

GIBBS MEASURES IN STATISTICAL PHYSICS AND COMBINATORICS (DRAFT)

WILL PERKINS

ABSTRACT. These are notes for lectures on Gibbs measures in statistical physics and combinatorics presented in Athens, Greece, May 2017, as part of the ‘Techniques in Random Discrete Structures’ summer school.

CONTENTS

1. Introduction	2
2. Gibbs measures and phase transitions	2
2.1. The hard sphere model	2
2.2. The hard-core model	5
2.3. The monomer-dimer model	6
2.4. The Potts model	8
2.5. Gibbs measures	8
Further reading	8
Exercises	8
3. Gibbs measures in extremal combinatorics	9
3.1. Graph polynomials and partition functions	9
3.2. Extremal results	10
3.3. Extensions and reductions	13
Further reading	14
Exercises	15
4. Occupancy fractions and optimization	15
4.1. The occupancy fraction of the hard-core model	16
4.2. The general method	20
4.3. Graph convergence and optimization	21
4.4. Matchings	21
4.5. Independent sets in bipartite vertex-transitive graphs	22

Date: July 17, 2017.

WP supported in part by EPSRC grant EP/P009913/1. The summer school was organized by Agelos Georgakopoulos and Dimitris Cheliotis and sponsored by the ERC project *Random Graph Geometry and Convergence*, PI Agelos Georgakopoulos.

Exercises	23
5. Lower bounds, Ramsey theory, and sphere packings	24
5.1. Ramsey theory: $R(3, k)$	24
5.2. Sphere packing densities	28
Further reading	32
Exercises	32
6. Generalizations and extensions	32
6.1. Adding local constraints	32
6.2. Colorings	34
Exercises	36
7. Further directions and open problems	36
7.1. Independent sets and matchings of a given size	36
7.2. General results and conjectures on graph homomorphisms	38
Exercises	38
Acknowledgements	38
References	38
Appendix A. Notation	40
Appendix B. Background material	41
B.1. Statistical physics and probability	41
B.2. Entropy	41
B.3. Linear programming	42
Appendix C. An alternative proof of Theorem 12	42

1. INTRODUCTION

2. GIBBS MEASURES AND PHASE TRANSITIONS

2.1. The hard sphere model. Consider the most basic model of a gas or a fluid: particles represented by a random configuration of identical spheres in a large container, subject to no forces between them except for the hard constraint that no two spheres can overlap. This is the *hard sphere* model from statistical physics. If we imagine the container is a large box (or torus) in \mathbb{R}^d of volume n and each sphere is of volume 1, then as we vary the number of spheres N , the occupation density, $\alpha = N/n$, varies from 0 to the maximum sphere packing density in d -dimensions, call this $\theta(d)$.

Let S be a bounded measurable region in \mathbb{R}^d . Let r_d be the radius of the ball of volume 1 in \mathbb{R}^d .

Definition. *The **hard sphere model** on S with k particles is a collection \mathbf{X} of k unordered points in S , uniformly distributed conditioned on the event that all pairs of points are at distance at least $2r_d$.*

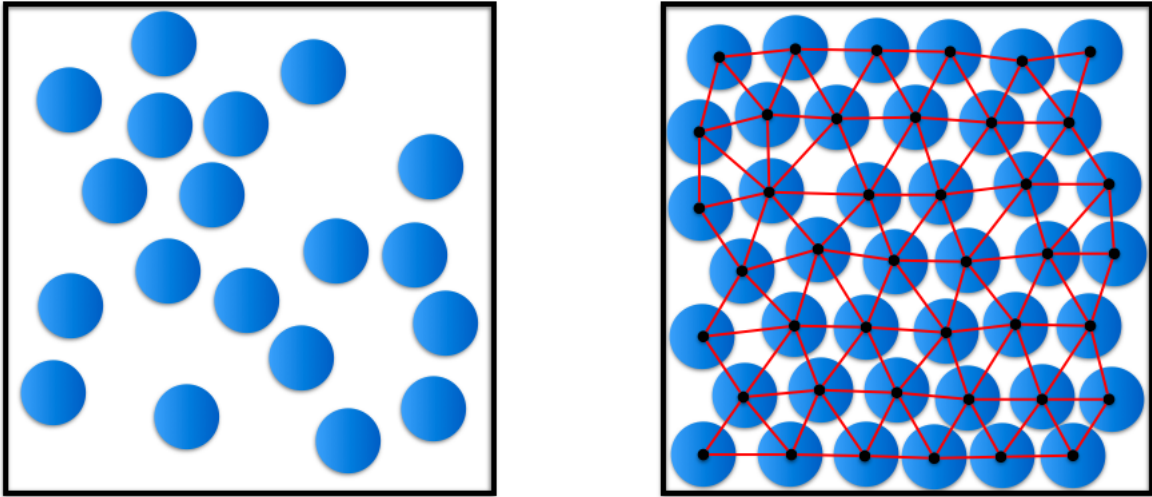


FIGURE 1. The hard sphere model at low and high density

We call this model the *canonical ensemble* to distinguish it from a slightly different model introduced below.

The points \mathbf{X} represent the centers of a sphere packing of spheres of volume 1. Note that we don't require that the entire sphere around a center fit in the set S - we can have centers arbitrarily close to the boundary of S .

Suppose S is a large box of volume n in \mathbb{R}^d . What do we expect to see in a typical configuration of spheres at density α , that is $k \sim \alpha n$? If α is small we'd expect to see *disorder* - the spheres look randomly jumbled, with no long-range order or organization. If α is large, near the maximum sphere packing fraction $\theta(d)$, we might expect a typical configuration from the model to look lattice-like and exhibit long range correlations. In fact this is exactly what physicists have observed, through both computer experiments as well as actual physical experiments with hard-sphere-like particles called colloids. Such a dramatic shift in macroscopic properties on varying a parameter is called a *phase transition*. The phase transition in the hard sphere model represents the transition of a gas into a solid or crystal.

Unfortunately, despite many years of study, it has not been proved mathematically that such a phase transition occurs in this simple model.

What is the mathematical definition of a phase transition? We will present several equivalent definitions later, but the most relevant definition for this course involves the *partition function* of the model.

For $x_1, \dots, x_k \in \mathbb{R}^d$, let $\mathcal{D}(x_1, \dots, x_k)$ be the event that $d(x_i, x_j) > 2r_d$ for all $i \neq j \in [k]$.

Definition. The partition function of the canonical hard sphere model on a bounded, measurable $S \subset \mathbb{R}^d$ with k spheres is

$$\hat{Z}_S(k) = \frac{1}{k!} \int_{S^k} \mathbf{1}_{\mathcal{D}(x_1, \dots, x_k)} dx_1 \cdots dx_k$$

The probability that k uniformly chosen unordered points in S form a packing of balls of volume 1 is simply $\hat{Z}_S(k)/\text{vol}(S)^k$.

Let $C_k(S)$ be the set of configurations of unordered k -tuples of points from a bounded, measurable region $S \subset \mathbb{R}^d$. The probability that the centers \mathbf{X}_k of the hard sphere model on S at fugacity λ belong to A_k is

$$\mathbb{P}[\mathbf{X}_k \in A_k] = \frac{\frac{1}{k!} \int_{S^k} \mathbf{1}_{\{\{x_1, \dots, x_k\} \in A_k\}} \cdot \mathbf{1}_{\mathcal{D}(x_1, \dots, x_k)} dx_1 \cdots dx_k}{\hat{Z}_S(k)},$$

and so the partition function acts as the normalizing constant of the probability distribution. Despite this innocuous sounding role, partition functions will be the central object of interest in this course. To give one example of why they play such an important role, let's see how they are connected with the idea of a phase transition in the hard sphere model.

Let $B_n = B_n(d)$ be the box of volume n centered at the origin in \mathbb{R}^d . For $\alpha \in (0, \theta(d))$, let

$$\hat{f}_{\mathbb{R}^d}(\alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \hat{Z}_{B_n}(\lfloor \alpha n \rfloor).$$

This is the *free energy* of the hard sphere model at density α . The fact that this limit exists follows from some general theory (subadditivity) but for now take it for granted. (Note that up to a constant this is the large deviation rate function of the event that αn random points placed in B_n are at pairwise distance at least $2r_d$). The limit is robust in the shape of the growing region - we could have used a ball of volume n or some other shape instead.

The free energy in fact encompasses many of the interesting properties of the model. For instance, it provides our first definition of a phase transition.

Definition. *The canonical hard sphere model on \mathbb{R}^d undergoes a phase transition at density α^* if the function $\hat{f}_{\mathbb{R}^d}(\alpha)$ is non-analytic at α^* ; that is, either $\hat{f}_{\mathbb{R}^d}$ or one of its higher derivatives is discontinuous at α^* .*

Note that a *phase transition* is only a property of a space of infinite volume like \mathbb{R}^d or later an infinite graph like \mathbb{Z}^d . For a finite region like B_n , the partition function is a polynomial (and thus analytic) and so there can be no 'finite volume' phase transition.

The grand canonical ensemble. It will often be more convenient to work with a slightly different hard sphere model, one in which the number of spheres itself is a random variable. In statistical physics, the first model we defined is the canonical ensemble, and the next is the grand canonical ensemble.

Definition. *The grand canonical hard sphere model on a bounded measurable $S \subset \mathbb{R}^d$ at fugacity λ is a set of centers \mathbf{X} distributed according to a Poisson point process with intensity λ conditioned on the event that all centers are pairwise distance at least $2r_d$.*

Definition. *The partition function of the grand canonical hard sphere model is*

$$\begin{aligned} Z_S(\lambda) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \int_{S^k} \mathbf{1}_{\mathcal{D}(x_1, \dots, x_k)} dx_1 \cdots dx_k \\ &= \sum_{k=0}^{\infty} \lambda^k \cdot \hat{Z}_S(k). \end{aligned}$$

We can again define the free energy of the hard sphere model:

$$f_{\mathbb{R}^d}(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_{B_n}(\lambda).$$

Definition. *The grand canonical hard sphere model on \mathbb{R}^d undergoes a phase transition at fugacity λ^* if the function $f_{\mathbb{R}^d}(\lambda)$ is non-analytic at λ^* ; that is, either $f_{\mathbb{R}^d}$ or one of its higher derivatives is discontinuous at λ^* .*

Open Problem. *Prove the hard sphere model exhibits a phase transition in some dimension d . That is, prove that the limiting free energy $f_{\mathbb{R}^d}(\lambda)$ (or $\hat{f}_{\mathbb{R}^d}(\alpha)$) has a non-analytic point.*

2.2. The hard-core model. While the hard sphere model presents fascinating mathematical challenges, we do not have a rigorous answer to its most pressing question, whether or not it can help explain the universal phenomenon of freezing and crystallization. Is there any way for us to simplify the model in some way that allows us to analyze the phase transition?

Happily, the answer is yes, that by discretizing the model in a specific way, we obtain a model that does indeed provably exhibit a crystallization phase transition.

Definition. *An independent set $I \subseteq V(G)$ is a set of vertices of a graph G that induce no edges. That is $(u, v) \notin E(G)$ for all $u, v \in I$.*

The hard-core model is a probability distribution over the independent sets $\mathcal{I}(G)$ of a graph G .

Definition. *Let G be a finite graph and $\lambda > 0$. The **hard-core model** on G at fugacity λ is a random independent set \mathbf{I} drawn from $\mathcal{I}(G)$ according to the distribution*

$$\Pr[\mathbf{I} = I] = \frac{\lambda^{|I|}}{Z_G(\lambda)},$$

where

$$Z_G(\lambda) = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|}$$

is the **partition function** of the hard-core model.

Let $\Lambda_{d,n}$ be the $n^{1/d} \times \cdots \times n^{1/d}$; that is the d -dimensional square grid with n vertices.

We can define the free energy of the hard-core model on \mathbb{Z}^d to be:

$$f_{\mathbb{Z}^d}(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_{\Lambda_{d,n}}(\lambda).$$

The hard-core model on \mathbb{Z}^d exhibits a phase transition at λ^* if $f_{\mathbb{Z}^d}(\lambda)$ is non-analytic there.

Let v_0 be the center of the box $\Lambda_{d,n}$. Let $\Lambda_{d,n}^{\text{even}}$ be $\Lambda_{d,n}$ with all the even sites on the boundary occupied, and likewise with $\Lambda_{d,n}^{\text{odd}}$.

We say the hard-core model on \mathbb{Z}^d at fugacity λ is in the *uniqueness* regime if

$$\lim_{n \rightarrow \infty} \Pr_{\Lambda_{d,n}^{\text{even}}, \lambda} [v_0 \in I] = \lim_{n \rightarrow \infty} \Pr_{\Lambda_{d,n}^{\text{odd}}, \lambda} [v_0 \in I].$$

In other words, there is a unique *infinite volume hard-core distribution* at fugacity λ . A phase transition occurs at λ^* if the model switches from uniqueness to non-uniqueness (or vice versa).

We can also define the phase transition in terms of decay of correlations.

