \sum_k \phi_k(G)\lambda^k.$$

Indeed, Friedland, Krop and Markström have conjectured that the same is true not only for the entire matching polynomial but even a the level of its coefficients.

Conjecture 5.4 (Upper Matching Conjecture [31])**.** *Among simple (bipartite) d-regular graphs on $2kd$ vertices, the number of matchings of size t (t-matchings) is maximized for a disjoint union of k copies of $K_{d,d}$.*

Friedland et al. proved their conjecture for bipartite graphs in the cases $d = 2$ and $t \leq 4$, and of course the result is known for perfect matchings and for all matchings both regardless of bipartiteness. There is also asymptotic evidence for the truth of the (nonbipartite) upper matching conjecture for matchings of a fixed size [42, 18] (see also [32]).

[1]There are several other definitions of the matching polynomial, all obtainable from each other by suitable transformations, but this one will serve our purposes best.

Friedland et al. also gave a lower matching conjecture, positing that the number of t-matchings is minimized (among r-regular bipartite multigraphs) for a random regular graph. This was first verified asymptotically [38], and has been recently proven in its non-asymptotic form [21].

5.2 Matchings as subgraph counts

In this section we show how the argument of Friedland et al. that Conjecture 5.4 holds for matchings of size ≤ 4 in bipartite graphs can be extended to the nonbipartite setting.

Indeed, it is clear that among d-regular graphs the number of matchings of size 1 (edges) is independent of the graph. To see that the same is true of matchings of size 2, it suffices to consider all pairs of edges and subtract pairs of incident edges—both of which are independent of the graph beyond its size and regularity. Friedland et al. extend this argument up to matchings of size 4. We will give a generalization of this argument which proves Conjecture 5.4 for $t \leq 5$ and further shows that $K_{d,d}$s has the most matchings of each size ≤ 4 among all (not necessarily bipartite) d-regular graphs on n vertices.

We will work in the somewhat more general setting of counting subgraphs of G which are isomorphic to some fixed subgraph F. In the case where $F = tK_2$ is a disjoint union of t edges, subgraphs isomorphic to F are precisely matchings of size t. First we define some useful functions for counting subgraphs in different ways. (The reader may wish to refer to Section 1.2 for notation.)

Definition 5.5. A function $\chi : V(F) \to V(G)$ is a *homomorphism from F to G* if $\{\chi(u), \chi(v)\} \in E(G)$ whenever $\{u, v\} \in E(F)$. Denote the number of homomorphisms by $\hom(F, G)$.

Definition 5.6. A function $\chi : V(F) \to V(G)$ is an *embedding of F into G* if is an injective homomorphism. Denote the number of embeddings by $\mathrm{emb}(F, G)$.

Definition 5.7. Let $\mathrm{sub}(F, G)$ denote the number of (not necessarily induced) subgraphs of G isomorphic to F.

Clearly,

$$\mathrm{sub}(F, G) = \frac{\mathrm{emb}(F, G)}{\mathrm{emb}(F, F)}, \tag{3}$$

where $\mathrm{emb}(F, F)$ is simply the number of automorphisms of F, so for a given subgraph F it suffices to count embeddings in place of subgraphs. The various connections between these functions are treated more thoroughly in Section 7.2 of [50], but for our purposes it is sufficient to note that we can count homomorphisms according to which vertices of F are mapped to the same image:

$$\mathrm{hom}(F, G) = \sum_P \mathrm{emb}(F/P, G), \tag{4}$$

where P ranges over all partitions of $V(F)$ and F/P is the graph obtained from F by contracting each block of P to a single vertex (so we may think of the vertices of F/P as the blocks of P, with two blocks adjacent if there is at least one edge in F between them). The Möbius inversion of this identity is (see [9] for more details)

$$\mathrm{emb}(F, G) = \sum_P \mu_p \,\mathrm{hom}(F/P, G),$$

where

$$\mu_P = (-1)^{n(F)-n(F/P)} \prod_{B \in P} (|B| - 1)!.$$

While we are mostly interested in subgraphs (and hence embeddings), homomorphisms are in some ways much easier to deal with. For example,

$$\mathrm{hom}(F_1 \cup F_2, G) = \mathrm{hom}(F_1, G)\,\mathrm{hom}(F_2, G).$$

Another advantage is that if we know that G is d-regular then the degree 1 vertices of F contribute very predictably to $\mathrm{hom}(F, G)$. In particular,

$$\mathrm{hom}(F, G) = d^{n(F)-n(F^*)}(n/d)^{\mathrm{cc}(F)-\mathrm{cc}(F^*)}\,\mathrm{hom}(F^*, G),$$

where F^* is the graph which is obtained from F by iteratively removing degree 1 and 0 vertices until none remain and $\mathrm{cc}(G)$ is the number of connected components of G. Adding back an isolated vertex contributes a factor of n to the number of homomorphisms (as it can map to any vertex of G), while adding a vertex of degree 1 yields a factor of d (as it can map to any neighbor of its neighbor in F). The graph F^* is commonly known as the 2-*core* of H, since it is the largest subgraph of H with no vertices of degree ≤ 1 (see [13]).

Let us return now to the identity (4). If F has t edges, then the possible graphs F/P all have at most t edges (some edges may coincide). Of course, many different partitions may yield the same graph H, but the number of such partitions depends only on F and H. This means that for some constants c_H we have

$$\mathrm{emb}(F, G) = \sum_{H:\ e(H)\leq t} c_H \,\mathrm{hom}(H, G).$$

With embeddings now written in terms of homomorphisms we can remove the degree-1 vertices.

$$\mathrm{emb}(F, G) = \sum_{H:\ e(H)\leq t} c_H \; d^{n(H)-n(H^*)}(n/d)^{\mathrm{cc}(H)-\mathrm{cc}(H^*)}\,\mathrm{hom}(H^*, G).$$

Finally, we return to embeddings:

$$\mathrm{emb}(F, G) = \sum_{H:\ e(H)\leq t} c_H \; d^{n(H)-n(H^*)}(n/d)^{\mathrm{cc}(H)-\mathrm{cc}(H^*)} \sum_{P'} \mathrm{emb}(H^*/P', G).$$

Unfortunately, it is not necessarily true that every vertex of H^*/P' has degree at least 2, since it is possible for all neighbors of a vertex to be in the same block of P. However, if this is the case then H^*/P' has strictly fewer edges than H. That is, we have written $\mathrm{emb}(F, G)$ as a linear combination of terms $\mathrm{emb}(H, G)$, where H is either a graph on t edges with no vertices of degree 1 or is a graph on strictly fewer than t edges. That is, we have proven

Lemma 5.8. *If G is a d-regular graph on n vertices and F is any graph on t edges, then there exist constants $\hat{c}_H = \hat{c}_H(F, n, d)$ such that*

$$\operatorname{emb}(F, G) = \sum_{H \in \mathcal{H}} \hat{c}_H \operatorname{emb}(H, G),$$

where $\mathcal{H}$ consists of graphs on $\leq t$ edges such that no graph on exactly t edges has any vertex of degree ≤ 1.

The main theorem of this section is the following, which comes as a corollary to Lemma 5.8 by a simple induction on t (we have already mentioned that such a representation is easy if $t = 1$), along with an application of equation (3) to convert embedding counts to subgraph counts.

Theorem 5.9. *If G is a d-regular graph on n vertices and F is any graph, then there exist constants $c_H = c_H(F, n, d)$ otherwise independent of G such that*

$$\operatorname{sub}(F, G) = \sum_{H \in \mathcal{H}} c_H(F, n, d) \operatorname{sub}(H, G),$$

where the sum is over the set $\mathcal{H}$ of isomorphism classes of subgraphs of G with at most t edges and no vertices of degree ≤ 1.

Note that the coefficients $\hat{c}_H$ and c_H are not generally the same, but they do coincide when H has the same number of edges as F.

A similar theorem was actually shown by Beezer & Farrell [7, 8] in a paper we were unable to access, although their intent was to invert the process to determine subgraph counts from the matching polynomial. If we were willing to sacrifice linearity we could also reduce this result further to rely only on counts for *connected* subgraphs, as in [70].

For our application to the UMC we are interested in the particular case where $F = tK_2$ is a union of t disjoint edges, so that $\phi_t(G) = \operatorname{sub}(F, G)$. In this case, Theorem 5.9 tells us that the number of matchings of size t in a d-regular graph G on

n vertices can be determined as a linear combination of subgraph counts for graphs on $\leq t$ edges with minimum degree 2. For example, when $t \leq 4$ the only such subgraphs are the cycles C_3 and C_4, and it is not too taxing to calculate their coefficients.

5.2.1 Small matchings in graphs of large girth

We have just seen that Theorem 5.9 gives an easy way to calculate the number of matchings of size ≤ 4 in bipartite graphs, but more generally it also provides an easy way to calculate the number of matchings in graphs of large girth.

Definition 5.10. The *girth* $\gamma(G)$ of a graph G is the length of the shortest cycle in G. If G is acyclic, we say $\gamma(G) = \infty$.

If $\gamma(G) = \gamma$ then any subgraph of G also has girth at least γ. In particular, a subgraph on fewer than γ edges cannot contain any cycle at all, so for any F with $e(F) < \gamma$ the sum in Theorem 5.9 contains only the term corresponding to the empty graph and hence the number of subgraphs of G isomorphic to F is independent of G (except for the dependence on n and r). Indeed, this is still true as long as F has small *diameter*.

Definition 5.11. The *diameter* of a connected graph G is the maximum length of a shortest path between two vertices of G. If G is disconnected, the diameter of G is the sum of the diameters of its components.

The diameter of a homomomorphic image of F is at most the diameter of F itself, so if F has diameter strictly less than $\gamma(G)$ the number of subgraphs of G isomorphic to F is again independent of G. Taking this one step further, the only subgraphs of G on $< 3\gamma/2$ edges with no vertices of degree 1 are cycles C_k for $\gamma \leq k \leq 3\gamma/2$. (Note that with $3\gamma/2$ edges there are additional such subgraphs, such as the "theta" graph consisting of three paths of length $\gamma/2$ with the same endpoints.) Thus if F has diameter $< 3\gamma/2$ the number of subgraphs of G isomorphic to F depends only on the number of cycles of each length in that range. As applied to matchings,

Corollary 5.12. *Let G be a d-regular graph on n vertices with girth $\gamma(G) > 2t/3$. There are coefficients $a_k(t,n,d)$ so that for any such graph*

$$\phi_t(G) = \sum_{k \le t} a_k(t,n,d)\,\mathrm{sub}(C_k, G).$$

5.2.2 Computing the coefficients

If we take $F = tK_2$ to be a matching of size t, we can actually calculate some of these coefficients by hand for graphs H on exactly t edges. We focus on the case where F and H have the same number of edges, so that $c_H = \hat{c}_H$. Finding the coefficients c_H in Theorem 5.9 is much harder, but can be done recursively for any desired H and F ([70] gives a similar technique).

Proposition 5.13. *For any graph F on t edges and any H on t edges with all degrees 2 or 3,*

$$c_H(F,n,d) = \frac{\mathrm{cov}(F,H)}{\mathrm{emb}(F,F)}(-1)^{n_2(H)}(-2)^{n_3(H)},$$

where $\mathrm{cov}(F,H)$ is the number of homomorphisms from F to H with image isomorphic to H and $n_i(H)$ is the number of vertices in H with degree i.

Proof. If we keep track of the coefficients in the proof of Lemma 5.8, we get

$$\mathrm{sub}(F,G) = \sum_P \left(\mu_P d^{n(F/P)-n(F/P^*)}\left(\tfrac{n}{d}\right)^{\mathrm{cc}(F/P)-\mathrm{cc}(F/P^*)}\right)\frac{\mathrm{aut}((F/P^*)/P')}{\mathrm{aut}(F)}\,\mathrm{sub}((F/P^*)/P', G),$$

where $\mathrm{aut}(G) := \mathrm{emb}(G,G)$ is the number of automorphisms of G.

To determine the coefficient $\hat{c}_H$, we need to know what partitions P and P' yield H from F. Since F and H have the same number of edges, none of the three operations (contracting according to P, taking the 2-core, and contracting according to P') may remove any edges. In particular, we must have $F/P^* = F/P$, so the n and d terms have exponent 0 and $H = (F/P)/P'$. That is, $H = F/P''$ for some partition P'' of $V(F)$, and P may be any refinement of P'' in which all blocks have size at least 2

(since the block sizes are the degrees in F/P). If H has maximum degree 3 then P'' has maximum block size 3, and so we must in fact have $P = P''$.

Now note that there are $\mathrm{cov}(F,H)/\mathrm{emb}(H,H)$ partitions P with F/P isomorphic to H: to get a cover of H by F, first determine a partition P in the set above, then map the parts to vertices of H in any of $\mathrm{emb}(H,H)$ ways. Therefore,

$$\begin{aligned} c_H(F,n,d) &= \frac{\mathrm{emb}(H,H)}{\mathrm{emb}(F,F)} \sum_{P:\ F/P\cong H} \mu_P \\ &= \frac{\mathrm{emb}(H,H)}{\mathrm{emb}(F,F)}(-1)^{n(H)} \frac{\mathrm{cov}(F,H)}{\mathrm{emb}(H,H)} \prod_{v\in V(H)} (d_H(v)-1)! \\ &= \frac{\mathrm{cov}(F,H)}{\mathrm{emb}(F,F)}(-1)^{n_2(H)}(-2)^{n_3(H)}. \end{aligned}$$

□

Indeed, if $F = tK_2$ we can further calculate $\mathrm{cov}(F,H) = \mathrm{emb}(F,F) = 2^t k!$, so for example if H is a union of cycles with total length t then $c_H(F,n,d) = (-1)^t$.

It is now easy to reproduce the result of [31] that Conjecture 5.4 holds for $t \leq 4$. In particular, if one restricts to bipartite graphs then for $t \leq 3$ the expression in Theorem 5.9 has only the term for the empty graph and the number of matchings is independent of the graph. When $t = 4$ it also has a term for the number of 4-cycles in G, with coefficient 1, so one just needs to see that among d-regular graphs on $2kn$ vertices $kK_{d,d}$ has the most 4-cycles. But this is clear, since every edge in $K_{d,d}$ is in $(d-1)^2$ 4-cycles, which is the most possible in a d-regular graph. The graph C_4 is also the only nonempty subgraph which appears in the sum for matchings of size 5. Wanless [70] calculates that the coefficient of C_4 in matchings of size 5 is $nd+n-8d$, which is positive whenever $n \geq 10$ (as necessary to have any matching of size 5), so again to maximize the number of matchings of size 5 it suffices to maximize the number of 4-cycles and this is still achieved by a union of $K_{d,d}$s.

We can also see that Conjecture 5.4 holds even among non-bipartite graphs for $t = 3$, since the coefficient of the 3-cycle is negative and so non-bipartiteness can only

hurt the number of matchings. Wanless also gives the coefficient of C_3 in matchings of size 4 as $-(nd+n-6d)$, which is negative when $n \geq 8$ (as necessary to have any matching of size 4). In other words, to show that the nonbipartite upper matching conjecture holds for $t = 4$ one just has to show that among d-regular graphs on $2kd$ vertices $kK_{d,d}$ maximizes $\text{sub}(C_4, G) - (nd+n-6d)\,\text{sub}(C_3 G)$. But this is again clear, since $kK_{d,d}$ has the most 4-cycles possible and the fewest 3-cycles.

Unfortunately, these calculations become more cumbersome for larger sizes of matchings. Still, this seems to lend some extra evidence in favor of extending Conjecture 5.4 to nonbipartite graphs.

CHAPTER VI

INDEPENDENT SETS AND HOMOMORPISMS

6.1 Extremal problems for independent sets and homomorphisms

The results outlined in the previous chapter suggest a range of similar extremal problems for configurations other than matchings. For example, one could ask which d-regular graphs have the most independent sets.

Definition 6.1. An *independent set* in a graph G is a set of pairwise nonadjacent vertices. Write $i_k(G)$ for the number of size-k independent sets in G and $i(G) = \sum_k i_k(G)$ for the total number of independent sets (including the empty independent set).

Kahn showed that the problem for d-regular, bipartite graphs behaves similarly to the case for matchings.

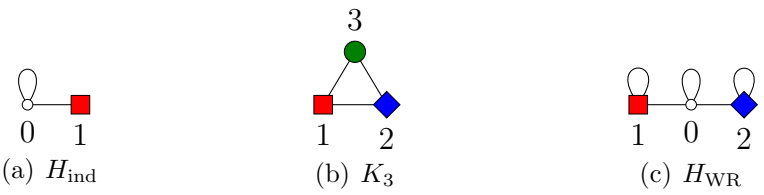
Theorem 6.2. *For any d-regular, bipartite graph on n vertices*

$$i(G) \leq i(K_{d,d})^{n/2d}.$$

Equality holds in Theorem 6.2 if $G = \frac{n}{2d}K_{d,d}$ is a union of $n/2d$ disjoint copies of $K_{d,d}$. In other words, unions of $K_{d,d}$s maximize the total number of independent sets over all d-regular, bipartite graphs on a fixed number of vertices.

More generally, we can consider any labeling of the vertices of G, subject to constraints on the labels. Recall Definition 5.5, repeated here for clarity:

Definition 6.3. Let G and H be graphs. A function $\chi : V(G) \to V(H)$ is a *graph homomorphism* if $\{\chi(u), \chi(v)\} \in E(H)$ for every $\{u, v\} \in E(G)$. Write $\mathrm{Hom}(G, H)$ for the set of graph homomorphisms from G to H and denote $\hom(G, H) := |\mathrm{Hom}(G, H)|$.



(a) H_{ind} (b) K_3 (c) H_{WR}

Figure 13: Homomorphism target graphs yielding (a) independent sets, (b) 3-colorings, and (c) Widom-Rowlinson configurations.

That is, in a homomorphism edges of G must map to edges of H (but there is no requirement on non-edges of G). While most of our graphs in this work will be simple graphs, it is important to consider loops in H.

We saw in Chapter 5 that if we think of G as a small, fixed graph then $\hom(G, H)$ is closely linked to the number of G-subgraphs of H (e.g., counting triangles in H). However, in this chapter we will usually think of H as being a small fixed graph, in which case $\hom(G, H)$ counts *H-colorings* of G: labelings of the vertices of G subject to constraints described by H. For instance, if H is the graph H_{ind} consisting of a single edge $\{0, 1\}$ with a loop at 0, then homomorphisms from G to H correspond to independent sets of G as the preimages of 1. Similarly, homomorphisms from G to the complete (loopless) graph K_t correspond to t-colorings of G. Another target graph which will be of interest is H_{WR}, a path of length three with loops at each vertex. Homomorphisms to H_{WR} are called *Widom-Rowlinson* configurations and model interactions between two types of particles which repel each other. See Figure 13.

For every such model, there is an associated extremal problem asking which d-regular, n-vertex graph G maximizes $\hom(G, H)$. If we wish to compare graphs on different numbers of vertices, it is natural to ask which graph maximizes the scaled quantity $\hom(G, H)^{1/n}$.

In a broad generalization of Kahn's result, Galvin & Tetali [35] showed that the analog of Theorem 6.2 holds for d-regular, bipartite G and *all* target graphs H (including, for example, H_{WR}). And using a cloning construction and a limiting argument,

they showed that in fact the partition function of such models (a weighted count of homomorphisms) is maximized by $K_{d,d}$.

Theorem 6.4 ([35]). *For any d-regular, bipartite graph G on n vertices and any graph H with nonnegative vertex weights $\vec{\lambda} \in \mathbb{R}_+^{V(H)}$,*

$$\hom_{\vec{\lambda}}(G, H)^{1/n} \leq \hom_{\vec{\lambda}}(K_{d,d}, H)^{1/2d},$$

where

$$\hom_{\vec{\lambda}}(G, H) := \sum_{\chi \in \mathrm{Hom}(G,H)} \prod_{v \in V(G)} \lambda_{\chi(v)}$$

is the total weight (or partition function*) of $\vec{\lambda}$-weighted homomorphisms from G to H.*

Note that the case $\vec{\lambda} \equiv 1$ is the counting result.

There is no such sweeping statement for the class of all d-regular graphs with the bipartiteness restriction removed. In [72] and [73], Zhao showed that the bipartiteness restriction on G in Theorem 6.2 and Theorem 6.4 can be removed for some class of graphs H, including H_{ind}. But such an extension is not possible for all graphs H: for example, K_{d+1} has more homomorphisms to H_{WR} than does $K_{d,d}$ (after normalizing for the different numbers of vertices). In fact Galvin conjectured the following:

Conjecture 6.5 ([33, 34]). *For any d-regular graph G on n vertices*

$$\hom(G, H_{WR})^{1/n} \leq \hom(K_{d+1}, H_{WR})^{1/(d+1)}.$$

The more general Conjecture 1.1 of [33] that the maximizing G for any H is either $K_{d,d}$ or K_{d+1} has been disproved by Sernau [64]. However, as we will show in the next section, Conjecture 6.5 is correct.

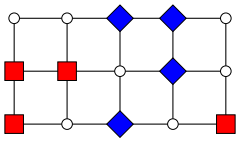


Figure 14: A configuration for the Widom-Rowlinson model on a grid. Vertices mapping to 1 and 2 are shown as squares and diamonds, respectively (corresponding to Figure 13).

6.2 The Widom-Rowlinson model

Homomorphisms $G \to H_{\mathrm{WR}}$, called *Widom-Rowlinson assignments* (or configurations) can be interpreted as modeling configurations of two types of repelling particles on the vertices of G, with the constraint that particles of different types should not occupy adjacent vertices. We think of vertices assigned 1 and 2 as "colored" (or "occupied") and those assigned 0 as "uncolored" (see Figure 14). The (symmetric) *Widom-Rowlinson model* on G is the probability distribution over $\Omega = \mathrm{Hom}(G, H_{\mathrm{WR}})$ suggested by the λ-weighting above:

$$\Pr[\chi] = \frac{\lambda^{X_1(\chi)+X_2(\chi)}}{P_G(\lambda)},$$

where $X_i(\chi)$ are the number of vertices colored i under χ, and

$$P_G(\lambda) = \sum_{\chi \in \Omega} \lambda^{X_1(\chi)+X_2(\chi)}$$

is the partition function for the weighting which assigns weight λ to occupied vertices and weight 1 to unoccupied vertices.

The above theorems of Kahn and Galvin and Tetali are based on the *entropy method* (see [62] and [34] for a survey, or Section 7.2 for an application), but in this context bipartiteness seems essential for the effectiveness of that method. We will approach the problem differently, using the *occupancy method* of [23].

Definition 6.6. Define the *occupancy fraction* $\alpha_G(\lambda)$ to be the expected fraction of vertices which receive a (nonzero) color in the Widom-Rowlinson model:

$$\alpha_G(\lambda) := \frac{\mathrm{E}[X_1 + X_2]}{n}$$

where X_i is the number of vertices colored i by the random assignment χ.

A calculation shows that $\alpha_G(\lambda)$ is in fact the scaled logarithmic derivative of the partition function:

$$\alpha_G(\lambda) = \frac{\lambda}{n} \cdot \frac{P_G'(\lambda)}{P_G(\lambda)} = \frac{\lambda \cdot (\log P_G(\lambda))'}{n}. \tag{5}$$

Our main result is that for any λ, $\alpha_G(\lambda)$ is maximized over all d-regular graphs G by K_{d+1}.

Theorem 6.7. *Let G be any d-regular graph and $\lambda > 0$. Then*

$$\alpha_G(\lambda) \leq \alpha_{K_{d+1}}(\lambda)$$

with equality if and only if G is a union of copies of K_{d+1}.

We will prove this by introducing local constraints on random configurations induced by the Widom-Rowlinson model on a d-regular graph G, then solving a linear programming relaxation of the optimization problem over all d-regular graphs.

Theorem 6.7 implies maximality of the normalized partition function:

Corollary 6.8. *Let G be a d-regular graph on n vertices and $\lambda > 0$. Then*

$$\frac{1}{n} \log P_G(\lambda) \leq \frac{1}{d+1} \log P_{K_{d+1}}(\lambda),$$

or equivalently,

$$P_G(\lambda) \leq P_{K_{d+1}}(\lambda)^{n/(d+1)},$$

with equality if and only if G is a union of K_{d+1}'s.

The quantity $\frac{1}{n} \log P_G(\lambda)$ is known in statistical physics as the *free energy per unit volume.*

Proof of Corollary 6.8. Since $\frac{1}{n}\log P_G(0) = 0$ for every G,

$$\begin{aligned}\frac{1}{n}\log P_G(\lambda) &= \frac{1}{n}\int_0^\lambda (\log P_G(t))'\,dt \\ &\leq \frac{1}{d+1}\int_0^\lambda \left(\log P_{K_{d+1}}(t)\right)'\,dt = \frac{1}{d+1}\log P_{K_{d+1}}(\lambda)\end{aligned}$$

where the inequality follows from Theorem 6.7 and (5). Exponentiating both sides gives Corollary 6.8, and taking $\lambda = 1$ in Corollary 6.8 gives the counting result to prove Conjecture 6.5. □

The method we use in this section is more probabilistic than the entropy method in the sense that Theorem 6.7 gives information about an observable of the model; in some statistical physics models, the analog of $\alpha_G(\lambda)$ would be called the *mean magnetization*. We also work directly in the statistical physics model, instead of counting homomorphisms.

Davies et al. [23] applied the occupancy method to two central models in statistical physics: the hard-core model of a random independent set described above, and the monomer-dimer model of a randomly chosen matching from a graph G. In both cases they showed that $K_{d,d}$ maximizes the occupancy fraction over all d-regular graphs. In the case of independent sets this gives a strengthening of the results of Kahn, Galvin and Tetali, and Zhao, while for matchings, it was not known previously that unions of $K_{d,d}$ maximizes the partition function or the total number of matchings. Another application of the occupancy method can be found in Section 7.3, where it is applied to counting independent sets in hypergraphs.

The idea of calculating the log partition function by integrating a partial derivative is not new of course; see for example, the interpolation scheme of Dembo, Montanari, and Sun [24] in the context of Gibbs distributions on locally tree-like graphs. The method is powerful because it reduces the computation of a very global quantity, $P_G(\lambda)$, to that of a locally estimable quantity, $\alpha_G(\lambda)$.

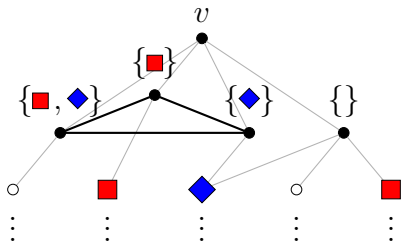


Figure 15: An example configuration with boundary conditions based on a coloring χ. The graph H consists of the four neighbors of v along with the black edges, and the list L_u is shown above each vertex u of H. The colors assigned by χ to v and its neighbors are immaterial and so are not shown.

Some partial results towards the Widom-Rowlinson counting problem were obtained by Galvin [33], who showed that a graph with more homomorphisms than a union of K_{d+1}'s must be close in a specific sense to a union of K_{d+1}'s.

6.2.1 Proof of Theorem 6.7

To prove Theorem 6.7, we will use the following experiment: for a d-regular graph G, we first draw a random χ from the Widom-Rowlinson model, then select a vertex v uniformly at random from $V(G)$. We then write our objective function, the occupancy fraction, in terms of local probabilities with respect to this experiment, and add constraints on the local probabilities that must hold for all G. We then relax the optimization problem to all distributions satisfying the local constraints, and optimize using linear programming.