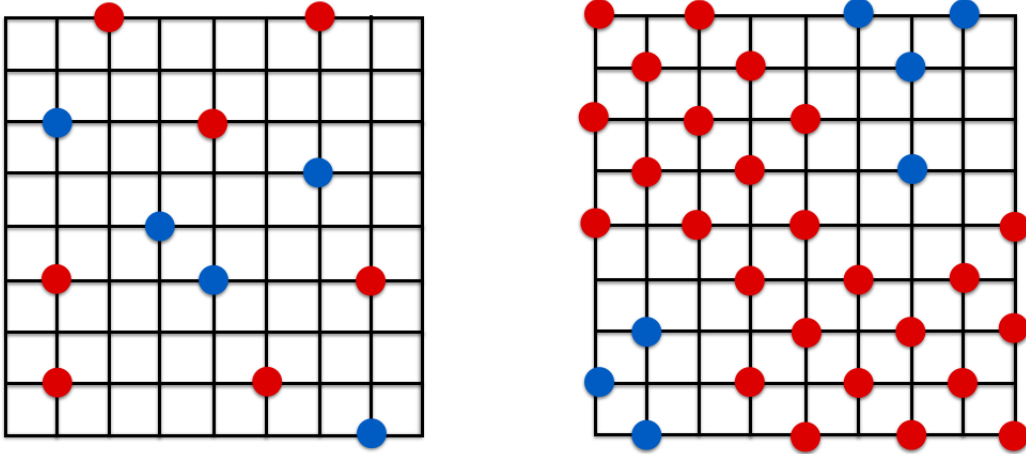


FIGURE 2. The hard-core model at low and high density

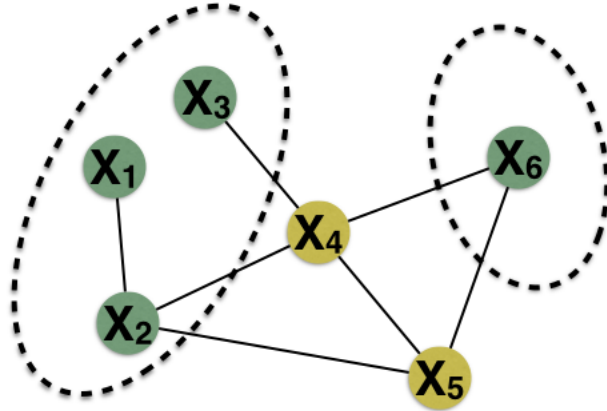


FIGURE 3. The spatial Markov property

The spatial Markov property. For a set of vertices $U \subseteq V(G)$, the neighborhood of U is $N(U) = \{x : x \notin U, x \sim v \text{ for some } v \in U\}$. Then for any subset $s_U \subseteq U$,

$$\Pr[\mathbf{I} \cap U = s_U | \mathbf{I} \cap U^c] = \Pr[\mathbf{I} \cap U = s_U | \mathbf{I} \cap N(U)].$$

If we condition on the ‘spins’ (occupied / unoccupied) of the boundary of a set U , then what happens on U and outside the boundary are independent.

Open Problem. Let $\lambda_c(d)$ be the smallest λ at which $f_{\mathbb{Z}^d}(\lambda)$ is not analytic. Determine the asymptotics of $\lambda_c(d)$ as $d \rightarrow \infty$. It is known that $\lambda_c(d) = \Omega(1/d)$ [54] and $\lambda_c(d) = O(\log^2 d/d^{1/3})$ [41].

2.3. The monomer-dimer model.

Definition. A matching M in a graph G is a collection of edges which share no endpoints.

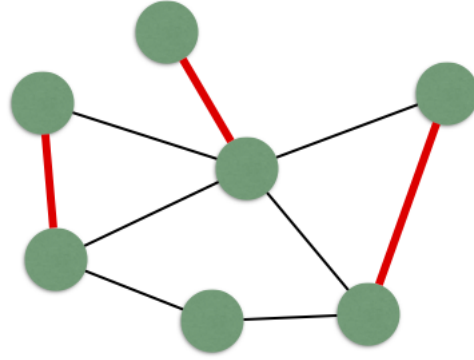


FIGURE 4. The monomer-dimer model

We consider the empty set of edges to be a matching. The size of a matching M , $|M|$, is its number of edges. A perfect matching is a matching that saturates all vertices, that is $|M| = |V(G)|/2$.

Let $\mathcal{M}(G)$ be the set of all matchings of G (we consider the empty set of edges to be a matching). The *monomer-dimer model* is a random matching \mathbf{M} drawn from $\mathcal{M}(G)$ according to the probability distribution

$$\Pr[\mathbf{M} = M] = \frac{\lambda^{|M|}}{Z_G^{\text{match}}(\lambda)},$$

where $|M|$ is the number of edges in the matching M and the partition function is

$$Z_G^{\text{match}}(\lambda) = \sum_{M \in \mathcal{M}(G)} \lambda^{|M|}.$$

There are many similarities between the hard-core model and the monomer-dimer model; in fact, the monomer-dimer model on G is the hard-core model on the *line graph* $L(G)$, the graph with vertex set $E(G)$ and edges between edges of G that share a common vertex.

In some important ways, however, the monomer-dimer model is very different than the hard-core model. It is a deep theorem of Heilmann and Lieb [32] that the monomer-dimer model does not exhibit a phase transition on *any* lattice. The proof uses the Lee-Yang theory of zeros of partition functions and phase transitions and the result follows from the fact that all of the roots of the equation

$$Z_G^{\text{match}}(\lambda) = 0$$

are real numbers for any graph G . Or in other words, the hard-core partition function $Z_G(\lambda)$ has only real roots as long as G is the line graph of some other graph H . This result was generalized by Chudnovsky and Seymour [10] who showed that $Z_G(\lambda)$ has only real roots for any *claw-free* graph G ; that is, any graph G that avoids an induced star with three leaves.

Open Problem. Consider infinite, vertex-transitive graphs. We know that bipartite graphs with sufficient expansion (eg. \mathbb{Z}^d) exhibit a hard-core phase transition. We know that line graphs (and claw-free graphs) do not (Heilmann-Lieb theorem, Chudnovsky Seymour). What about all the graphs in between? That is, graphs that have claws but whose number of ground states is infinite.

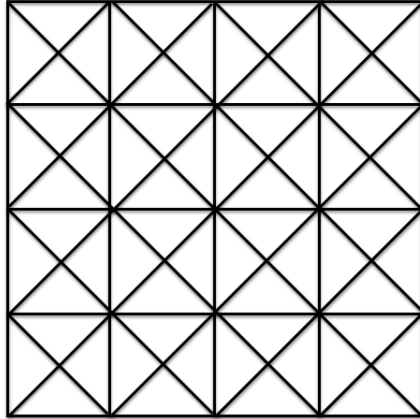


FIGURE 5. The lattice for the discrete hard square model

2.4. The Potts model. The Potts model was first studied by Renfrey Potts [44] in his PhD thesis, directed by his advisor Cyril Domb. It is a q -spin generalization of the 2-spin Ising model [33] and is an idealization of a magnetic material.

For a graph G and an assignment χ of q -colors to each vertex of G , let $m(G, \chi)$ denote the number of monochromatic edges of G under χ . The Potts model at inverse temperature β is a random q -coloring χ of G chosen according to the distribution

$$\Pr_{G,\beta}[\chi = \chi] = \frac{e^{-\beta m(G,\chi)}}{Z_G^q(\beta)},$$

where

$$Z_G^q(\beta) = \sum_{\chi: V(G) \rightarrow [q]} e^{-\beta m(G,\chi)}$$

is the q -color Potts model partition function.

When $\beta > 0$ the model is *anti-ferromagnetic*: colorings with few monochromatic edges are preferred. When $\beta < 0$, the model is *ferromagnetic*: colorings with many monochromatic edges are preferred.

2.5. Gibbs measures.

Further reading. For an introduction to the hard sphere model by a physicist aimed at mathematicians, see Harmut Löwen's survey [40]

For more on phase transition in systems with hard constraints and their relation to combinatorics, see Brightwell and Winkler [8]

For background on the Potts model and its partition function see the surveys of Wu [59] and Sokal [51], and for applications in combinatorics see the survey of Welsh and Morino [58].

Exercises.

- (1) Show that the grand canonical hard sphere partition function on any bounded measurable S satisfies $Z_S(\lambda) \leq e^{\lambda \cdot \text{vol}(S)}$.

- (2) Compute the hard-core partition function $Z_{S_k}(\lambda)$ for the graph S_k , the star on k leaves. Compute the probability that the center of the star is in the random independent set \mathbf{I} .
- (3) Compute the hard-core partition function $Z_{K_{d,d}}(\lambda)$ where $K_{d,d}$ is the complete d -regular bipartite graph on $2d$ vertices.
- (4) Compute the hard-core partition function $Z_G(\lambda)$ of the cycles C_3, C_4, C_5 .
- (5) Compute the hard-core partition function of the cycle C_n .
- (6) Prove that the 1-dimensional hard-core model on \mathbb{Z} does not exhibit a phase transition, by computing the limiting free energy and showing that it is a real analytic function of λ .
- (7) Suppose G is the disjoint union of two graphs H_1 and H_2 . Show that $Z_G(\lambda) = Z_{H_1}(\lambda) \cdot Z_{H_2}(\lambda)$.
- (8) Suppose v is a vertex in some graph with no edges in its neighborhood. Call a vertex u *uncovered* with respect to an independent set I if $N(u) \cap I = \emptyset$. Consider the hard-core model on G at fugacity λ and calculate the probability that v is occupied given that v has exactly j uncovered neighbors.
- (9) Describe the line graph of the infinite graph \mathbb{Z}^d .
- (10) Construct an infinite graph that is claw-free but not the line graph of any graph.
- (11) Prove that the following is an alternative description of the hard-core model on G . Let \mathbf{I} be a random subset of $V(G)$, each vertex included independently with probability $\frac{\lambda}{1+\lambda}$, conditioned on the event that the vertices form an independent set.

3. GIBBS MEASURES IN EXTREMAL COMBINATORICS

While Gibbs measures were originally developed in statistical physics, their mathematical properties have proved tremendously useful across many different scientific disciplines, including statistics, computer science, and machine learning where they appear under various names including Markov random fields, probabilistic graphical models, Boltzmann distributions, and log-linear models.

The majority of this course will focus on how Gibbs measures can be used to prove results in extremal graph theory.

3.1. Graph polynomials and partition functions. We begin by observing that certain partition functions from statistical physics arise in graph theory as graph polynomials.

The hard-core partition function $Z_G(\lambda)$ is the *independence polynomial*. The independence polynomial can also be thought of as the generating function for the sequence $i_0(G), i_1(G), \dots$ where $i_k(G)$ denotes the number of independent sets of size k in G :

$$Z_G(\lambda) = \sum_{k \geq 0} i_k \lambda^k.$$

Important parameters in graph theory can be computed using the independence polynomial. For example, $Z_G(1)$ counts the total number of independent sets of G , $i(G)$. The order of the highest term is the independence number, $\alpha(G)$, the size of the largest independent set of G . The highest order coefficient of the independence polynomial counts the number of maximum independent sets in G .

Likewise, the monomer-dimer partition function is the *matching polynomial* (or the matching polynomial) and is also a generating function:

$$Z_G^{\text{match}}(\lambda) = \sum_{k \geq 0} m_k(G) \lambda^k$$

where $m_k(G)$ is the number of matchings of size k in G . The matching number $\nu(G)$ is the order of the highest term. If $\nu(G) = \frac{1}{2}|V(G)|$, then G has a *perfect matching*, a matching that saturates all vertices. In that case $m_{|V(G)|/2}(G)$ counts the number of perfect matchings of G .

Towards the end of the course we'll see another partition function, the q -color Potts model partition function, defined as

$$Z^q(G) = \sum_{k \geq 0} c_k^q(G)$$

where $c_k^q(G)$ counts the number of different colorings of $V(G)$ with q -colors that result in exactly k monochromatic edges. The coefficient $c_0^q(G)$ counts the number of *proper* q -colorings of G . The Potts model partition function is a specification of the Tutte polynomial and is closely related to the chromatic polynomial.

3.2. Extremal results. The field of extremal combinatorics asks for the maximum and minimum of various graph parameters over different classes of graphs. Some examples of classic theorems from extremal combinatorics are Mantel's Theorem which answers the question which graph on n vertices containing no triangles has the most edges? (Answer: the complete balanced bipartite graph with $\sim n^2/4$ edges); or Dirac's Theorem: which graph on n vertices containing no Hamilton cycle has the largest minimum degree? (Answer: minimum degree $n/2$ guarantees a Hamilton cycle; this is tight by taking two disjoint cliques of size $n/2$ each). In Lecture 4 we will discuss an important branch of extremal combinatorics, Ramsey theory.

Here we focus on extremal results for bounded-degree graphs.

3.2.1. Independent sets in regular graphs. Which d -regular graph has the most independent sets? This question was first raised in the context of number theory by Andrew Granville, and the first approximate answer was given by Noga Alon [2] who applied the result to problems in combinatorial group theory.

Jeff Kahn gave a tight answer in the case of d -regular bipartite graphs.

Theorem 1 (Kahn [36]). *Let $2d$ divide n . Then for any d -regular, bipartite graph G on n vertices,*

$$i(G) \leq i(H_{d,n}) = \left(2^{d+1} - 1\right)^{n/2d},$$

where $H_{d,n}$ is the graph consisting of $n/2d$ copies of $K_{d,d}$.

In terms of the independence polynomial, we can rephrase this as: for any d -regular, bipartite G ,

$$Z_G(1) \leq Z_{K_{d,d}}(1)^{n/2d}.$$

Kahn's proof is via the *entropy method*. Recall the *entropy* of a discrete random variable Y :

$$H(Y) = - \sum_y \Pr[Y = y] \log \Pr[Y = y].$$

See Appendix B.2 for the basics of entropy from information theory. See also Galvin's lecture notes on the entropy method [28] for an exposition of Theorem 1 and extensions. The main tool we will use is Shearer's Lemma.