Fix d and λ. Define a *configuration with boundary conditions* $C = (H, \mathcal{L})$ to be a graph H on d vertices with family of lists $\mathcal{L} = \{L_u\}_{u \in H}$, where each $L_u \subseteq \{1, 2\}$ is a set of allowed colors for the vertex u. Here H represents the neighborhood structure of a vertex $v \in V(G)$ and the color lists L_u represent the colors permitted to neighbors of v, given an assignment χ on the vertices outside of $N(v) \cup \{v\}$. (See Figure 15.) Denote by $\mathcal{C}$ the set of all possible configurations with boundary conditions in any d-regular graph.

We now pick the assignment χ at random from the Widom-Rowlinson model on a

fixed d-regular graph G, pick a vertex v uniformly at random from $V(G)$, and consider the probability distribution induced on $\mathcal{C}$.

For example, if $G = K_{d+1}$ then with probability 1 the random configuration C is $H = K_d$ with $L_u = \{1, 2\}$ for all $u \in V(H)$. If $G = K_{d,d}$ then H is always d isolated vertices and the color lists can be any (possibly empty) subset of $\{1, 2\}$, but the lists must be the same for all $u \in V(H)$.

For a configuration $C = (H, \mathcal{L})$, define

$$\alpha_i^v(C) = \Pr[\chi(v) = i \mid C]$$
$$\alpha_i^u(C) = \frac{1}{d} \sum_{u \in V(H)} \Pr[\chi(u) = i \mid C],$$

where the probability is over the Widom-Rowlinson model on G given the boundary conditions $\mathcal{L}$. Note that the spatial Markov property of the model means that these probabilities are "local" in the sense that they can be computed knowing only C, without any dependence on the remainder of the graph. Let $\alpha^v(C) = \alpha_1^v(C) + \alpha_2^v(C)$ and $\alpha^u(C) = \alpha_1^u(C) + \alpha_2^u(C)$. Then we have

$$\begin{aligned} \alpha_G(\lambda) &= \frac{1}{n} \sum_{v \in V(G)} \Pr[\chi(v) \in \{1, 2\}] = \mathop{\mathrm{E}}_{C}[\alpha^v(C)] \\ &= \frac{1}{nd} \sum_{v \in V(G)} \sum_{u \sim v} \Pr[\chi(u) \in \{1, 2\}] = \mathop{\mathrm{E}}_{C}[\alpha^u(C)], \end{aligned} \tag{6}$$

where the expectations are over the probability distribution induced on $\mathcal{C}$ by our experiment of drawing χ from the model and v uniformly at random from $V(G)$, and the last sum is over all neighbors of v in G. Equality of the two expressions for α follows since sampling a uniform neighbor of a uniform vertex in a regular graph is equivalent to sampling a uniform vertex. We will show that this expectation is maximized when the graph G is K_{d+1}.

We can in fact write explicit formulae for $\alpha^v(C)$ and $\alpha^u(C)$. For a configuration $C = (H, \mathcal{L})$, let $P_C^{(0)}(\lambda)$ be the total weight of colorings of H satisfying the boundary conditions given by the lists $\mathcal{L}$ (corresponding to the partition function for the

neighborhood of v conditioned on $\chi(v) = 0$). Also, write $P_C^{(i)}(\lambda)$ for the total weight of colorings of H satisfying the boundary conditions and using only color i and 0 (corresponding to the partition functions for the neighborhood of v conditioned on $\chi(v) = i$). Finally, let $P_C^{(12)}(\lambda) = P_C^{(1)}(\lambda) + P_C^{(2)}(\lambda)$ and let

$$P_C(\lambda) = P_C^{(0)}(\lambda) + \lambda P_C^{(12)}(\lambda)$$

be the partition function of $N(v) \cup \{v\}$ conditioned on the boundary conditions given by C. Note that if $\mathcal{L}$ has a_1 lists containing 1 and a_2 lists containing 2, then $P_C^{(i)}(\lambda) = (1+\lambda)^{a_i}$.

Now we can write

$$\alpha^v(C) = \frac{\lambda P_C^{(12)}}{P_C} \quad \text{and} \quad \alpha^u(C) = \frac{\lambda\Big((P_C^{(0)})' + \lambda(P_C^{(12)})'\Big)}{d\, P_C}, \tag{7}$$

where P' is the derivative of P in λ. We will suppress the dependence of the partition functions on λ from now on.

For $G = K_{d+1}$, we have

$$P_{K_{d+1}} = 2(1+\lambda)^{d+1} - 1$$
$$\alpha_{K_{d+1}}(\lambda) = \frac{2\lambda(1+\lambda)^d}{2(1+\lambda)^{d+1} - 1}.$$

If $G = K_{d+1}$ then the only possible configuration is $C_{K_{d+1}}$, the complete neighborhood K_d with full boundary lists, so we also have $\alpha^u(K_d) = \alpha^v(K_d) = \alpha_{K_{d+1}}(\lambda)$ (we can also compute these directly). Since this quantity will arise frequently, we will use the notation $\alpha_K := \alpha_{K_{d+1}}(\lambda)$.

Let $q : \mathcal{C} \to [0,1]$ denote a probability distribution over the set of all possible configurations. We consider the following optimization problem over the variables

$q(C)$, $C \in \mathcal{C}$.

$$\begin{aligned} \alpha^* = \max \sum_{C \in \mathcal{C}} q(C)\alpha^v(C) \qquad & \text{subject to} \\ \sum_{C \in \mathcal{C}} q(C) = 1 & \\ \sum_{C \in \mathcal{C}} q(C)[\alpha^v(C) - \alpha^u(C)] = 0 & \\ q(C) \geq 0 \quad \forall C \in \mathcal{C}. & \end{aligned} \tag{8}$$

This linear program is a relaxation of our optimization problem of maximizing $\alpha_G(\lambda)$ over all d-regular graphs: any such graph induces a probability distribution on $\mathcal{C}$, and as we have seen above in (6), the constraint asserting the equality $\mathrm{E}[\alpha^v(C)] = \mathrm{E}[\alpha^u(C)]$ must hold in all d-regular graphs.

We will show that for any $\lambda > 0$ the unique optimal solution of this linear program is $q(C_{K_{d+1}}) = 1$, where $C_{K_{d+1}}$ is the configuration induced by K_{d+1} (i.e., $H = K_d$ and $L_u = \{1, 2\}$ for all $u \in H$).

The dual of the above linear program is

$$\begin{aligned} \alpha^* = \min \Lambda_p \qquad & \text{subject to} \\ \Lambda_p + \Lambda_c(\alpha^v(C) - \alpha^u(C)) \geq \alpha^v(C) \quad \forall C \in \mathcal{C}, & \end{aligned}$$

with decision variables Λ_p and Λ_c.

To show that the optimum is attained by $C_{K_{d+1}}$, we must find a feasible solution to the dual program with $\Lambda_p = \alpha_K = \frac{2\lambda(1+\lambda)^d}{2(1+\lambda)^{d+1}-1}$. Note that with $\Lambda_p = \alpha_K$ the constraint for $C_{K_{d+1}}$ holds with equality for any choice of Λ_c. In other words, it suffices to find some convex combination of the two local estimates α^u and α^v which is maximized by $C_{K_{d+1}}$ over all $C \in \mathcal{C}$.

Let C_0 be a configuration with $L_u = \emptyset$ for all $u \in H$ (in which case the edges of H are immaterial, and so abusing notation we will refer to any one of these configurations as C_0). We find a candidate Λ_c by solving the constraint corresponding to C_0 with

equality:

$$\begin{aligned}\alpha_K &= \Lambda_c(\alpha^u(C_0) - \alpha^v(C_0)) + \alpha^v(C_0) \\ &= (1-\Lambda_c)\frac{2\lambda}{1+2\lambda}.\end{aligned}$$

This gives

$$\Lambda_c = 1 - \frac{\alpha_K}{2\lambda}(1+2\lambda) = \frac{\alpha_K}{2\lambda}\frac{(1+\lambda)^d - 1}{(1+\lambda)^d}.$$

With this choice of Λ_c, the general dual constraint is

$$\alpha_K \geq \frac{\alpha_K}{2\lambda}\frac{(1+\lambda)^d - 1}{(1+\lambda)^d}\alpha^u(C) + \frac{\alpha_K}{2\lambda}(1+2\lambda)\alpha^v(C).$$

Plugging in the formulae for α^u and α^v from (7), this becomes

$$\frac{(P_C^{(0)})' + \lambda(P_C^{(12)})'}{2P_C^{(0)} - P_C^{(12)}} \leq \frac{d(1+\lambda)^d}{(1+\lambda)^d - 1}. \tag{9}$$

From this point on we may assume that C has some non-empty color list, since otherwise the configuration is equivalent to C_0 and the constraint holds with equality by our choice of Λ_c. This assumption tells us, among other things, that $(P_C^{(0)})' > 0$ and $2P_C^{(0)} - P_C^{(12)} > 0$.

Our goal is now to show that (9) holds for all C. We consider the two terms separately.

Claim 6.9. *For any $C \neq C_0$,*

$$\frac{\lambda(P_C^{(12)})'}{2P_C^{(0)} - P_C^{(12)}} \leq \frac{d\lambda(1+\lambda)^{d-1}}{(1+\lambda)^d - 1}$$

with equality if and only if the lists L_u are all equal and C has no dichromatic colorings.

Proof. Since the partition function $P_C^{(0)}$ is at least the total weight $P_C^{(1)} + P_C^{(2)} - 1$ of monochromatic colorings (with equality when C has no dichromatic colorings), we have

$$\frac{(P_C^{(12)})'}{2P_C^{(0)} - P_C^{(12)}} \leq \frac{(P_C^{(12)})'}{P_C^{(12)} - 2} = \frac{a_1(1+\lambda)^{a_1-1} + a_2(1+\lambda)^{a_2-1}}{(1+\lambda)^{a_1} + (1+\lambda)^{a_2} - 2} \tag{10}$$

(where, as above, a_i is the number of vertices in H allowed color i under the given boundary conditions), and so we need to show that

$$\frac{a_1(1+\lambda)^{a_1-1} + a_2(1+\lambda)^{a_2-1}}{(1+\lambda)^{a_1} + (1+\lambda)^{a_2} - 2} \leq \frac{d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1}.$$

In general, to show that $(a+b)/(c+d) \leq t$ it suffices to show that $a/c \leq t$ and $b/d \leq t$. Thus it is enough to show that

$$\frac{a(1+\lambda)^{a-1}}{(1+\lambda)^a - 1} \leq \frac{d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} \tag{11}$$

whenever $1 \leq a \leq d$. (Note that if either $a_1 = 0$ or $a_2 = 0$ then (10) reduces to (11), and if both $a_1, a_2 = 0$ then the configuration is C_0). Indeed, it is not hard to check via calculus that the left hand side of (11) is increasing with a. This completes the proof of the inequality in Claim 6.9.

We have equality in this final step when $a_1 = a_2 = d$ or when one is 0 and the other is d. So we have equality overall whenever the lists are all equal and there are no dichromatic colorings (recall that we are assuming C has some non-empty coloring list). □

Claim 6.10. *For any $C \neq C_0$,*

$$\frac{(P_C^{(0)})'}{2P_C^{(0)} - P_C^{(12)}} \leq \frac{d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1},$$

with equality if and only if the lists L_u are all equal and C has no dichromatic colorings.

Proof. We can write

$$\begin{aligned}\frac{\lambda(P_C^{(0)})'}{2P_C^{(0)} - P_C^{(12)}} &= \frac{\lambda(P_C^{(0)})'}{P_C^{(0)}} \cdot \frac{P_C^{(0)}}{(P_C^{(0)} - P_C^{(1)}) + (P_C^{(0)} - P_C^{(2)})} \\ &= \frac{\mathrm{E}_C[X_1] + \mathrm{E}_C[X_2]}{\mathrm{Pr}_C[X_1 > 0] + \mathrm{Pr}_C[X_2 > 0]},\end{aligned}$$

where now X_i is the number of vertices colored i in a random coloring chosen from the Widom-Rowlinson model on C. Noting that $\mathrm{E}_C[X_1] = 0$ whenever $\mathrm{Pr}_C[X_1 > 0] = 0$,

it suffices as above to show that whenever color 1 is permitted anywhere in C,

$$\frac{\mathrm{E}_C[X_1]}{\Pr_C[X_1 > 0]} = \mathop{\mathrm{E}}_{C}[X_1 \mid X_1 > 0] \leq \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} = \mathop{\mathrm{E}}_{K_d}[X_1 \mid X_1 > 0], \tag{12}$$

and similarly for X_2, but this will follow by symmetry.

We can decompose the expectation as

$$\mathop{\mathrm{E}}_{C}[X_1 \mid X_1 > 0] = \sum_{S \subseteq V(H)} \Pr_C[\chi^{-1}(2) = S \mid X_1 > 0] \cdot \mathop{\mathrm{E}}_{C}[X_1 \mid X_1 > 0,\ \chi^{-1}(2) = S].$$

The partition function restricted to colorings satisfying $X_1 > 0$ and $\chi^{-1}(2) = S$ is just $P_S(\lambda) = \lambda^{|S|}((1+\lambda)^{a_S} - 1)$, where a_S is the number of vertices in $H \setminus S$ which are allowed color 1 and are not adjacent to any vertex of S. The conditional expectation is then

$$\mathop{\mathrm{E}}_{C}[X_1 \mid X_1 > 0,\ \chi^{-1}(2) = S] = \frac{a_S \lambda (1+\lambda)^{a_S - 1}}{(1+\lambda)^{a_S} - 1} \leq \frac{d\lambda(1+\lambda)^{d-1}}{(1+\lambda)^d - 1}$$

with equality precisely when S is empty and 1 is available for every vertex. That is,

$$\mathop{\mathrm{E}}_{C}[X_1 \mid X_1 > 0] \leq \sum_{S \subseteq V(H)} \Pr_C[\chi^{-1}(2) = S \mid X_1 > 0] \cdot \frac{d\lambda(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} = \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1},$$

as desired. We have equality in (12) when $\Pr_C[a_S = d \mid X_1 > 0] = 1$, which holds for the configurations where 1 is available to every vertex but which have no dichromatic colorings. That is, for equality to hold in the claim C must have no dichromatic colorings, and any color which is available to some vertex u must be available to every vertex (so the lists must be identical). □

Adding the inequalities in Claims 6.9 and 6.10 shows that (9) holds for all C, proving optimality of K_{d+1}.

6.2.2 Uniqueness

Lemma 6.11. *The distribution induced by K_{d+1} is the unique optimum of the LP relaxation* (8)*.*

Proof. Complementary slackness for our dual solution says that any optimal primal solution is supported only on configurations C with identical boundary lists and no dichromatic colorings. These fall into three categories:

Case 0 $L_u = \emptyset$ for all u. In this case the edges of H are immaterial, as none of H can be colored. This is the configuration C_0 above.

Case 1 $L_u = \{i\}$ for all u (for $i = 1$ or 2). The edges of H are again immaterial, as every coloring of H with only color i is allowed. Call this configuration C_1.

Case 2 $L_u = \{1, 2\}$ for all u. In this case the prohibition on dichromatic colorings requires that $C = C_{K_{d+1}}$.

We can calculate $\alpha^v(C)$ and $\alpha^u(C)$ for each case. For Case 0 we have

$$\alpha^v(C_0) = \frac{2\lambda}{1+2\lambda} \qquad \text{and} \qquad \alpha^u(C_0) = 0.$$

For Case 1 we have

$$\alpha^v(C_1) = \frac{\lambda + \lambda(1+\lambda)^d}{\lambda + (1+\lambda)^{d+1}} \qquad \text{and} \qquad \alpha^u(C_1) = \frac{\lambda(1+\lambda)^d}{\lambda + (1+\lambda)^{d+1}}.$$

And of course, for Case 2 we have

$$\alpha^v(K_d) = \alpha^u(K_d) = \alpha_K.$$

In both Case 0 and Case 1 we have $\alpha^u < \alpha^v$, so the only convex combination q of the three cases giving $\sum_C q(C)\alpha^u(C) = \sum_C q(C)\alpha^v(C)$ (as is required for feasibility) is the one which puts all of the weight on $C_{K_{d+1}}$. □

6.2.3 The asymmetric model

It is also natural to consider a weighted version of the Widom-Rowlinson model with distinct activities λ_1, λ_2 for the two colors, so that the configuration χ is chosen with probability proportional to $\lambda_1^{X_1}\lambda_2^{X_2}$, and where the partition function $P_G(\lambda_1, \lambda_2)$ is again the normalizing factor. We can ask which graphs maximize $P(\lambda_1, \lambda_2)^{1/n}$. We conjecture

Conjecture 6.12. *For any $\lambda_1, \lambda_2 > 0$, the* weighted occupancy fraction

$$\overline{\alpha}_G(\lambda_1, \lambda_2) := \frac{\lambda_2 \alpha_G^1(\lambda_1, \lambda_2) + \lambda_1 \alpha_G^2(\lambda_1, \lambda_2)}{\lambda_1 + \lambda_2}$$

is maximized over all d-regular graphs by K_{d+1}.

In fact, Conjecture 6.12 implies the following conjecture on the maximality of the partition function:

Conjecture 6.13. *For any nonnegative weights $\vec{\lambda} \in \mathbb{R}_+^{\{0,1,2\}}$, and any d-regular graph G on n vertices,*

$$\hom_{\vec{\lambda}}(G, H_{WR}) \leq \hom_{\vec{\lambda}}(K_{d+1}, H_{WR})^{n/(d+1)}.$$

To see this, assume without loss of generality that $\lambda_0 = 1$ and $\lambda_1 \geq \lambda_2$, and let $F_G(x) = \frac{1}{n} \log P_G(\lambda_1 - \lambda_2 + x, x)$. We have

$$\frac{1}{n} \log P_G(\lambda_1, \lambda_2) = F_G(\lambda_2) = F_G(0) + \int_0^{\lambda_2} \frac{dF_G}{dx}(x)\, dx$$

$F_G(0) = \frac{1}{n} \log P_G(\lambda_1 - \lambda_2, 0) = \log(1 + \lambda_1 - \lambda_2)$ for all graphs G, and so if we can show that for all $0 \leq x \leq \lambda_2$, $\frac{dF_G}{dx}(x)$ is maximized when $G = K_{d+1}$, then we obtain (the log of) Conjecture 6.13. We compute:

$$\begin{aligned}
\frac{dF_G}{dx}(x) &= \frac{1}{n} \frac{\frac{d}{dx} P_G(\lambda_1 - \lambda_2 + x, x)}{P_G(\lambda_1 - \lambda_2 + x, x)} \\
&= \frac{1}{n} \frac{\sum_\chi \frac{xX_1 + (\lambda_1 - \lambda_2 + x)X_2}{x(\lambda_1 - \lambda_2 + x)} (\lambda_1 - \lambda_2 + x)^{X_1} \cdot x^{X_2}}{P_G(\lambda_1 - \lambda_2 + x, x)} \\
&= \frac{1}{x(\lambda_1 - \lambda_2 + x)} \frac{1}{n} \frac{\sum_\chi (xX_1 + (\lambda_1 - \lambda_2 + x)X_2)(\lambda_1 - \lambda_2 + x)^{X_1} \cdot x^{X_2}}{P_G(\lambda_1 - \lambda_2 + x, x)} \\
&= \frac{1}{x(\lambda_1 - \lambda_2 + x)} \left[x \alpha_G^{(1)}(\lambda_1 - \lambda_2 + x, x) + (\lambda_1 - \lambda_2 + x) \alpha_G^{(2)}(\lambda_1 - \lambda_2 + x, x) \right].
\end{aligned}$$

Conjecture 6.12 implies that this is maximized by K_{d+1}.

6.3 Other nonbipartite extremal graphs

Shortly after the work of the previous section, another method was found to prove Conjecture 6.5 (and more generally the weighted version Conjecture 6.13). The new

approach uses a bijective argument similar to the one applied by Zhao [72] to extend Theorem 6.2 to nonbipartite graphs, but with a very different outcome. Like Zhao's argument, this method actually gives a class of target graphs H for which unions of K_{d+1} maximize the number of homomorphisms. The method is very similar to the approach of Sernau [64] (which after correcting a small error was later discovered to yield the same result).

CHAPTER VII

MATCHINGS AND INDEPENDENT SETS IN HYPERGRAPHS

7.1 Unified setting for hypergraphs

A *hypergraph* $G = (V, E)$ is a set V of vertices along with a collection E of subsets of V called *hyperedges*. The hypergraph is called *r-uniform* if every edge contains exactly r vertices, and *d-regular* if every vertex is in exactly d edges. Furthermore, a hypergraph is *linear* (or *simple*) if no two edges intersect in more than one vertex. For instance, a 2-uniform hypergraph is just a graph, and it is simple if and only if the graph has no multiple edges.

There are several ways to extend the usual definition of an independent set in a graph to the hypergraph setting. We will use the following.

Definition 7.1. An *independent set* in a hypergraph G is a set $I \subseteq V$ of vertices such that $|I \cap e| \leq 1$ for every hyperedge $e \in E$, i.e., no two vertices in I are in a common hyperedge of G.

A different definition of a hypergraph independent set might only require that I should not contain any edge in its entirety, or perhaps more generally (for r-regular hypergraphs) that $|I \cap e| \leq t$ for some fixed $1 \leq t < r$. We choose to use the strong requirement above in part because of its close relationship to matchings: if one considers the hypergraph $G^{\mathsf{T}} = (E, V')$ obtained by exchanging the roles of vertices and hyperedges in G (where $V' = \{\{e | v \in e\}\}_{v \in V}$, so that the edge-vertex incidence matrix of G^{T} is the transpose of that of G) then an independent set in G corresponds precisely to a matching (set of pairwise disjoint hyperedges) in G^{T}. Furthermore, if G is a d-regular, r-uniform, linear hypergraph, then G^{T} is an r-regular, d-uniform,

linear hypergraph. That is, a tight bound on the maximum number of independent sets in a uniform, regular hypergraph is equivalent to a similar bound for matchings (and vice versa).

The best prior bound on independent sets in a uniform, regular hypergraph is due to Ordentlich & Roth.

Theorem 7.2 ([60]). *If G is an r-uniform, d-regular, linear hypergraph on n vertices then*

$$\frac{\lg i(G)}{n} \leq \frac{1}{r} + O\left(\frac{\log^2(rd)}{rd}\right).$$

This is not tight for $r = 2$, where the correct bound for independent sets (due to [45]) is

$$\frac{\lg i(G)}{n} \leq \frac{\lg i(K_{d,d})}{2d} = \frac{\lg(2^{d+1} - 1)}{2d} = \frac{1}{2} + \Theta\left(\frac{1}{d}\right).$$

Nor is it tight for $d = 2$, where the correspondence to matchings in an r-regular graph yields

$$\frac{\lg i(G)}{n} = \frac{\lg \mu(G^{\mathsf{T}})}{n} \leq \frac{\lg \mu(K_{r,r})}{r^2} = \Theta\left(\frac{\log r}{r}\right)$$

due to Davies et al.

Ordentlich & Roth were interested in bounding the number of independent sets in the *Hamming hypergraph* $H_{[r]^d}$ on vertex set $[r]^d$ with hyperedges consisting of sets of vertices differing only in a single coordinate (Hamming distance 1). Indeed, the Hamming hypergraph is the maximizer in the case of $d = 2$, since $H_{[r]^2} = K_{d,d}^{\mathsf{T}}$.

We give the following conjecture unifying the known bounds for independent sets and matchings in graphs.

Conjecture 7.3. *If G is a d-regular, r-uniform linear hypergraph on n vertices, then*

$$\frac{\lg i(G)}{n} \leq \frac{1}{r} + O\left(\frac{\log r}{rd}\right).$$

For $d, r > 2$ there is no obvious contender for the extremal hypergraph. There are several natural ways to extend the notion of complete bipartite graph to the

hypergraph setting, but among the hypergraphs we have considered the maximizer of $\frac{1}{n}\log P_G(\lambda)$ depends on the choice of λ. If it turns out to be the case that the overall maximizer does in fact depend on λ then it may indeed be very hard to pin down a general description of the maximizing hypergraph.

In this chapter we will give two approaches for improving this bound for independent sets in hypergraphs. The first, which improves upon the bound in Theorem 7.2 in the regime where r is small, is derived using the entropy technique pioneered by Radhakrishnan [61] in his elegant proof of the Bregman-Minc theorem, and (like the Bregman-Minc theorem itself) requires that the hypergraph satisfy a partiteness constraint. The second approach uses the occupancy fraction method pioneered by Davies et al. to work towards a bound which would indeed unify the known results for independent sets ($r = 2$) and matchings ($d = 2$), with a complete proof for the specific case of *triangle-free* 3-uniform hypergraphs. While these new results require slightly stronger hypotheses than Theorem 7.2, they both still apply to the motivating case of independent sets in the Hamming cube $H_{[r]^d}$.

7.2 Entropy bound

As in the other entropy-based bounds for independent sets and matchings, our result will require a bipartiteness condition. As with independent sets, there are several ways to extend the definition of graph bipartiteness to the hypergraph setting. The definition which seems the most useful in this context is the following, first given by Aharoni & Kessler [1].

Definition 7.4. A hypergraph $G = (V, E)$ is *bipartite* if there is a distinguished set $A \subseteq V$ of vertices such that $|A \cap e| = 1$ for every hyperedge $e \in E$.

Now we can state the main result of this section.

Theorem 7.5. *Let G be an r-uniform, d-regular, bipartite, linear hypergraph on n*

vertices. Then

$$\frac{\lg i(G)}{n} \leq \frac{1}{rd} \lg(r^d + 2^d - 1) = O\left(\frac{\log r}{r}\right).$$

In the regime where d is constant, the second term dominates in Theorem 7.2 and this result is tighter. On the other hand, if d is relatively large (e.g., $d > \lg^2 r$) Theorem 7.2 is tighter. In fact, the known extremal result for matchings shows that this bound gives the best possible order of growth (at the log level) for constant d. In particular, while this theorem requires a bipartiteness hypothesis which Theorem 7.2 does not, it still applies to their original motivating example of the Hamming hypergraph, giving a new, tighter bound in that case when d is constant and r is large.

7.2.1 Entropy preliminaries

In order to prove Theorem 7.5, we will need a few key definitions and lemmas. This treatment is largely copied from [45] (see [54] for a fuller discussion).

Definition 7.6. The (binary) *entropy* of a discrete random variable X is

$$H(X) := \mathrm{E}(\lg \frac{1}{p(X)}) = \sum_x p(x) \lg \frac{1}{p(x)},$$

where we write $p(x) := \Pr(X = x)$. (We extend this notation in natural ways below.) The *conditional entropy* $H(X|Q)$ of X given an event Q is just the entropy of the random variable X conditioned on Q. More generally, the conditional entropy of X given another random variable Y is

$$H(X \mid Y) := \mathrm{E}\, H(X|\{Y = y\}) = \sum_y p(y) \sum_x p(x|y) \lg \frac{1}{p(x|y)}.$$

Roughly speaking, the binary entropy is the number of bits of information necessary to determine a random variable.

For a random vector $X = (X_1, \ldots, X_k)$ we have

$$H(X) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_k \mid X_1, \ldots, X_{k-1}). \tag{13}$$

We will make use of the classical inequalities

$$H(X) \leq \lg|\text{supp}(X)|$$

(with equality if X is uniform over its support) and

$$H(X \mid Y) \leq H(X). \tag{14}$$

In general if Y determines Z then $H(X|Y) \leq H(X|Z)$. For the random vector above, $H(X) \leq \sum_i H(X_i)$, and more generally

$$H(X|Y) \leq \sum_i H(X_i \mid Y). \tag{15}$$

Together, (13) and (14) imply

$$H(X) \leq H(Y) + H(X \mid Y)$$

Finally, we will take advantage of a less classical inequality due to Shearer [20] which generalizes (15). For a random variable $X = (X_1, \ldots, X_k)$ and $A \subseteq [k]$, let X_A denote $(X_i)_{i \in A}$.