Lemma 2 (Shearer's Lemma [11]). *Let \mathcal{F} be a family of subsets of $[n] = \{1, \dots, n\}$ so that each element $i \in [n]$ is contained in at least t sets of \mathcal{F} . For a (random) vector (X_1, \dots, X_n) and a set $F \subseteq [n]$, let $X_F = (X_i)_{i \in F}$. Then*

$$H(X_1, \dots, X_n) \leq \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F).$$

Proof of Theorem 1. Let G be a d -regular, n vertex graph with bipartition (L, R) . Let \mathbf{I} be an independent set chosen uniformly from all the independent sets of G . Since \mathbf{I} is uniform, we have

$$H(\mathbf{I}) = \log i(G),$$

and so we aim to show

$$H(\mathbf{I}) \leq \frac{n}{2d} \log \left(2^{d+1} - 1 \right).$$

Order the vertices of G v_1, \dots, v_n with the first $n/2$ in L and the rest in R . We can express the random independent set \mathbf{I} as a random vector of 0's and 1's, $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ where $\mathbf{X}_i = 1$ if $v_i \in \mathbf{I}$. Let $\mathbf{X}_L = (\mathbf{X}_1, \dots, \mathbf{X}_{n/2})$ and $\mathbf{X}_R = (\mathbf{X}_{n/2+1}, \dots, \mathbf{X}_n)$. For a vertex v , $\mathbf{X}_{N(v)}$ denotes $\mathbf{X}_{u_1}, \dots, \mathbf{X}_{u_d}$ where u_1, \dots, u_d are the neighbors of v . By the Chain Rule for entropy we have

$$\begin{aligned} H(\mathbf{I}) &= H(\mathbf{X}_1, \dots, \mathbf{X}_n) = H(\mathbf{X}_L) + H(\mathbf{X}_R | \mathbf{X}_L) \\ &\leq H(\mathbf{X}_1, \dots, \mathbf{X}_n) = H(\mathbf{X}_L) + \sum_{i=1}^{n/2} H(\mathbf{X}_i | \mathbf{X}_{N(v_i)}) \end{aligned}$$

by subadditivity and monotonicity of conditioning. Now we can apply Shearer's Lemma to $H(\mathbf{X}_L)$ with the family of sets \mathcal{F} being the neighborhoods $N(v_1), \dots, N(v_{n/2})$; each $i \in [n/2]$ is covered by exactly d sets since G is d -regular. This gives

$$(1) \quad H(\mathbf{I}) \leq \sum_{i=1}^{n/2} \frac{1}{d} H(\mathbf{X}_{N(v_i)}) + H(\mathbf{X}_i | \mathbf{X}_{N(v_i)}).$$

Now we bound each term by conditioning on the event that v_i is *uncovered*, that is, $N(v) \cap \mathbf{I} = \emptyset$. Let $q(v_i) = \Pr[N(v_i) \cap \mathbf{I} = \emptyset]$. Then

$$\begin{aligned} H(\mathbf{X}_{N(v_i)}) &= -q(v_i) \log q(v_i) - (1 - q(v_i)) \log(1 - q(v_i)) + (1 - q(v_i)) H(\mathbf{X}_{N(v_i)} | N(v) \cap \mathbf{I} = \emptyset) \\ &\leq q(v_i) \log \frac{1}{q(v_i)} + (1 - q(v_i)) \log \left(\frac{2^d - 1}{1 - q(v_i)} \right) \end{aligned}$$

using the bound $H(\mathbf{Y}) \leq \log(|\text{support}(\mathbf{Y})|)$ in the last step. We can also write

$$H(\mathbf{X}_i | \mathbf{X}_{N(v_i)}) = q(v_i) \log(2).$$

Putting these together gives

$$\begin{aligned} H(\mathbf{X}_{N(v_i)}) + d \cdot H(\mathbf{X}_i | \mathbf{X}_{N(v_i)}) &= q(v_i) \cdot \log\left(\frac{2^d}{q(v_i)}\right) + (1 - q(v_i)) \cdot \log\left(\frac{2^d - 1}{1 - q(v_i)}\right) \\ &\leq \log(2^{d+1} - 1) \quad \text{by Jensen's Inequality.} \end{aligned}$$

Substituting this back into (1), we obtain

$$H(\mathbf{I}) \leq \frac{n}{2d} \log(2^{d+1} - 1)$$

as desired. \square

3.2.2. Bregman's Theorem.

Theorem 3 (Bregman [7]). *Let A be an $n \times n$ matrix with $\{0, 1\}$ -valued entries and row sums d_1, \dots, d_n . Then*

$$\text{perm}(A) \leq \prod_{i=1}^n (d_i!)^{1/d_i}.$$

Let $m_{\text{perf}}(G)$ denote the number of perfect matchings of a graph G .

Corollary 4. *Suppose G is a bipartite graph on two parts of $n/2$ vertices each, with left degrees $d_1, \dots, d_{n/2}$, then*

$$m_{\text{perf}}(G) \leq \prod_{i=1}^{n/2} (d_i!)^{1/d_i}.$$

In particular, if $2d$ divides n , and G is a d -regular bipartite graph, then

$$m_{\text{perf}}(G) \leq m_{\text{perf}}(H_{d,n}).$$

In the case of d -regular bipartite graphs, Bregman's theorem states that

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

Proof of Theorem 3. The proof we present is due to Radhakrishnan [45]. Let G be a bipartite graph on two sets (L, R) of $n/2$ vertices each with left degrees $d_1, \dots, d_{n/2}$. Let \mathbf{M} be a uniformly random perfect matching from G .

Suppose $V(G) = L \cup R$ with $L = \{u_1, \dots, u_{n/2}\}$ and $R = \{v_1, \dots, v_{n/2}\}$. We view a perfect matching in G as a permutation σ of $[n/2]$ so that $(u_i, v_{\sigma(i)}) \in E(G)$. Let S be the set of all such σ and let σ be a uniformly random element of S .

Since σ is chosen uniformly from S , and $|S| = m_{\text{perf}}(G)$, we have

$$\log m_{\text{perf}}(G) = H(\sigma).$$

So our goal is to prove the upper bound

$$H(\sigma) \leq \sum_{i=1}^{n/2} \frac{\log d_i!}{d_i}.$$

Let τ be a permutation of $[n/2]$. Then we will uncover the permutation σ in the order determined by τ :

$$\sigma(\tau(1)), \sigma(\tau(2)), \dots, \sigma(\tau(n/2)).$$

By the chain rule of entropy, we have

$$H(\boldsymbol{\sigma}) = \sum_{i=1}^{n/2} H(\boldsymbol{\sigma}(\tau(i)) | \boldsymbol{\sigma}(\tau(1)), \dots, \boldsymbol{\sigma}(\tau(i-1))).$$

Since this is true for every τ , we can take the expectation over a uniformly random τ .

$$H(\boldsymbol{\sigma}) = \sum_{i=1}^{n/2} \mathbb{E}_{\tau} H(\boldsymbol{\sigma}(\tau(i)) | \boldsymbol{\sigma}(\tau(1)), \dots, \boldsymbol{\sigma}(\tau(i-1))).$$

For τ and i fixed, let k be such that $\tau(k) = i$. Then we can write

$$H(\boldsymbol{\sigma}) = \sum_{i=1}^{n/2} \mathbb{E}_{\tau} H(\boldsymbol{\sigma}(i) | \boldsymbol{\sigma}(\tau(1)), \dots, \boldsymbol{\sigma}(\tau(k-1))).$$

Now let $R_i(\boldsymbol{\sigma}, \tau)$ be the number of neighbors of u_i that have not been revealed by $\boldsymbol{\sigma}(\tau(1)), \dots, \boldsymbol{\sigma}(\tau(k-1))$. Then using the fact that $H(Y) \leq \log |\text{supp}(Y)|$,

$$\begin{aligned} H(\boldsymbol{\sigma}) &\leq \sum_{i=1}^{n/2} \mathbb{E}_{\tau} \sum_{j=1}^{d_i} \Pr_{\boldsymbol{\sigma}}[|R_i(\boldsymbol{\sigma}, \tau)| = j] \log j \\ &= \sum_{i=1}^{n/2} \sum_{j=1}^{d_i} \Pr_{\boldsymbol{\sigma}, \tau}[|R_i(\boldsymbol{\sigma}, \tau)| = j] \log j. \end{aligned}$$

Now for any fixed $\boldsymbol{\sigma}$, $\Pr_{\tau}[|R_i(\boldsymbol{\sigma}, \tau)| = j] = 1/d_i$ by symmetry, and so

$$\begin{aligned} H(\boldsymbol{\sigma}) &\leq \sum_{i=1}^{n/2} \frac{1}{d_i} \sum_{j=1}^{d_i} \log j \\ &= \sum_{i=1}^{n/2} \frac{\log d_i!}{d_i} \end{aligned}$$

as desired. □

3.3. Extensions and reductions. Galvin and Tetali [30] proved a very broad generalization of Kahn's result. To describe it we need a definition.

Definition. Let G and H be two finite graphs. A map $\phi : V(G) \rightarrow V(H)$ is a **homomorphism** from G to H if for every $(u, v) \in E(G)$, $(\phi(u), \phi(v)) \in E(H)$. Let $\text{hom}(G, H)$ denote the number of homomorphisms from G to H .

Example. Suppose H^{ind} is the graph on two vertices w_1, w_2 , with an edge between w_1 and w_2 and a self-loop on w_2 . Then homomorphisms from G to H^{ind} correspond to independent sets, with the vertices of the independent set given by $\phi^{-1}(w_1)$, and so $\text{hom}(G, H^{\text{ind}}) = i(G)$, the number of independent sets of G .

Example. Homomorphisms from G to K_q , the complete graph on q vertices, correspond to proper q -colorings of $V(G)$.

Jeff Kahn's result on independent sets can be restated as the fact that

$$\text{hom}(G, H^{\text{ind}}) \leq \text{hom}(K_{d,d}, H^{\text{ind}})^{n/2d}$$

for any d -regular, bipartite G . What Galvin and Tetali show is that this result holds in complete generality over the choice of the target graph H , at least over bipartite regular graphs.

Theorem 5. *Let H be any graph, with or without self-loops, and let G be any d -regular, bipartite graph. Then*

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

Theorem 6 (Kahn [36], Galvin-Tetali [30], Zhao [60]). *For any d -regular graph G and any $\lambda > 0$,*

$$\frac{1}{|V(G)|} \log Z_G(\lambda) \leq \frac{1}{2d} \log Z_{K_{d,d}}(\lambda).$$

Removing the bipartite restriction, as Zhao did for independent sets, is not possible in general. (For a simple example see the exercises). Zhao [61] found a large family of H for which the statement for general d -regular graphs reduces to the bipartite statement. Sernau [48] found a large such family and disproved a conjecture of Galvin [27] that either $K_{d,d}$ or K_{d+1} is always the maximizing graph.

One specific target graph H that arises in statistical physics is H^{WR} , the graph on three vertices, w_1, w_2, w_3 , each with a self-loop, and w_2 joined to both w_1 and w_3 . Homomorphisms from G to H^{WR} correspond to valid configurations in the Widom-Rowlinson model in which particles of two different types can occupy sites on a lattice, but particles of different types are forbidden from neighboring each other. (The vertices w_1 and w_3 correspond to the two different types, while w_2 corresponds to an empty site). The extremal graph for Widom-Rowlinson configurations is the clique.

Theorem 7 (Cohen, Perkins, Tetali [13]; Sernau [48]). *For all d -regular G ,*

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

This theorem was first proved with the methods of Section 4, but it was later observed [12, 48] that it follows from a reduction from the case of independent sets.

Perhaps the most outstanding unresolved case of target graph H is that of proper colorings ($H = K_q$). Galvin and Tetali conjectured that a union of $K_{d,d}$'s maximize the number of q -colorings over all d -regular graphs.

Conjecture 8 (Galvin, Tetali [30]). *For all d -regular G , and all $q \geq 2$,*

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

Galvin and Tetali prove this for bipartite G with their general theorem. Partial results towards this conjecture have been proved [61, 27, 29]. Recently the case $d = 3$ and arbitrary q was resolved [21]. The proof proceeds via the Potts model from statistical physics; we will return to this in Lecture 5.

Further reading. Zhao wrote a recent survey on extremal problems for homomorphism counts in regular graph [62]. See also Csikvári [17] for further results and open problems in this area.

Exercises.

- (1) Let H be the graph consisting of two vertices each with self loops but no edge joining the two. Show that $\text{hom}(K_{d+1}, H)^{1/(d+1)} > \text{hom}(K_{d,d}, H)^{1/2d}$.
- (2) Let G be any graph. Let $G \times K_2$ be the *bipartite double cover* of G . That is, $G \times K_2$ has vertex set $V(G) \times \{0, 1\}$ with an edge between (u, i) and (v, j) if $uv \in E(G)$ and $i = 0, j = 1$ or $i = 1, j = 0$. Prove that

$$i(G)^2 \leq i(G \times K_2).$$

Deduce that Kahn's theorem on independent sets in d -regular bipartite graphs can be extended to all d -regular graphs. This is Zhao's proof [60].

- (3) Prove (in an elementary way) that for all d -regular G on n vertices,

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

- (4) What is the probability (in the hard-core model) that a vertex v is uncovered, given that the subgraph induced by its uncovered neighbors is isomorphic to a graph H ?
- (5) Let G, H be two graphs on n vertices each. Prove that if $Z_G^M(\lambda) \geq Z_H^M(\lambda)$ for all $\lambda > 0$, then $m_{\text{perf}}(G) \geq m_{\text{perf}}(H)$.
- (6) By going through the proof of Theorem 1, show that equality in the theorem is attained only by $K_{d,d}$ or $H_{d,n}$.

4. OCCUPANCY FRACTIONS AND OPTIMIZATION

In this section we present a new approach to proving extremal theorems for sparse graphs. We will prove extremal results for partition functions and graph polynomials by optimizing the logarithmic derivative of a given partition function over a given class of graphs. By integrating we obtain the corresponding result for the partition function. The logarithmic derivative has a probabilistic interpretation as the expectation of an observable of the relevant model from statistical physics, and so working directly with the model we impose local probabilistic constraints bases on the graph structure and parameters of the model, then relax the optimization problem to all local probability distributions satisfying these constraints.

An *observable* of a Gibbs measure is simply a random variable, a function of the random configuration. One particularly important observable is the energy. In general we can describe a Gibbs distribution in terms of a *Hamiltonian*, or energy function, $\mathcal{H}(\sigma)$ mapping configurations to real numbers. We then write

$$\Pr[\boldsymbol{\sigma} = \sigma] = \frac{e^{-\beta\mathcal{H}(\sigma)}}{Z(\beta)},$$

where β is called the *inverse temperature* and the partition function $Z(\beta) = \sum_{\sigma} e^{-\beta\mathcal{H}(\sigma)}$ is again the normalizing constant.

The energy of the random configuration $\mathcal{H}(\sigma)$ is one example of an observable. The expectation of this observable is

$$\begin{aligned}\mathbb{E}_\beta \mathcal{H}(\sigma) &= \sum_{\sigma} \mathcal{H}(\sigma) \Pr_{\beta}[\sigma = \sigma] \\ &= \sum_{\sigma} \mathcal{H}(\sigma) \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z(\beta)} \\ &= -\frac{Z'(\beta)}{Z(\beta)} \\ &= -(\log Z(\beta))' .\end{aligned}$$

This simple calculation is the key to the method that follows.

4.1. The occupancy fraction of the hard-core model. We start with independent sets and the hard-core model, where the relevant observable is the occupancy fraction.