Lemma 7.7 ([20])**.** *Let $X = (X_1, \ldots, X_n)$ be a random vector and let $\mathcal{A}$ be an m-*cover *of $[n]$, i.e., a collection of subsets (possibly with repeats) of $[n]$ such that each member of $[n]$ is contained in at least m members of $\mathcal{A}$. Then*

$$H(X) \leq \frac{1}{m} \sum_{A \in \mathcal{A}} H(X_A).$$

7.2.2 Proof of Theorem 7.5

Our proof of Theorem 7.5 is very much along the lines of Kahn's proof in the graph case. The idea is to bound the number of independent sets by bounding the entropy of a uniform random independent set. The random independent set, thought of as an indicator vector, can be decomposed into smaller vectors using Lemma 7.7, whose entropy can be bounded individually.

Proof of Theorem 7.5. For a pair of vertices $u \neq v \in V$ of a hypergraph G, say $u \sim v$ if there is some hyperedge $e \in E$ with $u, v \in e$, and let $N(v) = \{u \mid u \sim v\}$ denote the neighborhood of v. For a subset $S \subseteq V$ let $G[S]$ denote the induced subhypergraph of G on the vertices of S, i.e., with hyperedges $\{e \cap S\}_{e \in E}$.

Let X be the indicator vector for a uniform random independent set I of G, and let $A \subseteq V$ be the distinguished set guaranteed by biparteness of G. Note that since G is regular and exactly one of the r vertices in each edge is in A we must have $|A| = n/r$. For $v \in V$ let C_v be the event that v is covered by I, i.e., $I \cap N(v) \neq \emptyset$ (so that $I \cup \{v\}$ is not independent). Then (see below for more explanation)

$$
\begin{aligned}
\lg i(G) &= H(X) \\
&= H(X_{V \setminus A}) + H(X_A \mid X_{V \setminus A}) \\
&\leq \frac{1}{d} \sum_{v \in A} H(X_{N(v)}) + \sum_{v \in A} H(X_v \mid X_{V \setminus A}) && (16) \\
&\leq \frac{1}{d} \sum_{v \in A} \Big(H(C_v) + H(X_{N(v)} \mid C_v) + d H(X_v \mid C_v) \Big) && (17) \\
&\leq \frac{1}{d} \sum_{v \in A} \Pr[C_v] \left(\lg \frac{1}{\Pr[C_v]} + \lg |\mathrm{supp}(X_{N(v)} \mid C_v)| + d \lg |\mathrm{supp}(X_v \mid C_v)| \right) \\
&\qquad + \Pr[\overline{C_v}] \left(\lg \frac{1}{\Pr[\overline{C_v}]} + \lg |\mathrm{supp}(X_{N(v)} \mid \overline{C_v})| + d \lg |\mathrm{supp}(X_v \mid \overline{C_v})| \right) \\
&\leq \frac{1}{d} \sum_{v \in A} \Pr[C_v] \left(\lg \frac{1}{\Pr[C_v]} + \lg(i(G[N(v)]) - 1) + d \lg 1 \right) \\
&\qquad + \Pr[\overline{C_v}] \left(\lg \frac{1}{\Pr[\overline{C_v}]} + \lg 1 + d \lg 2 \right) && (18) \\
&\leq \frac{1}{d} \sum_{v \in A} \left(\Pr[C_v] \lg \frac{r^d - 1}{\Pr[C_v]} + \Pr[\overline{C_v}] \lg \frac{2^d}{\Pr[\overline{C_v}]} \right) && (19) \\
&\leq \frac{1}{d} \sum_{v \in A} \lg(r^d + 2^d - 1) && (20) \\
&= \frac{n}{rd} \lg(r^d + 2^d - 1)
\end{aligned}
$$

For (16) we use Lemma 7.7 with the cover $\mathcal{A} = \{N(v)\}_{v \in A}$, as linearity tells us that each vertex not in A has d distinct neighbors in A and so is counted in d of these

sets. The next inequality (17) follows from inequality (14) because $X_{V\setminus A}$ determines C_v for each $v \in V$. For (18), note that if v is uncovered then $X_{N(v)}$ must be all zeros, while if v is covered $X_{N(v)}$ can be any nonempty independent set of $G[N(v)]$. Similarly, if v is covered then X_v must be 0, while if it is uncovered X_v may be either 0 or 1. For (19) we note that $G[N(v)]$ consists of d disjoint edges of size $r-1$ (the remains of the edges incident to v) along with, perhaps, some more partial edges. An independent set contains at most one of the $r-1$ vertices from each, for at most r^d independent sets. Finally, for (20) we use concavity of the logarithm. □

We conjecture that (as is the case for graphs) the bipartiteness condition is not necessary, although it plays a crucial role in this proof. In the graph case the need for bipartiteness was alleviated by a clever bijection due to [72], which does not readily adapt to hypergraphs.

7.2.3 Further work

Actually, more recently this result has been extended to homomorphisms in hypergraphs (along the lines of [51]). For two hypergraphs G and F, $\chi : V(G) \to V(F)$ is a homomorphism if $\{\chi(v)\}_{v\in e} \in E(F)$ for every $e \in E(G)$.

Theorem 7.8 ([65])**.** *For any r-uniform, d-regular, linear hypergraph G on n vertices and any r-uniform hypergraph F (with loops allowed, so that the hyperedges may be multisets),*

$$\lg \hom(G, F) \leq \frac{1}{d} \sum_{v\in V(G)} \lg \hom(K_{r\times p(v)}, F)$$

where $p(v)$ is the number of edges in which v is the last vertex (according to some arbitrary fixed ordering), and $K_{r\times p}$ is the r-uniform, p-regular (non-linear) graph constructed by expanding each vertex on one side of $K_{p,p}$ into $r-1$ vertices.

In particular, if G is a bipartite hypergraph then putting the distinguished set (of

size n/r) at the end of the ordering gives

$$\frac{1}{n}\lg \hom(G,F) \leq \frac{1}{rd}\lg \hom(K_{r\times d},F).$$

We can obtain independent sets as homomorphisms by taking F to be the hypergraph with $V(F) = \{v_0, v_1\}$ and $E(F) = \{\{v_0,\ldots,v_0\},\{v_0,\ldots,v_0,v_1\}\}$. Since $i(K_{r\times d}) = 2^d + r^d - 1$, Theorem 7.5 follows as a corollary. Indeed, by using different target hypergraphs F this homomorphism result also gives bounds for the other definitions of hypergraph independent sets mentiond above.

While the same argument in the case of graphs immediately gives an extremal graph, this is not the case for hypergraphs, since we have bounded the number of independent sets in uniform, regular, *linear* hypergraphs by the number of independent sets in a hypergraph which is highly non-linear, so it is possible that the bound may be improved. In fact, while this argument agrees with the correct bound for $r = 2$, it is very far from the truth when $d = 2$ as r increases (i.e., graph matchings).

It may well be possible to remove the bipartiteness condition from these results for hypergraphs, just as for graphs. Zhao [72] achieves this in the graph case by giving an injection from independent sets in an arbitrary graph to independent sets in a bipartite version of that graph, an argument which he was later able to extend to a larger (but still relatively small) class of graph homomorphisms [73]. However, it is unclear how to extend this argument to hypergraphs, even in the simple case of independent sets when $r = 3$, since the most natural bipartite construction is not linear.

7.3 The Occupancy Method

To prove our second bound on independent sets in regular hypergraphs we will use the occupancy method [23], which we have already used in Section 6.2 to find the extremal graph for Widom-Rowlinson configurations. While in this instance we do not immediately find an extremal hypergraph, the basic method is very similar, and hinges

on the following observation about the hard-core distribution for random independent sets.

Consider a random independent set I chosen from the probability distribution

$$\Pr[I] := \frac{\lambda^{|I|}}{P_G(\lambda)}.$$

(This distribution is well known in the statistical physics literature as the *hard-core model*.) Call a vertex $x \in I$ *occupied* and let

$$\alpha_G := \mathrm{E}[|I|]/n$$

be the expected fraction of vertices which are occupied by this random independent set. Then we can write

$$\alpha_G = \frac{\lambda\, P_G'(\lambda)}{n\, P_G(\lambda)} = \lambda \frac{d}{d\lambda}\left[\frac{1}{n}\ln P_G(\lambda)\right]. \tag{21}$$

Since $\ln P_G(0) = \ln 1 = 0$ for all G, integrating over λ shows that a graph which maximizes α_G for all λ is also the maximizer for $\frac{1}{n}\log P_G(\lambda)$. More generally, to bound the normalized partition function at λ_0 it suffices to bound the occupancy fraction α_G for all $0 < \lambda < \lambda_0$. In what follows we will always assume $\lambda > 0$.

We can get a local estimate of α_G by examining (along with the independent set I) a uniformly random vertex v and a random edge e containing v, so that $\alpha_G = \Pr[v \in I]$. Because G is regular and uniform this is equivalent to picking e uniformly and then picking v uniformly from e.

Say a vertex x is *covered* by a vertex y if $y \in I$ and $x \sim y$. Note that any $x \in I$ is uncovered. Call an uncovered vertex which is also unoccupied *available*, and let A be the set of available vertices. Let $N(v)$ denote the neighborhood of v, and let $\hat{N}(v) = N(v) \cup \{v\}$.

Call a vertex *externally uncovered* if it is not covered by any vertex outside of $\hat{N}(v)$, and let C_v be the hypergraph G restricted to v and its externally uncovered neighbors (keeping all partial edges, including those of size 1, so that C_v is still d-regular, but

no longer uniform). Let $\mathcal{C}$ be the collection of all such possible configurations. For each $C \in \mathcal{C}$ write $p(C) = \Pr[C_v = C]$ for the distribution of C_v and let $P_C(\lambda)$ be the partition function for the hypergraph C. This partition function includes the one configuration with $v \in I$ (of weight λ).

We are interested in maximizing

$$\alpha_G = \sum_{C \in \mathcal{C}} p(C) \Pr[v \in I \mid C_v = C] = \sum_{C \in \mathcal{C}} p(C) \frac{\lambda}{P_C(\lambda)}$$

over all hypergraphs G. However, the only terms in this formula which depend at all on the original hypergraph G are the probabilities $p(C)$. Thus it will be useful to know more about which distributions p can actually arise from hypergraphs in this way.

Let $t(e) := |e \cap A|$ be the number of available vertices in e. We also know that

$$\Pr[v \in A \mid t(e) = t] = t/r$$

for each $0 \leq t \leq r$. Conditioning on $C_v = C$, we have

$$\begin{aligned} \frac{t}{r} \Pr[t(e) = t] &= \Pr[v \in A,\ t(e) = t] \\ &= \sum_C p(C) \Pr[v \in A,\ t(e) = t \mid C_v = C] \\ &= \sum_C p(C) \Pr[v \in A \mid C_v = C] \Pr[t(e) = t \mid v \in A,\ C_v = C]. \end{aligned}$$

We can calculate $\Pr[v \in A \mid C_v = C] = 1/P_C(\lambda)$ (since only the empty independent set on C leaves v available) and

$$\Pr[t(e) = t \mid v \in A,\ C_v = C] = \frac{d_t(C)}{d} =: \eta_t(C),$$

where $d_t(C)$ is the number of size-t edges containing v in C (since whenever v is available $t(e) = |e|$ and all d edges containing v are equally likely). Thus the probabilities $p(C)$ must satisfy

$$\sum_C p(C) \frac{\eta_t(C)}{P_C(\lambda)} = \frac{t}{r} \Pr[t(e) = t] = \frac{t}{r} \sum_C p(C) \Pr[t(e) = t \mid C_v = C]$$

giving linear constraints

$$\sum_C p(C) \left(t \Pr[t(e) = t \mid C_v = C] - \frac{r\,\eta_t(C)}{P_C(\lambda)} \right) = 0 \qquad \forall\, 0 \le t \le r.$$

When $t = r$ the constraint holds for any choice of $p(C)$, since $t(e) = r$ precisely when v is available and we pick an edge e of size r in C. It is also trivial for $t = 0$, since $\eta_t = 0$ (every edge containing v has size at least 1).

These linear constraints (along with the constraint that p should be a probability distribution over neighborhood configurations) give a linear program (LP) relaxation for the problem of maximizing the occupancy fraction over all d-regular, r-uniform linear hypergraphs G, and the optimal probability distribuition will give an upper bound on the occupancy fraction of such a graph—if we can solve the LP.

Remark 7.9. This LP relaxation generalizes both the relaxation for independent sets and that for matchings which yield tight bounds in [23], corresponding to the cases $r = 2$ and $d = 2$, respectively. In both prior uses the existence of triangles had no effect on the maximizing configuration. In this case triangles seem to have a more complicated effect on the constraints and we have not been able to show that allowing triangles does not affect the optimum configuration.

7.3.1 LP relaxation for triangle-free hypergraphs

It remains to calculate $\Pr[t(e) = t \mid C_v = C]$, which can be quite complicated. However, the computation is vastly simplified by assuming that the hypergraph is *triangle-free*, i.e., that no edge contains more than one vertex from $N(v)$ unless it also contains v itself. Note that the Hamming hypergraph $H_{[r]^d}$ which was the focus of Ordentlich & Roth's investigation satisfies this property, so restricting to the class of triangle-free hypergraphs still yields interesting results.

The possible neighborhood configurations C in a triangle-free hypergraph are completely parameterized by the number of edges $d_t(C)$ of each size t, as these are the

only nontrivial edges in C_v. For such a neighborhood C,

$$P_C(\lambda) = \lambda + \prod_{s=1}^{r} (1 + (s-1)\lambda)^{d_s(C)}.$$

For $t \neq 0$, one can obtain $t(e) = t$ given C either by picking e to be an edge of size t and taking the empty independent set or by picking an edge of size $t+1$ and covering v by one or more vertices outside that edge (the edge itself must be unoccupied, of course). That is, for $1 \leq t \leq r$

$$\Pr[t(e) = t \mid C_v = C] = \frac{\eta_t(C) + \eta_{t+1}(C)(P_{C_t}(\lambda) - 1)}{P_C(\lambda)},$$

where $P_{C_t}(\lambda)$ is the partition function for C with an edge of size $t+1$ and v removed; this is just a collection of disjoint edges. In particular,

$$P_{C_t}(\lambda) = \frac{P_C(\lambda) - \lambda}{1 + t\lambda}.$$

Finally, we can write a linear program relaxation of our problem with variables $p(C)$:

$$\frac{\alpha^*}{\lambda} = \max \sum_C p(C) \frac{1}{P_C(\lambda)} \qquad \text{subject to} \tag{22}$$

$$\sum_C \frac{p(C)}{P_C(\lambda)} \left(\eta_t(C) + \eta_{t+1}(C) \left(\frac{P_C - \lambda}{1 + t\lambda} - 1 \right) - \frac{r\,\eta_t}{t} \right) = 0 \qquad \forall\, 1 \leq t \leq r-1$$

$$\sum_C p(C) = 1$$

$$p(C) \geq 0 \qquad \forall C \in \mathcal{C}.$$

7.3.2 A proposed primal solution

For $d, r > 2$ the optimal solution for the relaxation does not seem likely to be feasible for the unrelaxed problem and so is probably not a tight bound, but it (or, more generally, *any* feasible solution to the dual LP) would still give an upper bound on the occupancy fraction (and hence the partition function).

Until recently, we believed the following conjecture, at least for the triangle-free case:

Conjecture 7.10. *The optimum of this relaxation is the distribution p^* supported on the neighborhoods I_t with with $\eta_t(I_t) = 1$ (so that all of the edges in I_t have size t).*

There is in fact a unique feasible solution with this support, which is realized in the case $r = 2$ by $K_{d,d}$ and in the case $d = 2$ by the $r \times r$ grid $H_{[r]^2}$. There is also a unique setting of the dual variables Λ^* satisfying complementary slackness for p^*, i.e., such that the constraints corresponding to the configurations I_t hold with equality. However, we have recently found counterexamples to dual feasibility showing that Λ^* is *not* always dual-feasible when $r \geq 8$ and d is sufficiently large. On the other hand, we have not yet found a counterexample for any $r \leq 7$. Indeed, we will see in Section 7.3.6 that Conjecture 7.10 is true for $r = 3$ and in Section 7.3.7 we will give an outline of a proof technique which may work for $4 \leq r \leq 6$.

If we enforce support only on configurations I_s the only nonzero terms in the primal constraint for t are those with $C \in \{I_t, I_{t+1}\}$. Writing $q(s) = p^*(I_s)/P_{I_s}$, the constraint then becomes

$$q(t)(1 - \tfrac{r}{t}) + q(t+1)((1+t\lambda)^{d-1} - 1) = 0 \qquad \forall 1 \leq t \leq r-1.$$

This is effectively a recursion in $q(t)$, along with the "boundary" condition that

$$\sum_{t=1}^{r} p^*(I_t) = \sum_{t=1}^{r} q(t) P_{I_t} = 1.$$

Writing

$$\begin{aligned} v(t) := \frac{q(t)}{q(r)} &= \prod_{i=t}^{r-1} \frac{q(i)}{q(i+1)} \\ &= \prod_{i=t}^{r-1} \Big(\frac{i}{r-i} \Big) ((1+i\lambda)^{d-1} - 1) = \binom{r-1}{t-1} \prod_{i=t}^{r-1} ((1+i\lambda)^{d-1} - 1) \end{aligned}$$

and $Z = \sum_t P_{I_t}(\lambda) v(t)$, the proposed solution to the primal is

$$\begin{aligned} q(t) &= \frac{v(t)}{Z} \\ p^*(I_t) &= P_{I_t}(\lambda) q(t) = \frac{P_{I_t}(\lambda) v(t)}{Z}. \end{aligned}$$

The objective function value at this point is

$$\frac{\alpha^*}{\lambda} = \sum_{t=1}^{r} \frac{p(I_t)}{P_{I_t}} = \sum_t q(t) = \frac{\sum_t v(t)}{Z}. \tag{23}$$

Unfortunately this conjectured bound is complicated to write and its asymptotics are not at all obvious. We will discuss its asymptotics in Section 7.3.8.

7.3.3 The dual LP

The LP dual to (22) is (in variables Λ and Λ_t, $1 \le t \le r-1$)

$$\frac{\alpha^*}{\lambda} = \min \Lambda \qquad \text{subject to}$$
$$\Lambda P_C(\lambda) + \sum_{t=1}^{r-1} \Lambda_t \left(\eta_{t+1}(C) \left(\frac{P_C(\lambda) - \lambda}{1 + t\lambda} - 1 \right) - \eta_t(C) \left(\frac{r}{t} - 1 \right) \right) \ge 1 \quad \forall C \in \mathcal{C}.$$

To show that the primal optimum is supported on the configurations I_s, we show that there is a feasible solution to the dual for which the corresponding constraints are tight. We can solve for candidate values Λ^* and Λ_t^* by setting these r constraints to equality. To simplify notation, we will write

$$Q_t := P_{I_t}(\lambda) - \lambda = (1 + (t-1)\lambda)^d.$$

Since for I_s only the $t = s$ and $t = s-1$ terms in the sum are nonzero, the corresponding constraint becomes

$$\Lambda^* P_{I_s} + \Lambda^*_{s-1} \left(\frac{Q_s}{1 + (s-1)\lambda} - 1 \right) - \Lambda^*_s \left(\frac{r}{s} - 1 \right) = 1, \tag{24}$$

where we take the convention that $\Lambda^*_s = 0$ whenever $s \le 0$ or $s \ge r$. This gives a system of linear equations for the dual variables which clearly has a unique solution.

We can rewrite this (for $0 \le s < r$) as

$$\Lambda^*_s = \left(\frac{s}{r-s} \right) \left[\Lambda^*_{s-1} \left(\frac{Q_t}{1 - (s-1)\lambda} - 1 \right) + \Lambda^*(Q_s + \lambda) - 1 \right].$$

Proposition 7.11. *The solution to the recurrence $a_t = f_t a_{t-1} + g_t$ with $f_t \neq 0$ is*

$$a_t = \left(\prod_{k=1}^{t} f_k \right) \left(a_0 + \sum_{m=1}^{t} \frac{g_m}{\prod_{k=1}^{m} f_k} \right).$$

We can use this to give an explicit formula for the Λ_t^*s, using

$$a_t = \Lambda_t^*, \qquad f_t = \left(\frac{t}{r-t}\right)\left(\frac{Q_t}{1-(t-1)\lambda} - 1\right), \qquad g_t = \left(\frac{t}{r-t}\right)\Lambda^*(Q_t + \lambda).$$

One slight hitch here is that by this definition we have $f_1 = 0$. However, since f_1 is only ever used to multiply by $\Lambda_0 = 0$ we can actually set it to whatever we like. In this case it is easiest to set $f_1 = 1$. Then the formula is

$$\Lambda_t^* = \left(\prod_{k=2}^{t} f_k\right)\left(\sum_{m=1}^{t} \frac{g_m}{\prod_{k=2}^{m} f_k}\right).$$

Furthermore,

$$\prod_{k=2}^{t} f_k = \frac{\prod_{k=1}^{t-1}((1+k\lambda)^{d-1} - 1)}{\binom{r-1}{t}} = \frac{t}{r-t}\left(\frac{v(1)}{v(t)}\right).$$

Plugging this in above gives (for $1 \le t < r$)

$$\begin{aligned} \Lambda_t^* &= \frac{t}{r-t}\sum_{s=1}^{t}\frac{v(s)}{v(t)}(\Lambda^* P_{I_s}(\lambda) - 1) \qquad (25)\\ &= \frac{1}{Z}\frac{t}{r-t}\sum_{s=1}^{t}\frac{v(s)}{v(t)}\sum_{i=1}^{r} v(i)(Q_s - Q_i). \end{aligned}$$

This formula of course fails for $t = r$ (because f_r is undefined), but plugging $\Lambda_r^* = 0$ into (24) for $s = r$ allows us to solve for Λ^* and verify that it is equal to α^*/λ from the primal solution, as expected from complementary slackness.

7.3.4 Dual feasibility

To prove Conjecture 7.10 we must show that the setting of the dual variables Λ^* and Λ_t^* is dual-feasible. In particular, for every neighborhood configuration C, we must show

$$\begin{aligned} \Lambda^* P_C(\lambda) + \sum_t \Lambda_t^* \eta_{t+1}(C)\frac{P_C(\lambda) - \lambda}{1+t\lambda} &\ge 1 + \sum_t \Lambda_t^* \eta_{t+1}(C) + \sum_t \Lambda_t^* \eta_t(C)\left(\frac{r}{t} - 1\right)\\ &= 1 + \sum_t \eta_t(C)\left[\Lambda_{t-1}^* + \Lambda_t^*\left(\frac{r}{t} - 1\right)\right]. \end{aligned}$$

Recalling the notation $Q_t = P_{I_t}(\lambda) - \lambda$, we will also slightly abuse this notation by writing $Q_\eta = P_C(\lambda) - \lambda = \prod_t Q_t^{\eta_t}$ (where $\eta = (\eta_t(C))_{t=1}^r$; we may also use this second formula to define Q_η for *any* $\eta \in \mathbb{R}^r$). Substituting

$$\Lambda^*_{t-1} + \Lambda^*_t \left(\frac{r}{t} - 1\right) = \Lambda^* P_{I_t}(\lambda) + \Lambda^*_{t-1} \frac{Q_t}{1 + (t-1)\lambda} - 1$$

from (24), the constraint for C becomes

$$\begin{aligned}
\Lambda^* P_C(\lambda) + \sum_t \eta_{t+1} \frac{\Lambda^*_t Q_\eta}{1 + t\lambda} &\geq 1 + \sum_t \eta_t \left[\Lambda^* P_{I_t}(\lambda) + \frac{\Lambda^*_{t-1} Q_t}{1 + (t-1)\lambda} - 1\right] \\
\sum_t \eta_t \left[\Lambda^* P_C(\lambda) + \frac{\Lambda^*_{t-1} Q_\eta}{1 + (t-1)\lambda}\right] &\geq \sum_t \eta_t \left[\Lambda^* P_{I_t}(\lambda) + \frac{\Lambda^*_{t-1} Q_t}{1 + (t-1)\lambda}\right] \\
\sum_t \eta_t Q_\eta \left[\Lambda^* + \frac{\Lambda^*_{t-1}}{1 + (t-1)\lambda}\right] &\geq \sum_t \eta_t Q_t \left[\Lambda^* + \frac{\Lambda^*_{t-1}}{1 + (t-1)\lambda}\right].
\end{aligned}$$

Finally, we can write the constraints simply as a slack constraint

$$\sum_t \eta_t (Q_\eta - Q_t) \left(\Lambda^* + \frac{\Lambda^*_{t-1}}{1 + (t-1)\lambda}\right) \geq 0 \tag{26}$$

for all convex combinations η such that $d\eta$ is integral.

Recalling the formulas (23) and (25) for $\Lambda^* = \alpha^*/\lambda$ and Λ^*_t and multiplying through by the nonnegative common denominator Z, the constraint (26) expands to

$$\sum_{t=1}^r \frac{\eta_t (Q_\eta - Q_t)}{1 + (t-1)\lambda} \sum_{i=1}^r v(i) \left(1 + (t-1)\lambda + \left(\frac{t-1}{r-t+1}\right) \sum_{s=1}^{t-1} \frac{v(s)}{v(t-1)} (Q_s - Q_i)\right) \geq 0$$

The left hand side is a polynomial in λ.

7.3.5 The slack constraint

We would like to show that

$$S(\eta) := \sum_t \eta_t c_t (Q_\eta - Q_t) \geq 0$$

for every convex combination η, where

$$c_t := \Lambda^* + \frac{\Lambda^*_{t-1}}{1 + (t-1)\lambda}.$$

Since we have equality (by construction) when η is a basis vector, it suffices to show that the only local minima of $S(\eta)$ on the simplex are at its vertices.

If η is not a vertex of the simplex, then there is a vector u such that the line segment $[\eta - u, \eta + u]$ is contained in the simplex. If η is also a local minimum of S, then for every such u the univariate function $\hat{S}(x) = S(\eta + xu)$ has a local minimum on $[-1, 1]$ at $x = 0$, so we must have $\hat{S}'(0) = 0$ and $\hat{S}''(0) > 0$. Letting $\hat{Q}(x) = Q_{\eta+xu}$,

$$\begin{aligned}
\ln(\hat{Q}(x)) &= \sum_t (\eta_t + u_t x) \ln Q_t \\
\frac{d}{dx}[\ln \hat{Q}(x)] &= \frac{\hat{Q}'(x)}{\hat{Q}(x)} = \sum_t u_t \ln Q_t = \ln Q_u \\
\hat{S}(x) &= \sum_t c_t(\eta_t + u_t x)(\hat{Q}(x) - Q_t) \\
\hat{S}'(x) &= \sum_t c_t \left[u_t(\hat{Q}(x) - Q_t) + (\eta_t + u_t x)\hat{Q}'(x)\right] \\
\hat{S}''(x) &= \sum_t c_t \left[2u_t Q'(x) + (\eta_t + u_t x)Q''(x)\right] \\
&= \sum_t c_t \left[2u_t \hat{Q}(x) \ln Q_u + (\eta_t + u_t x)\hat{Q}(x) \ln^2 Q_u\right] \\
&= \hat{Q}(x) \ln Q_u \sum_t c_t [2u_t + (\eta_t + u_t x) \ln Q_u] \\
&= \hat{Q}(x) \ln Q_u \sum_t c_t [u_t(2 + x \ln Q_u) + \eta_t \ln Q_u].
\end{aligned}$$

So at a non-vertex local minimum we would have

$$\begin{aligned}
0 = \hat{S}'(0) &= \sum_t c_t \left[u_t(Q_\eta - Q_t) + \eta_t Q_\eta \ln Q_u\right] \\
0 \leq \hat{S}''(0) &= \sum_t c_t \left[2u_t Q_\eta \ln Q_u + \eta_t Q_\eta \ln^2 Q_u\right] \\
&= \ln Q_u \sum_t c_t \left[2u_t Q_\eta + u_t(Q_t - Q_\eta)\right] \\
&= \ln Q_u \sum_t c_t u_t (Q_t + Q_\eta).
\end{aligned}$$

That is, at a non-vertex local minimum η,

$$\sum_t c_t u_t (Q_t + Q_\eta) \geq 0 \tag{27}$$

for *every* $u \in \mathbb{R}^r$ with $\sum_t u_t = 0$, $Q_u \geq 1$, and $\text{supp}(u) \subseteq \text{supp}(\eta)$.