Definition 9. *The occupancy fraction of the hard-core model on a graph G at fugacity λ is the expected fraction of vertices that appear in the random independent set:*

$$\bar{\alpha}_G(\lambda) = \frac{1}{|V(G)|} \mathbb{E}_{G,\lambda} |\mathbf{I}|.$$

We begin by collecting some basic facts about the occupancy fraction.

Proposition 10. *The occupancy fraction is λ times the derivative of the free energy:*

$$\bar{\alpha}_G(\lambda) = \lambda \cdot \left(\frac{1}{|V(G)|} \log Z_G(\lambda) \right)' .$$

Proof. We write

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{1}{|V(G)|} \mathbb{E}_{G,\lambda} |\mathbf{I}| \\ &= \frac{1}{|V(G)|} \sum_{I \in \mathcal{I}(G)} |I| \cdot \Pr_{G,\lambda}[\mathbf{I} = I] \\ &= \frac{1}{|V(G)|} \sum_{I \in \mathcal{I}(G)} |I| \cdot \frac{\lambda^{|I|}}{Z_G(\lambda)} \\ &= \frac{\lambda}{|V(G)|} \sum_{I \in \mathcal{I}(G)} |I| \cdot \frac{\lambda^{|I|-1}}{Z_G(\lambda)} \\ &= \frac{\lambda}{|V(G)|} \frac{Z'_G(\lambda)}{Z_G(\lambda)} \\ &= \lambda \cdot \left(\frac{1}{|V(G)|} \log Z_G(\lambda) \right)' .\end{aligned}$$

□

Proposition 11. *The occupancy fraction $\bar{\alpha}_G(\lambda)$ is an increasing function of λ .*

Proof. Writing Z_G for $Z_G(\lambda)$,

$$\begin{aligned} |V(G)| \cdot \bar{\alpha}'_G(\lambda) &= \left(\frac{\lambda Z'_G}{Z_G} \right)' = \frac{Z'_G}{Z_G} + \frac{\lambda Z_G Z''_G - \lambda (Z'_G)^2}{Z_G^2} \\ &= \frac{Z'_G}{Z_G} + \frac{1}{\lambda} \left(\frac{\lambda^2 Z''_G}{Z_G} - \left(\frac{\lambda Z'_G}{Z_G} \right)^2 \right) \\ &= \frac{\mathbb{E}_{G,\lambda}|\mathbf{I}| + \mathbb{E}_{G,\lambda}(|\mathbf{I}|^2) - \mathbb{E}_{G,\lambda}|\mathbf{I}| - (\mathbb{E}_{G,\lambda}|\mathbf{I}|)^2}{\lambda} \\ &= \frac{\text{var}_{G,\lambda}(|\mathbf{I}|)}{\lambda} > 0. \end{aligned}$$

□

Given Proposition 10, we make a simple observation. If for all $G \in \mathcal{G}$ and all $\lambda > 0$ we have $\bar{\alpha}_G(\lambda) \leq \bar{\alpha}_{G_0}(\lambda)$, then we have

$$\frac{1}{|V(G)|} \log Z_G(\lambda) \leq \frac{1}{|V(G_0)|} \log Z_{G_0}(\lambda)$$

for all $\lambda > 0$. This follows from a bit of calculus:

$$\begin{aligned} \frac{1}{|V(G)|} \log Z_G(\lambda) &= \frac{1}{|V(G)|} \log Z_G(0) + \int_0^\lambda \left(\frac{1}{|V(G)|} \log Z_G(t) \right)' dt \\ &= \int_0^\lambda \left(\frac{1}{|V(G)|} \log Z_G(t) \right)' dt \\ &\leq \int_0^\lambda \left(\frac{1}{|V(G)|} \log Z_{G_0}(t) \right)' dt \\ &= \frac{1}{|V(G_0)|} \log Z_{G_0}(\lambda). \end{aligned}$$

We can use this observation to prove a strengthening of Theorem 6.

Theorem 12 (Davies, Jenssen, Perkins, Roberts [22]). *For any d -regular graph G , and any $\lambda > 0$,*

$$\bar{\alpha}_G(\lambda) \leq \bar{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1}.$$

Proof of Theorem 12. We prove this first for triangle-free G to illustrate the method.

We say v is *uncovered* with respect to an independent set I if $N(v) \cap I = \emptyset$.

Consider the hard-core model on a d -regular, triangle-free G on n vertices.

Fact 1: $\Pr[v \in I | v \text{ uncovered}] = \frac{\lambda}{1+\lambda}$.

Fact 2: $\Pr[v \text{ uncovered} | v \text{ has } j \text{ uncovered neighbors}] = (1+\lambda)^{-j}$.

Fact 2 relies on the fact that G is triangle-free: the graph induced by the uncovered neighbors of v consists of isolated vertices.

Now we write $\bar{\alpha}_G(\lambda)$ in two ways:

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{1}{n} \sum_{v \in V(G)} \Pr[v \in \mathbf{I}] \\ &= \frac{1}{n} \frac{\lambda}{1 + \lambda} \sum_{v \in V(G)} \Pr[v \text{ uncovered}] \quad \text{by Fact 1} \\ &= \frac{1}{n} \frac{\lambda}{1 + \lambda} \sum_{v \in V(G)} \sum_{j=0}^d \Pr[v \text{ has } j \text{ uncovered neighbors}] \cdot (1 + \lambda)^{-j} \quad \text{by Fact 2,}\end{aligned}$$

and

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{1}{n} \frac{1}{d} \sum_{v \in V(G)} \sum_{u \sim v} \Pr[u \in \mathbf{I}] \quad \text{since } G \text{ is } d\text{-regular} \\ &= \frac{1}{n} \frac{1}{d} \frac{\lambda}{1 + \lambda} \sum_{v \in V(G)} \sum_{u \sim v} \Pr[u \text{ uncovered}] \quad \text{by Fact 1.}\end{aligned}$$

Now consider the following two-part experiment: pick \mathbf{I} from the hard-core model on G and independently choose \mathbf{v} uniformly at random from $V(G)$. Let \mathbf{Y} be the number of uncovered neighbors of \mathbf{v} with respect to \mathbf{I} . Now our two expressions for $\bar{\alpha}_G(\lambda)$ can be interpreted as expectations over \mathbf{Y} .

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{\lambda}{1 + \lambda} \mathbb{E}_{G,\lambda} (1 + \lambda)^{-\mathbf{Y}} \\ \bar{\alpha}_G(\lambda) &= \frac{1}{d} \frac{\lambda}{1 + \lambda} \mathbb{E}_{G,\lambda} \mathbf{Y}.\end{aligned}$$

Thus the identity

$$(2) \quad \mathbb{E}_{G,\lambda} (1 + \lambda)^{-\mathbf{Y}} = \frac{1}{d} \mathbb{E}_{G,\lambda} \mathbf{Y}$$

holds for all d -regular triangle-free graphs G .

Now the idea is to *relax* the maximization problem; instead of maximizing $\bar{\alpha}_G(\lambda)$ over all d -regular graphs, we can maximize $\frac{\lambda}{1 + \lambda} \mathbb{E} (1 + \lambda)^{-\mathbf{Y}}$ over all distributions of random variables \mathbf{Y} that are bounded between 0 and d and satisfy the constraint (2).

It is not too hard to see that to maximize $\mathbb{E} \mathbf{Y}$ subject to these constraints, we must put all of the probability mass of \mathbf{Y} on 0 and d . Because of the constraint (2), there is a unique such distribution.

Now fix a vertex v in $K_{d,d}$. If any vertex on v 's side of the bipartition is in I , then v has 0 uncovered neighbors. If no vertex on the side is in I , then v has d uncovered neighbors. So the distribution of \mathbf{Y} induced by $K_{d,d}$ (or $H_{d,n}$) is exactly the unique distribution satisfying the constraints that is supported on 0 and d . And therefore,

$$\bar{\alpha}_G(\lambda) \leq \bar{\alpha}_{K_{d,d}}(\lambda).$$

Now we give the full proof for graphs that may contain triangles.

Let G be a d -regular n -vertex graph (with or without triangles). Do the following two part experiment: sample \mathbf{I} from the hard-core model on G at fugacity λ , and independently choose \mathbf{v} uniformly from $V(G)$. Previously we considered the random variable \mathbf{Y} counting

the number of uncovered neighbors of \mathbf{v} . When G was triangle-free we knew there were no edges between these uncovered vertices, but now we must consider these potential edges. Let \mathbf{H} be the graph induced by the uncovered neighbors of \mathbf{v} ; \mathbf{H} is a random graph over the randomness in our two-part experiment.

We now can write $\bar{\alpha}_G(\lambda)$ in two ways, as expectations involving \mathbf{H} .

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{\lambda}{1+\lambda} \Pr_{G,\lambda}[\mathbf{v} \text{ uncovered}] = \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda} \left[\frac{1}{Z_{\mathbf{H}}(\lambda)} \right] \\ \bar{\alpha}_G(\lambda) &= \frac{1}{d} \mathbb{E}_{G,\lambda}[\mathbf{I} \cap N(\mathbf{v})] = \frac{\lambda}{d} \mathbb{E}_{G,\lambda} \left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)} \right],\end{aligned}$$

and so for any d -regular graph G , we have the identity

$$(3) \quad \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda} \left[\frac{1}{Z_{\mathbf{H}}(\lambda)} \right] = \frac{\lambda}{d} \mathbb{E}_{G,\lambda} \left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)} \right].$$

Now again we can relax our optimization problem from maximizing $\bar{\alpha}_G$ over all d -regular graphs, to maximizing $\frac{\lambda}{1+\lambda} \mathbb{E} \left[\frac{1}{Z_{\mathbf{H}}(\lambda)} \right]$ over all possible distributions \mathbf{H} on \mathcal{H}_d , the set of graphs on at most d vertices, satisfying the constraint (3).

We claim that the unique maximizing distribution is the one distribution supported on the empty graph, \emptyset , and the graph of d isolated vertices, \overline{K}_d . This is the distribution induced by $K_{d,d}$ (or $H_{d,n}$) and is given by

$$\begin{aligned}\Pr_{K_{d,d}}(\mathbf{H} = \emptyset) &= \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1} \\ \Pr_{K_{d,d}}(\mathbf{H} = \overline{K}_d) &= \frac{(1+\lambda)^d}{2(1+\lambda)^d - 1}.\end{aligned}$$

To show that this distribution is the maximizer we will use linear programming (see Appendix B.3 for some basic facts about linear programming).

Both our objective function and our constraint are linear functions of the variables $\{p(H)\}_{H \in \mathcal{H}_d}$, so we can pose the relaxation as a linear program.

$$\begin{aligned}\text{maximize} \quad & \sum_{H \in \mathcal{H}_d} p(H) \cdot \frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} \\ \text{subject to} \quad & p(H) \geq 0 \quad \forall H \in \mathcal{H}_d \\ & \sum_{H \in \mathcal{H}_d} p(H) = 1 \\ & \sum_{H \in \mathcal{H}_d} p(H) \left[\frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} - \frac{\lambda Z'_H(\lambda)}{d Z_H(\lambda)} \right] = 0.\end{aligned}$$

The first two constraints insure that the variables $p(H)$ form a probability distribution; the last is constraint (3).

Our candidate solution is $p(\emptyset) = \frac{(1+\lambda)^{d-1}}{2(1+\lambda)^{d-1}}$, $p(\overline{K_d}) = \frac{(1+\lambda)^d}{2(1+\lambda)^{d-1}}$, with objective value $\overline{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^{d-1}}$. To prove that this solution is optimal (and thus prove the theorem), we need to find some feasible solution to the dual with objective value $\overline{\alpha}_{K_{d,d}}(\lambda)$.

The dual linear program is:

$$\begin{aligned} & \text{minimize} && \Lambda_p \\ & \text{subject to} && \Lambda_p + \Lambda_c \cdot \left[\frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} - \frac{\lambda}{d} \frac{Z'_H(\lambda)}{Z_H(\lambda)} \right] \geq \frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} \quad \text{for all } H \in \mathcal{H}_d. \end{aligned}$$

For each variable of the primal, indexed by $H \in \mathcal{H}_d$, we have a dual constraint. For each constraint in the primal (not including the non-negativity constraint), we have a dual variable, in this case Λ_p corresponding to the probability constraint (summing to 1) and Λ_c corresponding to the remaining constraint. (Note that we do not have non-negativity constraints $\Lambda_p, \Lambda_c \geq 0$ in the dual because the corresponding primal constraints were equality constraints).

Now our task becomes: find a feasible dual solution with $\Lambda_p = \overline{\alpha}_{K_{d,d}}(\lambda)$. What should we choose for Λ_c ? By complementary slackness in linear programming, the dual constraint corresponding to any primal variable that is strictly positive in an optimal solution must hold with equality in an optimal dual solution. In other words, we expect the constraints corresponding to $H = \emptyset, \overline{K_d}$ to hold with equality. This allows us to solve for a candidate value for Λ_c . Using $Z_\emptyset(\lambda) = 1$ and $Z'_\emptyset(\lambda) = 0$, we have the equation

$$\overline{\alpha}_{K_{d,d}}(\lambda) + \Lambda_c \left[\frac{\lambda}{1+\lambda} - 0 \right] = \frac{\lambda}{1+\lambda}.$$

Solving for Λ_c gives

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

Now with this choice of Λ_c , and $\Lambda_p = \overline{\alpha}_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^{d-1}}$, our dual constraint for $H \in \mathcal{H}_d$ becomes:

$$(4) \quad \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1} + \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1} \left[\frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)} - \frac{\lambda}{d} \frac{Z'_H(\lambda)}{Z_H(\lambda)} \right] \geq \frac{\lambda}{1+\lambda} \frac{1}{Z_H(\lambda)}.$$

Multiplying through by $Z_H(\lambda) \cdot (2(1+\lambda)^d - 1)$ and simplifying, (4) reduces to

$$(5) \quad \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} \geq \frac{\lambda Z'_H(\lambda)}{Z_H(\lambda) - 1},$$

and we must show this holds for all $H \in \mathcal{H}_d$ (except for $H = \emptyset$ for which we know already the dual constraint holds with equality). Luckily (5) has a nice probabilistic interpretation: the RHS is simply $\mathbb{E}_{H,\lambda} [|\mathbf{I}| \mid |\mathbf{I}| \geq 1]$, the expected size of the random independent set given that it is not empty, and the LHS is the same for the graph of d isolate vertices, $\overline{K_d}$. Proving (5) is left for the exercises, and this completes the proof. \square

4.2. The general method. Here we give an overview of a general method suggested by the proof of Theorem 12 above.

4.3. Graph convergence and optimization. In the realm of dense graphs (graphs on n vertices with $\Theta(n)$ edges), there is a beautiful notion of convergence for a sequence of graphs based on subgraph densities due to L S and others (see Lovasz's textbook [39] as well as ..).