In particular, if $i = \max(\text{supp}(\eta))$ and $j = \min(\text{supp}(\eta))$ it suffices to show that the sum in (27) is negative for $u = e_i - e_j$ (where e_k is the kth basis vector), i.e., that

$$c_i(Q_i + Q_\eta) < c_j(Q_j + Q_\eta).$$

Since Q_η can be anything between Q_j and Q_i and the c_ks are (as we will show) decreasing in k, this is the same as showing that

$$c_i(Q_i + Q_j) < 2c_j Q_j$$

whenever $i > j$.

We will show that this is true when $r = 3$. Unfortunately, when $r \geq 4$ it is not, and indeed it does not even seem to be the case that $\hat{S}''(0) - \ln Q_u \hat{S}'(0) > 0$ for some $u = e_i - e_j$.

7.3.6 The case $r = 3$

When $r = 3$, we can write

$$\begin{aligned}
Zc_1 &= (1+\lambda)^{d-1}(1+2\lambda)^{d-1} + (1+2\lambda)^{d-1} - (1+\lambda)^{d-1} \\
Zc_2 &= \frac{3(1+2\lambda)^{d-1} - 1}{2(1+\lambda)} \\
Zc_3 &= \frac{2(1+\lambda)^{d-1} - 1}{1+2\lambda}
\end{aligned}$$

We must show that $c_1 > c_2 > c_3 > 0$ and that

$$c_i(Q_i + Q_j) < 2c_j Q_j$$

whenever $i > j$.

Proposition 7.12. $c_1 > c_2 > c_3$.

Proof. To see the first inequality it suffices to see that

$$\begin{aligned}(1+\lambda)^d((1+2\lambda)^{d-1}-1)+(1+\lambda)(1+2\lambda)^{d-1} &> 2(1+2\lambda)^{d-1}-1\\ &> \frac{3}{2}(1+2\lambda)^{d-1}-\frac{1}{2}.\end{aligned}$$

For the second it suffices to show that

$$3(1+2\lambda)^d-(1+2\lambda) > 4(1+\lambda)^d-2(1+\lambda).$$

Indeed, this is true termwise as polynomials in λ:

$$2+(6d-2)\lambda+\sum_{k=2}^{d}3\binom{d}{k}2^k\lambda^k > 2+(4d-2)\lambda+\sum_{k=2}^{d}4\binom{d}{k}\lambda^k.$$

since $3(2^k)\geq 4$ when $k\geq 2$.

Finally, it is clear from inspection that $c_3 > 0$. □

Proposition 7.13. $c_i(Q_i+Q_j) < 2c_jQ_j$ *whenever* $1\leq j<i\leq 3$.

Proof. For $i=2$ and $j=1$ we must show that

$$(3(1+2\lambda)^{d-1}-1)((1+\lambda)^d+1) < 4(1+\lambda)^d((1+2\lambda)^{d-1}-1)+4(1+\lambda)(1+2\lambda)^{d-1}).$$

This simplifies to showing

$$(1+\lambda)^d(1+2\lambda)^{d-1}+(1+2\lambda)^d+2\lambda(1+2\lambda)^{d-1}-3(1+\lambda)^d+1>0.$$

As before, this is true termwise as a polynomial in λ:

$$\begin{aligned}&(1+\lambda)^d(1+2\lambda)^{d-1}+(1+2\lambda)^d+2\lambda(1+2\lambda)^{d-1}-3(1+\lambda)^d+1\\ &=\lambda(d+2(d-1)+2d-3d)+\sum_{k\geq 2}\lambda^k\left(\textstyle\sum_\ell \binom{d}{k-\ell}\binom{d-1}{\ell}2^\ell+\binom{d}{k}2^k+\binom{d-1}{k-1}2^k-3\binom{d}{k}\right)\\ &>0\end{aligned}$$

since $2^k>3$ when $k\geq 2$.

For $i = 3$ and $j = 1$, we must show that

$$(2(1+\lambda)^{d-1} - 1)((1+2\lambda)^d + 1) < 2((1+\lambda)^{d-1} + 1)(1+2\lambda)^d - 2(1+\lambda)^{d-1}(1+2\lambda).$$

This simplifies to

$$3(1+2\lambda)^d - 4(1+\lambda)^d + 1 > 0,$$

which is once again true termwise:

$$3(1+2\lambda)^d - 4(1+\lambda)^d + 1 = \sum_{k\geq 1} \lambda^k \Big(3\binom{d}{k}2^k - 4\binom{d}{k}\Big) > 0$$

since $3(2^k) > 4$ when $k \geq 1$.

For $i = 3$ and $j = 2$, we must show that

$$(2(1+\lambda)^d - (1+\lambda))((1+2\lambda)^d + (1+\lambda)^d) < (3(1+2\lambda)^d - (1+2\lambda))(1+\lambda)^d.$$

This simplifies to

$$(1+2\lambda)^d - (1+\lambda)^{d-1}\Big(2(1+\lambda)^d - (1+2\lambda)^d + \lambda\Big) > 0,$$

so it suffices to see that

$$2(1+\lambda)^d - (1+2\lambda)^d = \sum_{k\geq 0}(2-2^k)\lambda^k < 1.$$

□

Thus we have proven

Theorem 7.14. *For any d-regular, 3-uniform, linear, triangle-free hypergraph G and for any $\lambda > 0$,*

$$\frac{\alpha_G(\lambda)}{\lambda} \leq \Lambda^* = \frac{(1+2\lambda)^{d-1}((1+\lambda)^{d-1}+1) - (1+\lambda)^{d-1}}{3(1+2\lambda)^{d-1}(\lambda + (1+\lambda)^d) - 3(1+\lambda)^d + 1}.$$

7.3.7 An approach for $r = 4$

We have seen that at a local minimum η of the slack it must be the case that

$$c_i(Q_i + Q_\eta) < c_j(Q_j + Q_\eta) \qquad \forall j < i \in \text{supp}(\eta). \tag{28}$$

In the case $r = 3$ we saw that this was impossible, but when $r \geq 4$ (28) can be true for $j = 1$, so we will have to rule out those values of η separately. From this point the proposed proof method is as follows:

(a) Show that $c_i < c_j$ when $j < i$.

(b) Show that if $\eta_1 \geq 1/2$, taking $u = \eta - e_1$ (i.e., considering the 1-D slice through η and e_1) yields $\hat{S}'(0) < 0$, so no such point can be a local minimum.

(c) Show that for $2 \leq j < i \leq r$

$$c_i Q_i - c_j Q_j < (c_j - c_i) Q_j,$$

so that inequality (28) cannot hold when $\min(\text{supp}(\eta)) = j \geq 2$.

(d) Show that inequality (28) cannot hold when $0 < \eta_1 < 1/2$ by showing that

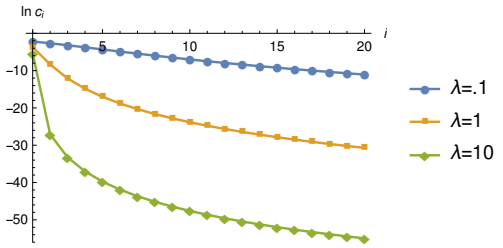
$$c_i Q_i - c_1 \leq (c_1 - c_i)\sqrt{Q_2} < (c_1 - c_i) Q_\eta$$

for all $2 \leq i \leq r$.

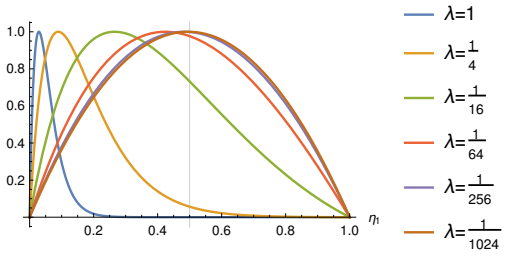
These statements all seem to be true for $r \leq 6$ (see See Figure 16 for some plots for $r = 4$). However this approach is not sustainable either, as (d) fails for $r \geq 7$ and (c) fails for $r \geq 8$ (see Figure 17). It nevertheless seems likely that this approach would succeed in proving the case $r = 4$.

7.3.8 Asymptotics of the occupancy bound

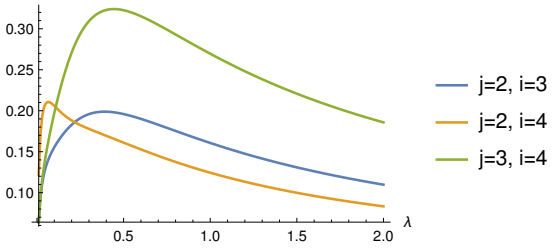
We have just shown that when $r = 3$ the value α^*/λ obtained from the LP above is an upper bound on the occupancy fraction of any d-regular, r-uniform, linear, triangle-free hypergraph. We want to find the asymptotics of α^*/λ and, more importantly,



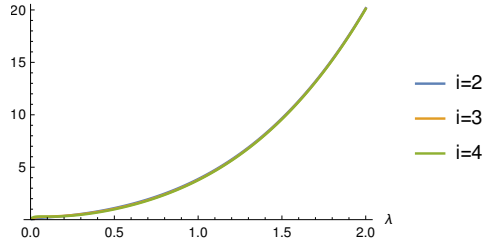
(a) The logs of the coefficients c_i for $r = 20$, $d = 10$ and $\lambda = 0.1, 1, 10$



(b) The slack $S(\eta)$ along the line through e_1 and e_3, normalized for presentability, for $r = 4$ and $d = 50$.

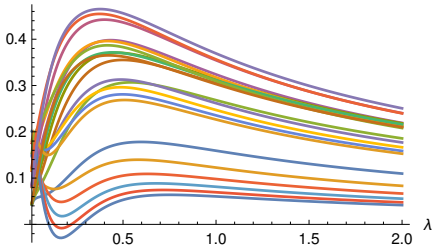


(c) The differences $2c_jQ_j - c_i(Q_i + Q_j)$ for $2 \leq j < i$ as a function of λ for $r = 4$ and $d = 10$

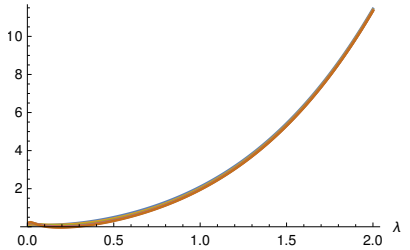


(d) The differences $(c_1 - c_i)\sqrt{Q_2} - (c_iQ_i - c_1)$ for $i \geq 2$ as a function of λ for $r = 4$ and $d = 10$

Figure 16: Plots of the steps in the approach for $r = 4$.



(a) The differences $2c_jQ_j - c_i(Q_i + Q_j)$ for $2 \leq j < i$ as a function of λ for $r = 8$ and $d = 10$



(b) The differences $(c_1 - c_i)\sqrt{Q_2} - (c_iQ_i - c_1)$ for $i \geq 2$ as a function of λ for $r = 7$ and $d = 10$

Figure 17: Failure of the approach outlined above when $r \geq 7$.

the asymptotics of the corresponding bound

$$B^*(\mu) := \int_0^\mu \frac{\alpha^*}{\lambda} d\lambda$$

on the scaled log partition function.

Theorem 7.15. *For any d-regular, 3-uniform, linear, triangle-free hypergraph G on n vertices,*

$$\frac{\ln P_G(\lambda)}{n} \leq \mathcal{B}^*(\lambda) = \frac{\ln(1+\lambda)}{3} + O\left(\frac{1}{d}\right).$$

Proof. The inequality follows from Theorem 7.14 and formula (21) for the occupancy fraction, so it remains to show the asymptotics of the bound. Let us parameterize $\lambda = x/d$, and let

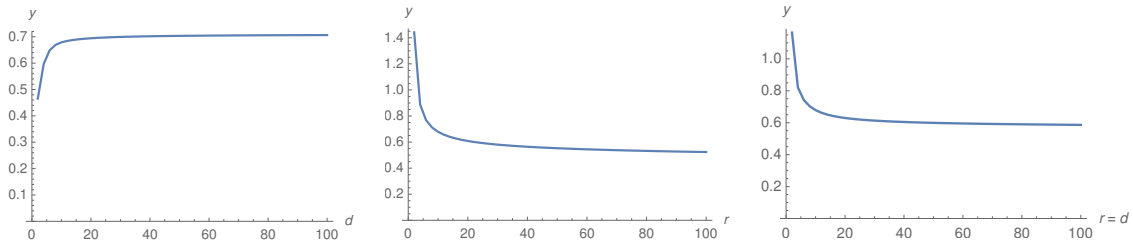
$$\begin{aligned} G(d,\lambda) &= \frac{\alpha^*}{\lambda} - \frac{1}{3(1+\lambda)} \\ &= \frac{3(1+2\lambda)^{d-1} - 1}{3(1+\lambda)\Big(3\big((1+\lambda)^d + \lambda\big)\big((1+2\lambda)^{d-1} - 1\big) + (1+3\lambda)\Big)} \\ &= \frac{3e^{2x} - 1}{3(3e^{3x} - 3e^x + 1)}(1 + o_d(1)), \end{aligned}$$

which approaches 0 exponentially fast as $x \to \infty$. Then

$$\begin{aligned} \int_0^1 \frac{\alpha^*}{\lambda} d\lambda &= \int_0^1 \frac{d\lambda}{3(1+\lambda)} + \int_0^1 G(d,\lambda) d\lambda \\ &= \frac{\ln(2)}{3} + \frac{1}{d}\int_0^d G(d, x/d) dx \\ &\leq \frac{\ln(2)}{3} + \frac{1+o(1)}{d}\int_0^\infty \left(\frac{3e^{2x} - 1}{3(3e^{3x} - 3e^x + 1)}\right) dx \\ &= \frac{\ln(2)}{3} + \frac{O(1)}{d}. \end{aligned}$$

□

The asymptotics in r are harder to pin down, but we conjecture



(a) Varying d at $r = 10$ (b) Varying r at $d = 10$ (c) Varying $d = r$

Figure 18: Plots of $y = \frac{rd}{\ln^2 r}\left(B^*(1) - \frac{\ln 2}{r}\right)$ showing that the limits seem to converge. The actual exponent on $\ln r$ in the bound may be less than 2, but it is certainly not always 1 as in Conjecture 7.3.

Conjecture 7.16. *The bound $B^*(\lambda)$ on the normalized log partition function obtained from Conjecture 7.10 satisfies*

$$\frac{\ln P_G(\lambda)}{n} \leq B^*(\lambda) = \frac{\ln(1+\lambda)}{r} + O\left(\frac{\log^2 r}{rd}\right).$$

Of course this cannot be the correct asymptotics in terms of both r and d, since when $d = 2$ we know that $\log r$ is correct rather than $\log^2 r$ (since the method of [23] shows that the bound B^* is both true and tight when $d = 2$). However, for large d the squared logarithm does seem to be correct asymptotics for this bound. See Figure 18. On the other hand, the asymptotics of B^* in r seem to be rather a moot point, since Conjecture 7.10 does not hold for large r.

The discovery that Conjecture 7.10 is not always true does, however, raise the interesting question of what the actual optimum to our LP relaxation is. In particular, it may in fact be the case that when both r and d are large the maximizing hypergraphs look very different from either of the known optima for small r or d.

REFERENCES

[1] Aharoni, R. and Kessler, O., "On a possible extension of Hall's theorem to bipartite hypergraphs," *Discrete Math.*, vol. 84, no. 3, pp. 309–313, 1990.

[2] Aldous, D., "Mixing times for the branch-rotation chain on cladograms (or the triangulation walk)."

[3] Aldous, D., "Random walks on finite groups and rapidly mixing Markov chains," in *Séminaire de Probabilités*, vol. 17, pp. 243–297, Springer, 1983.

[4] Aldous, D., "Triangulating the circle, at random," *Am. Math. Mon.*, vol. 101, no. 3, pp. 223–233, 1994.

[5] Alon, N. and Friedland, S., "The maximum number of perfect matchings in graphs with a given degree sequence," *Elec. J. Comb.*, vol. 15, 2008.

[6] Ardila, F., "The Catalan matroid," *J. Comb. Theory, Ser. A*, vol. 104, pp. 49–62, Oct. 2003.

[7] Beezer, R. A., "The number of subgraphs of a regular graph," 1994.

[8] Beezer, R. A. and Farrell, E. E. J., "The matching polynomial of a regular graph," *Discrete Math.*, vol. 137, pp. 7–18, Jan. 1995.

[9] Bender, E. A. and Goldman, J. R., "On the applications of Möbius inversion in combinatorial analysis," *Am. Math. Mon.*, vol. 82, no. 8, pp. 789–803, 2012.

[10] Bern, M. and Eppstein, D., "Mesh generation and optimal triangulation," *Computing in Euclidean geometry*, vol. 4, pp. 47–123, 1995.

[11] Bernardi, O. and Bonichon, N., "Intervals in Catalan lattices and realizers of triangulations," *J. Comb. Theory. Ser. A*, vol. 116, no. 1, pp. 55–75, 2009.

[12] Billoire, A. and David, F., "Scaling properties of randomly triangulated planar random surfaces: A numerical study," *Nuc. Phys. B*, vol. 275, no. 4, pp. 617–640, 1986.

[13] Bollobás, B., "The evolution of random graphs," *Transactions of the American Mathematical Society*, vol. 286, no. 1, pp. 257–274, 1984.

[14] Bonin, J. E. and de Mier, A., "Lattice path matroids: structural properties," *Eur. J. Comb.*, vol. 27, pp. 701–738, Jul. 2006.

[15] Bregman, L. M., "Some properties of nonnegative matrices and their permanents," in *Soviet Mathematics Doklady*, vol. 14, pp. 945–949, 1973.

[16] Bubley, R. and Dyer, M., "Path coupling: A technique for proving rapid mixing in Markov chains," in *Foundations of Computer Science*, vol. 38, pp. 223–231, IEEE, 1997.

[17] Caputo, P., Restrepo, R., and Tetali, P., "Random transpositions on cyclic permutations of lattice paths," *Unpubl. notes*, no. 1, pp. 1–3, 2010.

[18] Carroll, T., Galvin, D., and Tetali, P., "Matchings and independent sets of a fixed size in regular graphs," *JCT, Ser. A*, vol. 116, pp. 1219–1227, Oct. 2009.

[19] Choe, Y. and Wagner, D. G., "Rayleigh matroids," *Comb. Probab. Comput.*, vol. 15, p. 765, Jul. 2006.

[20] Chung, F. R. K., Graham, R. L., Frankl, P., and Shearer, J. B., "Some intersection theorems for ordered sets and graphs," *J. Comb. Theory, Ser. A*, vol. 43, pp. 23–37, Sep. 1986.

[21] Csikvári, P., "Lower matching conjecture, and a new proof of Schrijver's and Gurvits's theorems," *arXiv preprint arXiv:1406.0766*, 2014.

[22] Cutler, J. and Radcliffe, A., "An entropy proof of the Kahn-Lovasz theorem," *Elec. J. Comb.*, vol. 18, pp. 1–8, 2011.

[23] Davies, E., Jenssen, M., Perkins, W., and Roberts, B., "Independent sets, matchings, and occupancy fractions," *arXiv Prepr.*, pp. 1–18, 2015.

[24] Dembo, A., Montanari, A., and Sun, N., "Factor models on locally tree-like graphs," *Ann. Probab.*, vol. 41, no. 6, pp. 4162–4213, 2013.

[25] Diaconis, P. and Saloff-Coste, L., "Comparison techniques for random walks on finite groups," *Ann. Probab.*, vol. 21, no. 4, pp. 2131–2156, 1993.

[26] Diaconis, P. and Saloff-Coste, L., "Logarithmic Sobolev inequalities for finite Markov chains," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 695–750, 1996.

[27] Diaconis, P. and Shahshahani, M., "Time to reach stationarity in the Bernoulli-Laplace diffusion model," *SIAM J. Math. Anal.*, vol. 18, no. 1, pp. 208–218, 1987.

[28] Ding, J. and Mossel, E., "Mixing under monotone censoring," *Electron. Commun. Probab.*, vol. 19, no. 46, pp. 1–6, 2014.

[29] Feder, T. and Mihail, M., "Balanced matroids," in *ACM Symp. Theory Comput.*, (New York, New York, USA), pp. 26–38, ACM Press, 1992.

[30] Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J., "Correlation inequalities on some partially ordered sets," *Comm. Math. Phys.*, vol. 22, no. 2, pp. 89–103, 1971.

[31] Friedland, S., Krop, E., and Markström, K., "On the number of matchings in regular graphs," *Elec. J. Comb.*, vol. 15, pp. 1–28, 2008.

[32] Galvin, D., "Bregman's theorem and extensions," 2011.

[33] Galvin, D., "Maximizing H-colorings of a regular graph," *J. Graph Theory*, vol. 73, no. 1, pp. 66–84, 2013.

[34] Galvin, D., "Three tutorial lectures on entropy and counting," in *Lake Michigan Work. Comb. Graph Theory*, 2014.

[35] Galvin, D. and Tetali, P., "On weighted graph homomorphisms," *DIMACS*, vol. 63, pp. 97–104, 2004.

[36] Glauber, R. J., "Time-dependent statistics of the Ising model," *J. Math. Phys.*, vol. 4, no. 2, pp. 294–307, 1963.

[37] Goldreich, O., Goldwasser, S., Lehman, E., Ron, D., and Samorodnitsky, A., "Testing monotonicity," *Combinatorica*, vol. 20, pp. 301–337, Mar. 2000.

[38] Gurvits, L., "Unleashing the power of Schrijver's permanental inequality with the help of the Bethe approximation," *arXiv preprint arXiv:1106.2844*, 2011.

[39] Heilmann, O. J. and Lieb, E. H., *Theory of Monomer-Dimer Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[40] Heitsch, C. E., "Combinatorics on plane trees, motivated by RNA secondary structure configurations."

[41] Heitsch, C. E. and Tetali, P., "Meander graphs," No. 01, pp. 469–480, 2011.

[42] Ilinca, L. and Kahn, J., "Asymptotics of the upper matching conjecture," *JCT Ser. A*, vol. 120, pp. 976–983, Jul. 2013.

[43] Jerrum, M. and Son, J.-B., "Spectral gap and log-Sobolev constant for balanced matroids," in *IEEE Symp. Found. Comput. Sci.*, vol. 14036, pp. 721–729, IEEE Comput. Soc, 2002.

[44] Jerrum, M., Son, J.-B., Tetali, P., and Vigoda, E., "Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains," *Ann. Appl. Probab.*, vol. 14, pp. 1741–1765, Nov. 2004.

[45] Kahn, J., "An entropy approach to the hard-core model on bipartite graphs," *Comb. Probab. Comput.*, vol. 10, pp. 219–237, Jun. 2001.

[46] Keough, L., *Extremal Results for the Number of Matchings and Independent Sets*. PhD thesis, 2015.

[47] Knuth, D. E., "The art of computer programming. generating all trees," 2006.

[48] Lawler, G. F. and Sokal, A. D., "Bounds on the l^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality," *Trans. Am. Math. Soc.*, vol. 309, no. 2, p. 557, 1988.

[49] Levin, D. A., Peres, Y., and Wilmer, E. L., *Markov chains and mixing times*. Am. Math. Soc., 2009.

[50] Lovász, L., *Large networks and graph limits*, vol. 60. American Mathematical Soc., 2012.

[51] Madiman, M. and Tetali, P., "Information inequalities for joint distributions, with interpretations and applications," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2699–2713, Jun. 2010.

[52] Madras, N. and Randall, D., "Markov chain decomposition for convergence rate analysis," *Ann. Appl. Probab.*, vol. 12, no. 2, pp. 581–606, 2002.

[53] Mathieu, P., "Log-Sobolev and spectral gap inequalities for the knapsack Markov chain," *Markov Process. Relat. Fields*, vol. 8, pp. 595–610, 2002.

[54] McEliece, R., *The theory of information and coding*. Cambridge University Press, 2002.

[55] McShine, L. and Tetali, P., "On the mixing time of the triangulation walk and other Catalan structures," *DIMACS Ser. Discret. Math. Theor. Comput. Sci.*, vol. 43, pp. 147–160, 1999.

[56] Molloy, M., Reed, B., and Steiger, W., "On the mixing rate of the triangulation walk," *DIMACS Ser. Discret. Math. Theor. Comput. Sci.*, vol. 43, pp. 179–190, 1999.

[57] Montenegro, R. and Tetali, P., "Mathematical aspects of mixing times in Markov chains," *Found. Trends Theor. Comput. Sci.*, vol. 1, no. 3, pp. 237–354, 2005.

[58] Morris, B. and Sinclair, A., "Random walks on truncated cubes and sampling 0-1 knapsack solutions," *SIAM J. Comp.*, vol. 34, no. 1, pp. 195–226, 2004.

[59] Narayana, T. V., "Cyclic permutation of lattice paths and the Chung-Feller theorem," *Scandinavian Actuarial Journal*, vol. 1967, no. 1-2, pp. 23–30, 1967.

[60] Ordentlich, E. and Roth, R. M., "Independent sets in regular hypergraphs and multidimensional runlength-limited constraints," *SIAM J. Discret. Math.*, vol. 17, no. 4, pp. 615–623, 2004.

[61] Radhakrishnan, J., "An entropy proof of Bregman's theorem," *JCT, Ser. A*, vol. 77, pp. 161–164, Jan. 1997.

[62] Radhakrishnan, J., "Entropy and counting," *Comput. Math. Model. Algorithms*, pp. 146–168, 2003.

[63] Randall, D. and Tetali, P., "Analyzing Glauber dynamics by comparison of Markov chains," *J. Math. Phys.*, vol. 41, no. 3, pp. 1598–1615, 2000.

[64] Sernau, L., "Graph operations and upper bounds on graph homomorphism counts," pp. 1–22, 2015.

[65] Shabanov, D., "On homomorphisms in hypergraphs," *Unpubl. notes*, no. 1, pp. 1–3, 2013.

[66] Sinclair, A. and Jerrum, M., "Approximate counting, uniform generation and rapidly mixing Markov chains," *Information and Computation*, vol. 82, no. 1, pp. 93–133, 1989.

[67] Sleator, D. D., Tarjan, R. E., and Thurston, W. P., "Rotation distance, triangulations, and hyperbolic geometry," *J. Am. Math. Soc.*, vol. 1, no. 3, pp. 647–647, 1988.

[68] Sohoni, M., "Rapid mixing of some linear matroids and other combinatorial objects," *Graphs Comb.*, vol. 15, pp. 93–107, Mar. 1999.

[69] Stanley, R. P., *Catalan Numbers*. Cambridge University Press, 2015.

[70] Wanless, I. M., "Counting matchings and tree-like walks in regular graphs," *Comb. Probab. Comput.*, vol. 19, no. 3, pp. 463–480, 2010.

[71] Wilson, D. B., "Mixing times of lozenge tiling and card shuffling in Markov chains," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 274–325, 2004.

[72] Zhao, Y., "The number of independent sets in a regular graph," *Comb. Probab. Comput.*, vol. 19, pp. 315–320, Nov. 2009.

[73] Zhao, Y., "The bipartite swapping trick on graph homomorphisms," *SIAM J. Discrete Math.*, vol. 25, no. 2, p. 660, 2011.