Graph convergence goes hand in hand with a powerful method for solving extremal problems for dense graphs, the method of Flag algebras developed by Razborov [46].

For sparse graphs, there is a notion of convergence due to Bejamini and Schramm [4].

Open Problem. *For a simple class of graphs and definition of local view (e.g. the number of occupied neighbors of a random vertex in the hard-core model on 3-regular graphs) can we completely characterize (the closure of) the set of all distributions on local views that are achievable by graphs?*

4.4. Matchings.

Theorem 13 (Davies, Jenssen, Perkins, Roberts [22]). *For any d -regular graph G , and any $\lambda > 0$,*

$$\alpha_G^{\text{match}}(\lambda) \leq \alpha_{K_{d,d}}^{\text{match}}(\lambda).$$

Corollary 14. *For any d -regular graph G , and any $\lambda > 0$,*

$$\frac{1}{|V(G)|} \log Z_G^{\text{match}}(\lambda) \leq \frac{1}{2d} \log Z_{K_{d,d}}^{\text{match}}(\lambda).$$

The proof of Theorem 13 follows the same idea as the proof of Theorem 12 but is more technically involved.

Consider sampling a random matching \mathbf{M} from the monomer-dimer model on a d -regular graph G at fugacity λ and independently choosing an edge \mathbf{e} uniformly at random from $E(G)$. The closed neighborhood of an edge e , $\bar{N}(e)$, is e along with all its incident edges. The *externally uncovered neighborhood* of \mathbf{e} with respect to the matching \mathbf{M} is the set of edges of $\bar{N}(\mathbf{e})$ that are not incident to any edge in $\mathbf{M} \cap \bar{N}(\mathbf{e})^c$. That is, the externally uncovered neighborhood of \mathbf{e} are those edges that are not blocked from being in the random matching by some edge f outside of $\bar{N}(\mathbf{e})$. Call the subgraph induced by the externally uncovered neighborhood $\mathbf{C} = \mathbf{C}(\mathbf{M}, \mathbf{e})$. \mathbf{C} always includes at least one edge, namely \mathbf{e} itself. Note that the presence or absence of edges in $\bar{N}(\mathbf{e})$ in the matching have no effect on \mathbf{C} .

Now we can write the edge occupancy fraction as an expectation over \mathbf{C} .

$$\begin{aligned} \alpha_G^{\text{match}}(\lambda) &= \frac{1}{|E(G)|} \sum_{e \in E(G)} \Pr_{G,\lambda}[e \in \mathbf{M}] \\ &= \frac{1}{|E(G)|} \sum_{e \in E(G)} \mathbb{E}_{G,\lambda} \left[\frac{\lambda}{Z_{\mathbf{C}(\mathbf{M},e)}(\lambda)} \right] \\ &= \mathbb{E}_{G,\lambda} \left[\frac{\lambda}{Z_{\mathbf{C}}(\lambda)} \right]. \end{aligned}$$

The second equality above uses the spatial Markov property of the monomer-dimer model. Alternatively, we can use the regularity of G to write

$$\begin{aligned} \alpha_G^{\text{match}}(\lambda) &= \frac{1}{2d-1} \frac{1}{|E(G)|} \sum_{e \in E(G)} \left[\Pr_{G,\lambda}[e \in \mathbf{M}] + \sum_{f \sim e} \Pr_{G,\lambda}[f \in \mathbf{M}] \right] \\ &= \frac{1}{2d-1} \frac{1}{|E(G)|} \sum_{e \in E(G)} \mathbb{E}_{G,\lambda} |\mathbf{M} \cap \bar{N}(e)| \\ &= \frac{1}{2d-1} \mathbb{E}_{G,\lambda} \left[\frac{\lambda Z'_C(\lambda)}{Z_C(\lambda)} \right]. \end{aligned}$$

The partition function of an externally neighborhood C can be written in terms of three parameters, L, R, K , where L, R are the number of edges incident to the left and right vertices of e respectively that are not in a triangle, and K denotes the number of triangles in C . Let $Z_{l,r,k}$ denote the partition function of C such that $L = l, R = r, K = k$. Then we can compute

$$Z_{l,r,k} = 1 + \lambda + (l + r + 2k)\lambda + [k^2 + k(l + r - 1) + lr]\lambda^2.$$

Open Problem. Determine the minimum of $\frac{1}{|V(G)|} \log Z_G^{\text{match}}(\lambda)$ over d -regular graphs. The answer may very well depend on λ and the parity of d .

- Is it true that for d even and all λ , K_{d+1} is the minimizer?
- Is it true that for all d and all $\lambda \leq 1$, K_{d+1} is the minimizer?

4.5. Independent sets in bipartite vertex-transitive graphs. Next, we see that the same methods can be used to prove lower bounds on the occupancy fraction in bipartite graphs, using very different kinds of constraints.

Now let $\bar{\alpha}_{T_d}(\lambda)$ be the solution to the equation

$$\frac{\alpha}{\lambda(1-\alpha)} = \left(\frac{1-2\alpha}{1-\alpha} \right)^d$$

We interpret $\bar{\alpha}_{T_d}(\lambda)$ as the occupancy fraction of the unique translation invariant hard-core measure on the infinite d -regular tree T_d at fugacity λ

Theorem 15 (Davies, Jensen, P., Roberts [22]). *Let G be a bipartite, vertex-transitive, d -regular graph. Then*

$$\bar{\alpha}_G(\lambda) > \bar{\alpha}_{T_d}(\lambda).$$

To prove Theorem 15 we need a correlation inequality.

Lemma 16. *Let G be a bipartite graph on bipartition $L \cup R$. Suppose $v_1, \dots, v_k \in L$. Then*

$$\Pr_{G,\lambda} \left[\bigwedge_{i=1}^k v_i \in \mathbf{I} \right] \geq \prod_{i=1}^k \Pr_{G,\lambda} [v_i \in \mathbf{I}]$$

There are several different proofs of this lemma; Van den Berg and Steif [54] show that it is a consequence of the FKG inequality.

Proof of Theorem 15. Fix a vertex $v \in V(G)$ and let \mathbf{Y} be the number of its uncovered neighbors when drawing \mathbf{I} according to the hard-core model on G . For each neighbor $u_1 \dots u_d$ of v , let \mathbf{Y}_{u_i} be the indicator random variable that u_i is uncovered.

$$\begin{aligned} \bar{\alpha}_G(\lambda) &= \Pr_{G,\lambda}[v \in \mathbf{I}] = \frac{\lambda}{1+\lambda} \mathbb{E}[(1+\lambda)^{-\mathbf{Y}}] \text{ since } G \text{ is triangle-free} \\ &= \frac{\lambda}{1+\lambda} \mathbb{E}[(1+\lambda)^{-\sum_{i=1}^d \mathbf{Y}_{u_i}}] \\ &= \frac{\lambda}{1+\lambda} \left(\bar{\alpha}_G(\lambda) + (1 - \bar{\alpha}_G(\lambda)) \mathbb{E}[(1+\lambda)^{-\sum_{i=1}^d \mathbf{Y}_{u_i}} | v \notin \mathbf{I}] \right), \end{aligned}$$

and rearranging gives

$$\frac{\bar{\alpha}_G(\lambda)}{\lambda(1 - \bar{\alpha}_G(\lambda))} = \mathbb{E}[(1+\lambda)^{-\sum_{i=1}^d \mathbf{Y}_{u_i}} | v \notin \mathbf{I}].$$

Now on the event $\{v \notin \mathbf{I}\}$, by vertex-transitivity the random variables $\mathbf{Y}_{u_1}, \dots, \mathbf{Y}_{u_d}$ each have a Bernoulli(p) distribution, where $p = \frac{1+\lambda}{\lambda} \frac{\bar{\alpha}_G(\lambda)}{1 - \bar{\alpha}_G(\lambda)}$. Moreover, Lemma 16 applied to $G \setminus v$, shows that these random variables are *positively correlated* (conditioned on $\{v \notin \mathbf{I}\}$). This gives

$$\begin{aligned} \frac{\bar{\alpha}_G(\lambda)}{\lambda(1 - \bar{\alpha}_G(\lambda))} &= \mathbb{E}[(1+\lambda)^{-\sum_{i=1}^d \mathbf{Y}_{u_i}} | v \notin \mathbf{I}] > \prod_{i=1}^d \mathbb{E}[(1+\lambda)^{-\mathbf{Y}_{u_i}} | v \notin \mathbf{I}] \\ &= \left(1 - p + \frac{p}{1+\lambda}\right)^d = \left(\frac{1 - 2\bar{\alpha}_G(\lambda)}{1 - \bar{\alpha}_G(\lambda)}\right)^d. \end{aligned}$$

The function $\frac{\alpha}{\lambda(1-\alpha)}$ is increasing in α , the function $\left(\frac{1-2\alpha}{1-\alpha}\right)^d$ is decreasing in α , and the two functions are equal at $\alpha = \alpha_{T_d}(\lambda)$, so we can conclude that $\bar{\alpha}_G(\lambda) > \alpha_{T_d}(\lambda)$. \square

The lower bound in Theorem 15 is in fact asymptotically tight when $\lambda \leq \lambda_c(T_d) = \frac{(d-1)^{d-1}}{(d-2)^d}$ (the uniqueness threshold of the hard-core model on T_d): there is a sequence of graphs G_n on n vertices that achieve the lower bound in the limit as $n \rightarrow \infty$.

Exercises.

- (1) Prove that for any graph G on at most d vertices,

$$\mathbb{E}_{G,\lambda} [|\mathbf{I}| | |\mathbf{I}| \geq 1] \leq \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1}.$$

- (2) Suppose G is a d -regular graph and suppose $v \in V(G)$ does not belong to a $K_{d,d}$ component. Give a lower bound (in terms of d, λ) on the probability that the uncovered neighborhood of v is not the empty graph or the graph of d isolated vertices. (This proves uniqueness in the independent set theorem in a quantitative way).
- (3) Suppose a vertex v of a graph G has d neighbors (but we make no assumption on the presence or absence of edges in its neighborhood). Do the hard-core model on G at fugacity λ , and let $p_k = \Pr[|\mathbf{I} \cap N(v)| = k]$. Give a lower bound on p_{k-1} in terms of p_k, d , and λ . Is the bound tight in any graph?
- (4) Use the tight bound from the previous question, for $k = 2, \dots, d$, to prove that $\alpha_G(\lambda) \leq \alpha_{K_{d,d}}(\lambda)$ for any d -regular G .
- (5) Prove that $\alpha_G(\lambda) \geq \alpha_{K_{d+1}}(\lambda)$ for any d -regular G .

5. LOWER BOUNDS, RAMSEY THEORY, AND SPHERE PACKINGS

5.1. Ramsey theory: $R(3, k)$.

Definition 17. *The Ramsey number $R(t, k)$ is the fewest number of vertices a graph must have to guarantee the existence of either a clique of size t or an independent set of size k .*

We could equivalently define the Ramsey numbers in terms of edge colorings of the complete graph: $R(t, k)$ is the smallest N so that any red/blue coloring of the edges of K_N contains either a monochromatic red or monochromatic blue clique.

Particularly important are the diagonal Ramsey numbers $R(k, k)$. Understanding the asymptotics of $R(k, k)$ as $k \rightarrow \infty$ is a central problem in extremal combinatorics.

Beyond the diagonal Ramsey numbers, the next most studied is $R(3, k)$: the minimum number of vertices that guarantee the existence of either a triangle or an independent set of size k . The major question here is to understand the asymptotics of $R(3, k)$ as $k \rightarrow \infty$. For an enjoyable history of the problem see Spencer's survey [52]. The asymptotic order $R(3, k) = \Theta(k^2/\log k)$ was determined in two seminal papers, the upper bound by Ajtai, Komlós, and Szemerédi [1] and the lower bound by Kim [37]. The constant in the upper bound was improved by Shearer [49] to

$$(6) \quad R(3, k) \leq (1 + o_k(1)) \frac{k^2}{\log k}.$$

Bohman and Keevash [5] and independently Fiz Pontiveros, Griffiths, and Morris [24] improved the constant in the lower bound to

$$R(3, k) \leq \left(\frac{1}{4} + o_k(1) \right) \frac{k^2}{\log k}$$

by analyzing the *triangle-free process*: begin with an empty graph G_0 on n vertices. To form G_{i+1} from G_i , add an edge (u, v) uniformly at random from the set of all potential edges that are not in $E(G_i)$ and would not create a triangle in G_{i+1} ; if there is no such edge to choose from, terminate the process.

Shearer's upper bound comes from the following theorem giving a lower bound on the independence number of triangle-free graphs.

Theorem 18 (Shearer [49]). *For every triangle-free graph G of average degree at most d ,*

$$\alpha(G) \geq (1 + o_d(1)) \frac{\log d}{d} \cdot |V(G)|.$$

Shearer's proof proceeds by analyzing a certain randomized greedy algorithm that constructs an independent set of the required size.

The bound (6) on $R(3, k)$ follows from Theorem 18 by the following observation. We must show that any triangle-free graph G on at least $(1 + \epsilon) \frac{k^2}{\log k}$ vertices has an independent set of size k . First suppose that the maximum degree of G is at least k . Then because G is triangle-free we can take the neighborhood of a vertex of degree at least k to obtain the desired independent set. Now suppose that all vertices have degree at most k (or at most

$k - 1$ if you prefer). Then Shearer's result states that

$$\begin{aligned}\alpha(G) &\geq (1 + o_k(1)) \frac{\log k}{k} \cdot |V(G)| \\ &\geq (1 + \epsilon)(1 + o_k(1)) \frac{\log k}{k} \frac{k^2}{\log k} \\ &\geq k\end{aligned}$$

for arbitrary $\epsilon > 0$ and large enough k , and therefore $R(3, k) \leq (1 + o_k(1)) \frac{k^2}{\log k}$.

Here we will apply the method of Section 4 to provide an alternative proof of this bound. We then give a conjecture suggested by the method; a proof of this conjecture would lead to a factor 2 improvement in the bound.

Theorem 19 (Davies, Jenssen, Perkins, Roberts [23]). *For $\lambda \geq 1$, and any triangle-free graph G of maximum degree d ,*

$$\bar{\alpha}_G(\lambda) \geq (1 + o_d(1)) \frac{\log d}{d}.$$

Compared to Theorem 18, Theorem 19 has a stronger condition: max degree d instead of average degree d ; but the conclusion is also stronger: instead of guaranteeing the existence of an independent set of size $(1 + o_d(1)) \frac{\log d}{d} \cdot n$, Theorem 19 says that the average size over *all* independent sets is at least this large. Below we will explore whether or not this should imply the existence of a significantly larger independent set.

Proof. Choose \mathbf{I} from the hard-core model on G at fugacity λ and independent choose \mathbf{v} uniformly at random from $V(G)$. Let $\mathbf{Y} = \mathbf{Y}(\mathbf{I}, \mathbf{v})$ be the number of uncovered neighbors of \mathbf{v} with respect to \mathbf{I} ; that is the number of neighbors of \mathbf{v} with no neighbors in \mathbf{I} .

Then as in Section 4, since G is triangle-free we have

$$(7) \quad \bar{\alpha}_G(\lambda) = \frac{\lambda}{1 + \lambda} \mathbb{E}(1 + \lambda)^{-\mathbf{Y}} \geq \frac{\lambda}{1 + \lambda} (1 + \lambda)^{-\mathbb{E}\mathbf{Y}},$$

where the inequality is an application of Jensen's inequality. Since G has maximum degree d we have

$$\begin{aligned}\bar{\alpha}_G(\lambda) &= \frac{1}{|V(G)|} \sum_{v \in V(G)} \Pr_{G, \lambda}[v \in \mathbf{I}] \\ &\geq \frac{1}{|V(G)|} \sum_{v \in V(G)} \frac{1}{d} \sum_{u \sim v} \Pr_{G, \lambda}[u \in \mathbf{I}] \\ (8) \quad &= \frac{1}{d} \frac{\lambda}{1 + \lambda} \mathbb{E}\mathbf{Y}.\end{aligned}$$

Combining (7) and (8), we get

$$\begin{aligned}\bar{\alpha}_G(\lambda) &\geq \frac{\lambda}{1 + \lambda} \max \left\{ (1 + \lambda)^{-\mathbb{E}\mathbf{Y}}, \frac{1}{d} \mathbb{E}\mathbf{Y} \right\} \\ &\geq \frac{\lambda}{1 + \lambda} \min_{y \geq 0} \max \left\{ (1 + \lambda)^{-y}, \frac{y}{d} \right\}.\end{aligned}$$

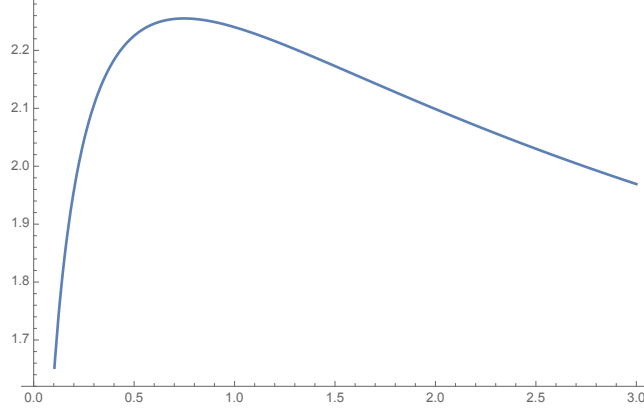


FIGURE 6. $\frac{\lambda}{1+\lambda}y^*$ as a function of λ with $d = 100$.

Now since $(1 + \lambda)^{-y}$ is a decreasing function and $\frac{y}{d}$ is an increasing function, we have

$$\bar{\alpha}_G(\lambda) \geq \frac{\lambda}{1 + \lambda} \frac{y^*}{d}$$

where y^* is the solution to the equation

$$(1 + \lambda)^{-y} = \frac{y}{d}$$

or,

$$y \cdot e^{y \log(1+\lambda)} = d.$$

The solution is

$$y^* = \frac{W(d \log(1 + \lambda))}{\log(1 + \lambda)}$$

where $W(\cdot)$ is the W-Lambert function. This gives

$$(9) \quad \bar{\alpha}_G(\lambda) \geq \frac{1}{d} \frac{\lambda}{1 + \lambda} \frac{W(d \log(1 + \lambda))}{\log(1 + \lambda)}.$$

Now although $\bar{\alpha}_G(\lambda)$ is monotone increasing in λ , somewhat surprisingly the bound (9) is not monotone in λ (see Figure 6 for example).

It turns out that it is best to take $\lambda = \lambda(d) \rightarrow 0$ as $d \rightarrow \infty$, but not as quickly as any polynomial, that is $\lambda(d) = \omega(d^{-\epsilon})$ for every $\epsilon > 0$.

We set $\lambda = 1/\log d$ and derive a bound asymptotically in d . We show in the exercises that the Lambert function satisfies

$$W(x) = \log(x) - \log \log(x) + o(1)$$

as $x \rightarrow \infty$. If $\lambda \rightarrow 0$ then $\frac{\lambda}{(1+\lambda)\log(1+\lambda)} \rightarrow 1$, and $W(d \log(1 + \lambda)) = (1 + o_d(1)) \log d$. This gives, for $\lambda = 1/\log d$,

$$\bar{\alpha}_G(\lambda) \geq (1 + o_d(1)) \frac{\log d}{d},$$

and by monotonicity this extends to all larger λ . □

5.1.1. *Tightness.* It is not known that Shearer's Theorem 18 is tight: the best we know is that a random d -regular graph has independence number $(2 + o_d(1)) \frac{\log d}{d} n$ whp. Theorem 19, however, is asymptotically tight in d . The random d -regular graph has *average* independent set size $(1 + o_d(1)) \frac{\log d}{d} n$ whp at $\lambda = 1$. In fact Theorem 19 is tight over a wider range of λ , all $\lambda = O_d(1)$. If we set $\lambda = d^c$ for c ranging from -1 to 1 , then the random d -regular graph has occupancy fraction $(1 + c) \frac{\log d}{d} + o(1)$ whp as $n \rightarrow \infty$. On the other hand, our lower bound states that for all triangle-free G of max degree d , and all $c \in (-1, 0]$, $\bar{\alpha}_G(\lambda) \geq (1 + c + o_d(1)) \frac{\log d}{d}$. A natural question is whether the lower bound can be extended to $c \in (0, 1]$.

5.1.2. *On the number of independent sets in triangle-free graphs.*

Theorem 20. *For any triangle-free graph G on n vertices with maximum degree at most d ,*

$$i(G) \geq e^{\left(\frac{1}{2} + o_d(1)\right) \frac{\log^2 d}{d} n}.$$

Proof. We integrate the bound (9) for $\lambda = 0$ to 1 to obtain a lower bound on the partition function.

$$\begin{aligned} \frac{1}{n} \log i(G) &= \frac{1}{n} \log Z_G(1) = \int_0^1 \frac{\bar{\alpha}_G(t)}{t} dt \\ &\geq \int_0^1 \frac{1}{d} \frac{1}{1+t} \frac{W(d \log(1+t))}{\log(1+t)} dt \quad \text{from (9)} \\ &= \frac{1}{d} \int_0^{W(d \log 2)} \frac{1}{1+u} du \quad \text{using the substitution } u = W(d \log(1+t)) \\ &= \frac{1}{d} \left[W(d \log 2) + \frac{1}{2} W(d \log 2)^2 \right] \\ &= \left(\frac{1}{2} + o_d(1) \right) \frac{\log^2 d}{d}. \end{aligned}$$

□

Theorem 20 is tight asymptotically in d , using the random d -regular graph as an example again.

Using a similar argument to the proof of the $R(3, k)$ upper bound, we can use Theorem 20 to give a lower bound on the number of independent sets in a triangle-free graph without degree restrictions.

Corollary 21. *For any triangle-free graph G on n vertices,*

$$i(G) \geq e^{\left(\frac{\sqrt{2 \log 2}}{4} + o(1)\right) \sqrt{n} \log n}.$$

Proof. Suppose the maximum degree of G is equal to d . Then $i(G) \geq 2^d$ since we can simply take all subsets of the neighborhood of the vertex with largest degree, and $i(G) \geq e^{\left(\frac{1}{2} + o_d(1)\right) \frac{\log^2 d}{d} n}$ from Theorem 20. As the first lower bound is increasing in d and the second is decreasing in d , we have

$$i(G) \geq \min_d \max \left\{ 2^d, e^{\left(\frac{1}{2} + o_d(1)\right) \frac{\log^2 d}{d} n} \right\} = 2^{d^*}$$

where d^* is the solution to $2^d = e^{(\frac{1}{2} + o_d(1)) \frac{\log^2 d}{d} n}$, that is,

$$d^* = (1 + o_d(1)) \frac{\sqrt{2} \sqrt{n} \log n}{4 \sqrt{\log 2}},$$

and so

$$i(G) \geq e^{\left(\frac{\sqrt{2} \log 2}{4} + o(1)\right) \sqrt{n} \log n}.$$

□

5.1.3. *Max vs. average independent set size?* Theorem 19 implies the bound (6) in exactly the same way as Shearer's bound, as the occupancy fraction is of course a lower bound on the independence ratio. But we might hope that it gives more – that in triangle-free graphs there is a significant gap between the independence number and the size of a uniformly random independent set (i.e. at $\lambda = 1$ in the hard-core model).

Open Problem. *Use Theorem 19 to improve the current asymptotic upper bound on $R(3, k)$.*

We give three specific conjectures whose resolution would improve the bound.

Conjecture 22 ([23]). *For any triangle-free graph G , we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \bar{\alpha}_G(1)} \geq 4/3.$$

Conjecture 23 ([23]). *For any triangle-free graph G of minimum degree d , we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \bar{\alpha}_G(1)} \geq 2 - o_d(1).$$

Conjecture 24 ([23]). *For any $\epsilon > 0$, there is λ small enough so that for any triangle-free graph G we have*

$$\frac{\alpha(G)}{|V(G)| \cdot \bar{\alpha}_G(\lambda)} \geq 2 - \epsilon.$$

Conjecture 22 would imply a factor 4/3 improvement on the current upper bound for $R(3, k)$, while Conjectures 23 and 24 would both imply a factor 2 improvement.

Open Problem. *Shearer [50] proved that any K_r -free graph G of average degree d on n vertices has independence number*

$$\alpha(G) \geq (C(r) + o_d(1)) \frac{\log d}{d \cdot \log \log d} \cdot n,$$

for some constant $C(r) > 0$. It is not known whether the $\log \log d$ in the denominator is necessary or not.

5.2. **Sphere packing densities.** In this section we return to hard spheres and suggest an analogy between Ramsey numbers and sphere packing densities. We then make the analogy partially rigorous by using the occupancy method in continuous space to prove a lower bound on the asymptotics of the maximum sphere packing density as the dimension tends to infinity.

5.2.1. *Sphere packing densities in Euclidean space.* Let $B_r(x)$ be the ball of radius r around $x \in \mathbb{R}^d$.

Definition 25. *The maximum sphere packing density of d -dimensional Euclidean space, $\theta(d)$, is*

$$\theta(d) = \sup_{\mathcal{P}} \limsup_{R \rightarrow \infty} \frac{\text{vol}(\mathcal{P} \cap B_R(0))}{\text{vol}(B_R(0))},$$

where the supremum is over all sphere packings \mathcal{P} of equal-sized spheres.

Of course $\theta(1) = 1$. $\theta(2) = \pi/\sqrt{12} = .9068\dots$ with the packing given by the hexagonal lattice and this was proved by Thue in 1894. $\theta(3) = \pi/\sqrt{18} = .7404\dots$ with the packing given by stacking hexagonal packings (exactly how you might try to stack oranges); this was Kepler's Conjecture and it was only proved in 2005 by Thomas Hales [31].

In breakthrough 2016 paper Maryna Viazovska proved that $\theta(8)$ is achieved by the E_8 lattice [57], and along with collaborators quickly proved that $\theta(24)$ is given by the Leech lattice [15]. All other dimensions are currently unknown.

What about optimal sphere packings in very high dimensions? Almost nothing is known! We do not know if the optimal packings are lattice packings or disordered. And our upper and lower bounds on $\theta(d)$ as $d \rightarrow \infty$ are very far apart.

A lower bound of $\theta(d) \geq 2^{-d}$ is trivial.

Proposition 26. *In all dimensions $\theta(d) \geq 2^{-d}$.*

Proof. Take any saturated (maximal) sphere packing, and double all the radii; because the original packing was saturated, the doubled balls must cover space (or else there would have been space for another center). Since the fraction of space covered increases by at most a factor 2^d , the original packing must have covered at least a 2^{-d} fraction of space. \square

Compare Proposition 26 (and its proof) to the following bound on the independence number of a graph. (We use D here to distinguish vertex degree from dimension of Euclidean space).

Proposition 27. *For all graphs G of maximum degree D , $\alpha(G) \geq |V(G)|/(D+1)$.*

Proof. Take any maximal independent set I of G . Let $B(I)$ be the set of vertices in I and their neighbors. Because I was maximal, we must have $B(I) = V(G)$. And because G has maximum degree D , $|B(I)| \leq (D+1)|I|$, and so $|I| \geq |V(G)|/(D+1)$. \square

This suggests an analogy between independent sets and sphere packings with the maximum degree D of a graph equivalent in some sense to the size of the excluded neighborhood of a center of a sphere packing (2^d times the volume of a sphere). Of course the centers of a sphere packing are in fact an independent set in the infinite graph with vertex set \mathbb{R}^d in which two vertices are joined if their distance is at most $2r_d$.

The 2^{-d} bound has been improved by a factor of d by Rogers [47], with subsequent constant factor improvements by Rogers and Davenport [19], Ball [3], Vance [55], and finally Venkatesh [56] who proved that $\theta(d) \geq (65963 + o_d(1))d \cdot 2^{-d}$. Venkatesh also gains an additional $\log \log d$ factor in a sparse sequence of dimensions.

An upper bound of $2^{-(.599\dots+o_d(1))\cdot d}$ was proved by Kabatiansky and Levenshtein [35]; there is a recent constant factor improvement by Cohn and Zhao [16].

The lower bounds mentioned above in fact show the existence of a lattice packing of the given density. This is clearly a stronger result than the existence of some packing, but it is not clear that by considering lattice packings only we will be able to close the gap in the bounds. There is a proof of $\theta(d) \geq .01 \cdot d2^{-d}$ by Krivelevich, Litsyn, and Vardy [38] using graph theory and the above mentioned result of Ajtai, Komlós, and Szemerédi [1], strengthening this analogy between Ramsey theory and sphere packing.

Open Problem. *Improve the exponential order of the upper or lower bound on $\theta(d)$.*

5.2.2. *A lower bound on the occupation density of the hard sphere model.* Recall the hard sphere model on a bounded, measurable set $S \subset \mathbb{R}^d$. It is a random set \mathbf{X} of centers in S at pairwise distance at least $2r_d$. Its partition function is

$$Z_S(\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \int_{S^k} \mathbf{1}_{\mathcal{D}(x_1, \dots, x_k)} dx_1 \cdots dx_k.$$

The analogue of the occupancy fraction is the expected number of centers per unit of volume.

Definition 28. *The occupation density of the hard sphere model on S at fugacity λ is*

$$\alpha_S(\lambda) = \frac{1}{\text{vol}(S)} \mathbb{E}_{S, \lambda} |\mathbf{X}|.$$

Theorem 29 (Joos, Jenssen, P. [34]). *Let B_n be the ball of volume n in \mathbb{R}^d . Then for $\lambda \geq 3^{-d/2}$,*

$$\alpha_{B_n}(\lambda) \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d}.$$

The corresponding bound $\theta(d) \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d}$ follows:

Lemma 30. *Let B_n be the ball of volume n in \mathbb{R}^d . Then for any $\lambda > 0$,*

$$\theta(d) \geq \limsup_{n \rightarrow \infty} \alpha_{B_n}(\lambda).$$

The proof of this lemma uses the fact that volume grows subexponentially fast in \mathbb{R}^d .

5.2.3. *Proof of Theorem 29.* Given a set of centers X in S , we can partition S into three sets: points that are covered, blocked, and free. A point $x \in S$ is covered if $d(x, X) \leq r_d$; it is blocked if $d(x, X) \in (r_d, 2r_d]$ and free if $d(x, X) > 2r_d$.

Definition 31. *The expected free volume of the hard sphere model on S is*

$$\text{FV}_S(\lambda) = \frac{1}{\text{vol}(S)} \int_S \Pr_{S, \lambda}[d(x, \mathbf{X}) > 2r_d] dx.$$

That is, $\text{FV}_S(\lambda)$ is the expected fraction of points in S that could be added to \mathbf{X} and still result in a packing.

Lemma 32.

$$\alpha_S(\lambda) = \lambda \cdot \text{FV}_S(\lambda).$$

Proof. □

Now consider the following two-part experiment. Pick \mathbf{X} from the hard sphere model on S and choose $\mathbf{v} \in S$ uniformly at random. Let \mathbf{T}_S be the externally uncovered volume in the $2r_d$ neighborhood of \mathbf{v} ; that is,

$$\mathbf{T}_S = \{x \in B_{2r_d}(\mathbf{v}) : d(x, y) > 2r_d \forall y \in \mathbf{X} \cap B_{2r_d}(\mathbf{v})^c\}.$$

Note that only centers outside of the $2r_d$ ball around \mathbf{v} affect the set \mathbf{T}_S .

Lemma 33. *Let S be bounded and measurable, and consider the above two-part experiment. Then*

$$(1) \alpha_S(\lambda) = \lambda \mathbb{E}_{S, \lambda} \left[\frac{1}{Z_{\mathbf{T}_S}(\lambda)} \right].$$

$$(2) \alpha_S(\lambda) \geq 2^{-d} \lambda \mathbb{E}_{S, \lambda} \left[\frac{\lambda Z'_{\mathbf{T}_S}(\lambda)}{Z_{\mathbf{T}_S}(\lambda)} \right].$$

Proof. □

Lemma 34. *Let S be bounded and measurable.*

$$(1) \log Z_S(\lambda) \leq \lambda \cdot \text{vol}(S).$$

$$(2) \alpha_S(\lambda) \geq \lambda \cdot e^{-\lambda \mathbb{E}_{S, \lambda} \text{vol}(\mathbf{T}_S)}.$$

Finally we need a simple geometric inequality about spheres in \mathbb{R}^d .

Lemma 35. *Let $S \subseteq B_{2r_d}(0)$ be measurable. Then*

$$\mathbb{E}[\text{vol}(B_{2r_d}(\mathbf{u}) \cap S)] \leq 2 \cdot 3^{d/2},$$

where \mathbf{u} is chosen uniformly from S .

Proof. □

Proof of Theorem 29. Let $B_n = B_{n^{1/d} r_d}(0)$. Let $\alpha_n = \alpha_{B_n}(\lambda)$.

We have

$$\alpha_n = \lambda \cdot \mathbb{E}_{B_n, \lambda} \left[\frac{1}{Z_{\mathbf{T}_{B_n}}(\lambda)} \right] \geq \lambda \cdot e^{-\lambda \mathbb{E}_{B_n, \lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda)}.$$

On the other hand,

$$\begin{aligned} \alpha_n &\geq 2^{-d} \mathbb{E}_{B_n, \lambda} \left[\frac{\lambda Z'_{\mathbf{T}_{B_n}}(\lambda)}{Z_{\mathbf{T}_{B_n}}(\lambda)} \right] \\ &= 2^{-d} \mathbb{E}_{B_n, \lambda} [\text{vol}(\mathbf{T}_{B_n}) \cdot \alpha_{\mathbf{T}_{B_n}}(\lambda)] \\ &\geq 2^{-d} \mathbb{E}_{B_n, \lambda} \left[\lambda \cdot \text{vol}(\mathbf{T}_{B_n}) \cdot e^{-\lambda \mathbb{E}_{\mathbf{T}_{B_n}} \text{vol}(\mathbf{U})} \right] \text{ by Lemma 34 part 2} \\ &\geq 2^{-d} \mathbb{E}_{B_n, \lambda} \left[\log Z_{\mathbf{T}_{B_n}}(\lambda) \cdot e^{-\lambda \mathbb{E}_{\mathbf{T}_{B_n}} \text{vol}(\mathbf{U})} \right] \text{ by Lemma 34 part 1} \\ &\geq 2^{-d} \mathbb{E}_{B_n, \lambda} \left[\log Z_{\mathbf{T}_{B_n}}(\lambda) \cdot e^{-\lambda 2 \cdot 3^{d/2}} \right] \text{ by Lemma 35} \\ &= 2^{-d} \cdot e^{-\lambda 2 \cdot 3^{d/2}} \mathbb{E}_{B_n, \lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda). \end{aligned}$$

Now with $z = \mathbb{E}_{B_n, \lambda} \log Z_{\mathbf{T}_{B_n}}(\lambda)$, we have

$$\alpha_n \geq \inf_{z \geq 0} \max\{\lambda e^{-z}, z \cdot 2^{-d} e^{-\lambda 2 \cdot 3^{d/2}}\}.$$

As before, one expression is decreasing in z and the other increasing and so the infimum is achieved at

$$z^* = W\left(\lambda \cdot \left(2e^{-\lambda 2 \cdot 3^{d/2}}\right)^d\right).$$

Now take $\lambda = d^{-2} 3^{d/2}$. Then

$$\begin{aligned} z^* &= W(\lambda 2^d e^{2/d}) \\ &= \log \lambda + d \log 2 - \log d - \log \log(2/\sqrt{3}) + o_d(1). \end{aligned}$$

This gives

$$\alpha_n \geq (1 + o_d(1)) \frac{\log(2/\sqrt{3}) \cdot d}{2^d}.$$

□

Further reading. Henry Cohn has lecture notes on sphere packings [14].

Exercises.

- (1) Show that $R(3, 3) = 6$.
- (2) Compute the asymptotics of $W(x)$ as $x \rightarrow \infty$ to the first two terms.
- (3) Find a triangle-free graph G for which $\frac{\alpha(G)}{\alpha_G(1)} < 3/2$.

6. GENERALIZATIONS AND EXTENSIONS

6.1. Adding local constraints.

6.1.1. *The Petersen graph.* The *Petersen graph*, $P_{5,2}$, has 10 vertices, is 3-regular and vertex-transitive, has girth 5 and is a $(3, 5)$ -Moore graph (see Figure 1). Its independence polynomial is

$$Z_{P_{5,2}}(\lambda) = 1 + 10\lambda + 30\lambda^2 + 30\lambda^3 + 5\lambda^4,$$

and its occupancy fraction is

$$\bar{\alpha}_{P_{5,2}}(\lambda) = \frac{\lambda(1 + 6\lambda + 9\lambda^2 + 2\lambda^3)}{Z_{P_{5,2}}(\lambda)}.$$

Our first result provides a tight lower bound on the occupancy fraction of triangle-free cubic graphs for every $\lambda \in (0, 1]$:

Theorem 36 (Perarnau, P. [42]). *For any triangle-free, cubic graph G , and for every $\lambda \in (0, 1]$,*

$$\bar{\alpha}_G(\lambda) \geq \bar{\alpha}_{P_{5,2}}(\lambda),$$

with equality if and only if G is a union of copies of $P_{5,2}$.

By integrating $\frac{\bar{\alpha}_G(\lambda)}{\lambda}$ from $\lambda = 0$ to 1 we obtain the corresponding counting result:

Corollary 37. *For any triangle-free, cubic graph G , and any $\lambda \in (0, 1]$,*

$$\frac{1}{|V(G)|} \log Z_G(\lambda) \geq \frac{1}{10} \log Z_{P_{5,2}}(\lambda),$$

and in particular,

$$\frac{1}{|V(G)|} \log i(G) \geq \frac{1}{10} \log i(P_{5,2}),$$

with equality if and only if G is a union of copies of $P_{5,2}$.

6.1.2. *The Heawood graph.* The Heawood graph, $H_{3,6}$, has 14 vertices, is 3-regular and vertex-transitive, has girth 6, and is a $(3, 6)$ -Moore graph (see Figure 1). It can be constructed as the point-line incidence graph of the Fano plane. Its independence polynomial is

$$Z_{H_{3,6}}(\lambda) = 1 + 14\lambda + 70\lambda^2 + 154\lambda^3 + 147\lambda^4 + 56\lambda^5 + 14\lambda^6 + 2\lambda^7,$$

and its occupancy fraction is

$$\bar{\alpha}_{H_{3,6}}(\lambda) = \frac{\lambda(1 + 10\lambda + 33\lambda^2 + 42\lambda^3 + 20\lambda^4 + 6\lambda^5 + \lambda^6)}{Z_{H_{3,6}}(\lambda)}.$$

Our second result provides a tight upper bound on the occupancy fraction of cubic graphs with girth at least 5:

Theorem 38 (Perarnau, P. [42]). *For any cubic graph G of girth at least 5, and for every $\lambda > 0$,*

$$\bar{\alpha}_G(\lambda) \leq \bar{\alpha}_{H_{3,6}}(\lambda),$$

with equality if and only if G is a union of copies of $H_{3,6}$.

And by integrating $\frac{\bar{\alpha}_G(t)}{t}$ from $t = 0$ to λ we obtain the corresponding counting results.

Corollary 39. *For any cubic graph G of girth at least 5, and for every $\lambda > 0$,*

$$\frac{1}{|V(G)|} \log Z_G(\lambda) \leq \frac{1}{14} \log Z_{H_{3,6}}(\lambda),$$

and in particular,

$$\frac{1}{|V(G)|} \log i(G) \leq \frac{1}{14} \log i(H_{3,6}),$$

with equality if and only if G is a union of copies of $H_{3,6}$.

Note that Theorem 38 applies to all positive λ , while Theorem 36 requires $\lambda \in (0, 1]$. Some bound on the interval for which $P_{5,2}$ minimizes the occupancy fraction is necessary: for large λ , $P_{7,2}$ has a smaller occupancy fraction, and in fact in the limit as $\lambda \rightarrow \infty$, it is minimal: Staton [53] proved the independence ratio of any triangle-free cubic graph is at least $5/14$ and this is achieved by $P_{7,2}$.

Open Problem. *Can Corollaries 37 or 39 be proved with the entropy method?*

6.2. **Colorings.** Let $N_q(G)$ be the number of proper q -colorings of a graph G .

Conjecture 40 (Galvin, Tetali [30]). *For any d -regular G and any $q \geq 2$,*

$$\frac{1}{|V(G)|} \log N_q(G) \leq \frac{1}{2d} \log N_q(K_{d,d}).$$

Galvin and Tetali proved the conjecture under the condition that G is bipartite. The cases $d = 2$ or $q = 2$ are easy exercises.

The next result resolves the conjecture for the case $d = 3$.

Theorem 41 (Davies, Jenssen, Perkins, Roberts [21]). *For any 3-regular G and any $q \geq 2$,*

$$\frac{1}{|V(G)|} \log N_q(G) \leq \frac{1}{6} \log N_q(K_{3,3}).$$

To prove Theorem 41, we proceed indirectly, via the Potts model.

6.2.1. *Optimizing the internal energy of the Potts model.* Recall the Potts model from Section 2.4, a random q -coloring of the vertices of G chosen according to

$$\Pr_{G,\beta}[\chi = \chi] = \frac{e^{-\beta m(G,\chi)}}{Z_G^q(\beta)}.$$

Let $U_G^q(\beta)$ be the *internal energy* (per edge) of the Potts model, or the expected fraction of monochromatic edges of G under the random coloring:

$$U_G^q(\beta) = \frac{1}{|E(G)|} \mathbb{E}_{G,\beta}[m(G, \chi)].$$

Up to scaling, the internal energy is (minus) the derivative of the free energy:

$$(10) \quad U_G^q(\beta) = -\frac{|V(G)|}{|E(G)|} \left(\frac{1}{|V(G)|} \log Z_G^q(\beta) \right)'.$$

Below we will show that for cubic graphs, $K_{3,3}$ minimizes the internal energy of the anti-ferromagnetic Potts model.

Theorem 42 ([21]). *For any 3-regular G , any $q \geq 2$, and any $\beta > 0$,*

$$U_G^q(\beta) \geq U_{K_{3,3}}^q(\beta).$$

By integrating this bound from 0 to β we immediately obtain the corresponding bound on the free energy (the direction of the inequality flips because of the minus sign in (10)):

$$\frac{1}{|V(G)|} \log Z_G^q(\beta) \leq \frac{1}{6} \log Z_{K_{3,3}}^q(\beta),$$

then sending $\beta \rightarrow \infty$ gives Theorem 41.

To prove Theorem 42 we first define a local view of the two-part experiment of drawing χ from the Potts model on a d -regular graph and choose v uniformly from $V(G)$. We record the graph structure at depth 1, that is the neighbors of v and any incident edges, and for each neighbor of v we record the multiset of colors it sees externally under χ (that is, the colors assigned to its neighbors outside of $v \cup N(v)$). See Figure 7 for a depiction of a local view.

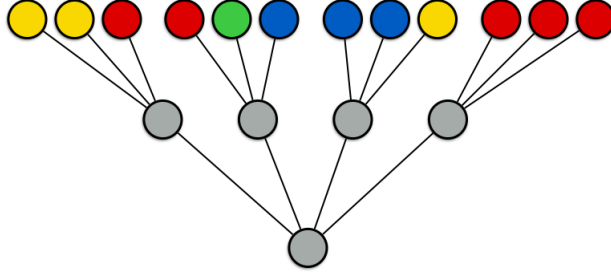


FIGURE 7. Local view for the Potts model program

Now we can add a family of consistency constraints. Let $\mathcal{S}_{q,d}$ be the set of all q -partitions of size d ; that is, non-negative integer vectors that sum to at most d . To take advantage of symmetries, we regard these vectors as unordered. A q -coloring χ of d vertices induces a q -partition; for instance if $q = 4$, $d = 6$, and χ assigns colors $\{3, 3, 2, 4, 1, 2\}$ to the d vertices, then the q -partition $S(\chi) = \{2, 2, 1, 1\} \in \mathcal{S}_{4,6}$. Our constraints will insist that for every $S \in \mathcal{S}_{q,d}$, the probability that the q -partition induced by χ on the neighbors of \mathbf{v} is S equals the average probability of the same for a neighbor of \mathbf{v} .

Both of these probabilities can be computed as expectations over the random local view. For a local view C and a q -partition $S \in \mathcal{S}_{q,d}$ we define

$$\begin{aligned} \gamma_C^{v,S} &:= \frac{1}{Z_C} \sum_{\chi \in [q]^{V_C}} \mathbf{1}_{\{H(\chi(N(v)))=S\}} \cdot e^{-\beta m(\chi)}, \\ \gamma_C^{N,S} &:= \frac{1}{d} \frac{1}{Z_C} \sum_{\chi \in [q]^{V_C}} \sum_{u \in N(v)} \mathbf{1}_{\{H(\chi(N(u)))=S\}} \cdot e^{-\beta m(\chi)}. \end{aligned}$$

Observe that for any graph and any q -partition S , we must have

$$\mathbb{E}_C[\gamma_C^{v,S}] = \mathbb{E}_C[\gamma_C^{N,S}].$$

Our minimization program becomes

$$(11) \quad \begin{aligned} &\text{Minimize } \sum_C p_C U_C^v && \text{subject to} \\ &p_C \geq 0 \quad \forall C, \\ &\sum_C p_C = 1, \\ &\sum_C p_C (\gamma_C^{v,S} - \gamma_C^{N,S}) = 0 \quad \text{for all } S \in \mathcal{S}_{q,d}. \end{aligned}$$

Open Problem. Prove that the optimal solution to the linear program (11) is given by the distribution induced by $K_{d,d}$ for all $q \geq d+1$ and all $\beta > 0$. This would prove Conjecture 40 for $q \geq d+1$.

We have found some examples with $q \leq d$ and β large enough for which the optimal solution to the linear program above is not the distribution induced by $K_{d,d}$ so we do not expect the program to be tight in those cases. Another open problem is to find additional constraints that show $K_{d,d}$ is optimal for $q \leq d$.

Exercises.

- (1) Compute $Z_{K_{d,d}}^q(\beta)$.
- (2) Come up with an alternative approach to proving Conjecture 40 by looking at random partial colorings (no monochromatic edges allowed, but some vertices can be left uncolored) instead of the Potts model.

7. FURTHER DIRECTIONS AND OPEN PROBLEMS

7.1. Independent sets and matchings of a given size. Instead of asking for maximality of a graph polynomial, like $Z_G(\lambda)$ or Z_G , we could ask for the stronger statement that a given graph maximizes each of the individual coefficients; that is, independent sets or matchings of a given size.

Conjecture 43 (Kahn [36]). *Let $2d$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/2$,*

$$i_k(G) \leq i_k(H_{d,n}).$$

Conjecture 44 (Friedland, Krop, Markstrom [26]). *Let $2d$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/2$,*

$$m_k(G) \leq m_k(H_{d,n}).$$

The case $k = n/2$ concerns perfect matchings and is a special case of Bregman's theorem. The best current progress on these conjectures are the following approximate results from [20]:

Theorem 45. *For every $\epsilon > 0$, $d > 0$, there exists n_0 large enough so that the following is true for all $n \geq n_0$. Let $2d$ divide n . Then for any d -regular G on n vertices, and any $\epsilon n \leq k \leq n/2$,*

$$\begin{aligned} i_k(G) &\leq i_k(H_{d,n}), \\ m_k(G) &\leq m_k(H_{d,n}). \end{aligned}$$

Here we give a short proof of a weaker approximate result that follows directly from the partition function result. We give the proof for matchings but the same proof works for independent sets as well.

Proposition. *Let $2d$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/2$,*

$$m_k(G) \leq n \cdot m_k(H_{d,n}).$$

Proof. Choose λ so that $\mathbb{E}_{H_{d,n},\lambda}|\mathbf{M}| = k$. Now since the size of the random matching drawn from $H_{d,n}$ is a log-concave probability distribution (this is true for the size of a matching drawn from any graph; it is also true for the size of an independent set drawn from $H_{d,n}$), we have that the mean is also the mode of the distribution, and so $\mathbb{P}_{H_{d,n},\lambda}[|\mathbf{M}| = k] \geq 1/n$. This gives

$$\begin{aligned} \lambda^k m_k(H_{d,n}) &\geq \frac{1}{n} Z_{H_{d,n}}^{\text{match}}(\lambda) \\ &\geq \frac{1}{n} Z_G^{\text{match}}(\lambda) \\ &\geq \frac{1}{n} \lambda^k i_k(G). \end{aligned}$$

Dividing through by λ^k completes the proof. \square

7.1.1. *Free volume.* Going back to the idea of free volume from Section 5.2, we give two conjectures are stronger than Conjectures 43 and 44 but also have a nice probabilistic interpretation that may help in proving them.

Given a graph G and an integer $1 \leq k \leq \alpha(G)$, let \mathbf{I}_k be a uniformly chosen independent set of G of size exactly k . Let $\text{FV}_G(k)$ be the expected number of free vertices with respect to \mathbf{I}_k ; that is, the expected number of vertices of V that are neither in \mathbf{I}_k nor neighboring some $u \in \mathbf{I}_k$.

$$\text{FV}_G(k) = \frac{1}{i_k(G)} \sum_{I \in \mathcal{I}(G): |I|=k} |\{u : d(u, I) \geq 2\}|$$

Since each independent set of size $k + 1$ contains exactly $k + 1$ independent sets of size k and the free vertices with respect to \mathbf{I}_k are exactly the vertices that can be added to form an independent set of size $k + 1$, we have

$$\text{FV}_G(k) = \frac{(k + 1)i_{k+1}(G)}{i_k(G)}.$$

We can define the analogous quantity $\text{FV}_G^{\text{match}}(k)$ for matchings.

Theorems 12 and 13 show in fact that when \mathbf{I} or \mathbf{M} are chosen from the hard-core and monomer-dimer models respectively, $K_{d,d}$ maximizes the expected fraction of free vertices and edges (as the expected fraction of free vertices is simply $\bar{\alpha}_G(\lambda)/\lambda$). We conjecture the same is true for \mathbf{I}_k and \mathbf{M}_k .

Conjecture 46. *Let $2d$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/2$,*

$$\frac{i_{k+1}(G)}{i_k(G)} \leq \frac{i_{k+1}(H_{d,n})}{i_k(H_{d,n})}.$$

Conjecture 47. *Let $2d$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/2$,*

$$\frac{m_{k+1}(G)}{m_k(G)} \leq \frac{m_{k+1}(H_{d,n})}{m_k(H_{d,n})}.$$

Conjectures 46 and 47 are of course stronger than Conjectures 43 and 44: we obtain the latter by successively multiplying.

If we ask to minimize the number of independent sets in a regular graph, then the entire hierarchy of statements is known: the minimizer is a union of K_{d+1} 's, cliques on $d + 1$ vertices.

Theorem 48 (Cutler, Radcliffe [18]). *Let $d + 1$ divide n . Then for any d -regular G on n vertices, and any $1 \leq k \leq n/(d + 1)$,*

$$\frac{i_{k+1}(G)}{i_k(G)} \geq \frac{i_{k+1}(CL_{n,d})}{i_k(CL_{n,d})}$$

where $CL_{n,d}$ is the union of $n/(d + 1)$ disjoint cliques on $d + 1$ vertices. As a consequence, for any d -regular graph G ,

$$\frac{1}{|V(G)|} \log i(G) \geq \frac{1}{d + 1} \log i(K_{d+1}) = \frac{\log(d + 2)}{d + 1}.$$

From the perspective of free volume the proof is a one sentence argument: in a d -regular graph each vertex in an independent set excludes d others, and in a union of K_{d+1} 's, there is no overlap of excluded vertices, and so the expected excluded volume when drawing a uniformly random independent set of size $k - 1$ is maximized by $CL_{n,d}$.

[26, 25, 9, 43]

7.2. General results and conjectures on graph homomorphisms.

Exercises.

- (1) Let a_0, a_1, \dots, a_n and b_0, b_1, \dots, b_n be two sequences of non-negative integers with $a_0 = b_0 = 1$. Let $A(x) = \sum_{k=0}^n a_k x^k$ and $B(x) = \sum_{k=0}^n b_k x^k$.
 - (a) Suppose $\frac{a_k}{a_{k-1}} \geq \frac{b_k}{b_{k-1}}$ for all $k = 1 \dots n$. Show that $\frac{A'(x)}{A(x)} \geq \frac{B'(x)}{B(x)}$ for all $x \geq 0$.
 - (b) Show that the statements ' $a_k \geq b_k$ for all $k = 1 \dots n$ ' and ' $\frac{A'(x)}{A(x)} \geq \frac{B'(x)}{B(x)}$ for all $x \geq 0$ ' are incomparable in general.

ACKNOWLEDGEMENTS

Thanks to Agelos Georgakopoulos and Dimitris Cheliotis for organizing the summer school and to all the participants at the National and Kapodistrian University of Athens for their enthusiastic participation.

Thanks to Ewan Davies, Matthew Jenssen, Barnaby Roberts, Emma Cohen, Prasad Tetali, and Guillem Perarnau who collaborated on many of the results presented here. Thanks to David Galvin, Peter Csikvári, and Yufei Zhao for their helpful and interesting conversations.

REFERENCES

- [1] M. Ajtai, J. Komlós, and E. Szemerédi. A note on Ramsey numbers. *Journal of Combinatorial Theory, Series A*, 29(3):354–360, 1980.
- [2] N. Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel Journal of Mathematics*, 73(2):247–256, 1991.
- [3] K. Ball. A lower bound for the optimal density of lattice packings. *International Mathematics Research Notices*, 1992(10):217–221, 1992.
- [4] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6, 2001.
- [5] T. Bohman and P. Keevash. Dynamic concentration of the triangle-free process. *arXiv preprint arXiv:1302.5963*, 2013.
- [6] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [7] L. Bregman. Some properties of nonnegative matrices and their permanents. *Soviet Math. Dokl*, 14(4):945–949, 1973.
- [8] G. Brightwell and P. Winkler. Hard constraints and the Bethe lattice: adventures at the interface of combinatorics and statistical physics. In *Proc. Int'l. Congress of Mathematicians*, volume III, pages 605–624, 2002.
- [9] T. Carroll, D. Galvin, and P. Tetali. Matchings and independent sets of a fixed size in regular graphs. *Journal of Combinatorial Theory, Series A*, 116(7):1219–1227, 2009.
- [10] M. Chudnovsky and P. Seymour. The roots of the independence polynomial of a clawfree graph. *Journal of Combinatorial Theory, Series B*, 97(3):350–357, 2007.
- [11] F. R. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Series A*, 43(1):23–37, 1986.
- [12] E. Cohen, P. Csikvári, W. Perkins, and P. Tetali. The Widom–Rowlinson model, the hard-core model and the extremality of the complete graph. *European Journal of Combinatorics*, 62:70–76, 2017.

- [13] E. Cohen, W. Perkins, and P. Tetali. On the Widom–Rowlinson occupancy fraction in regular graphs. *Combinatorics, Probability and Computing*, 26(2):183–194, 2017.
- [14] H. Cohn. Packing, coding, and ground states. *arXiv preprint arXiv:1603.05202*, 2016.
- [15] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. Viazovska. The sphere packing problem in dimension 24. *arXiv preprint arXiv:1603.06518*, 2016.
- [16] H. Cohn and Y. Zhao. Sphere packing bounds via spherical codes. *Duke Math. J.*, 163:1965–2002, 2014.
- [17] P. Csikvári. Extremal regular graphs: the case of the infinite regular tree. *arXiv preprint arXiv:1612.01295*, 2016.
- [18] J. Cutler and A. Radcliffe. The maximum number of complete subgraphs in a graph with given maximum degree. *Journal of Combinatorial Theory, Series B*, 104:60–71, 2014.
- [19] H. Davenport, C. Rogers, et al. Hlawka’s theorem in the geometry of numbers. *Duke Math. J.*, 14(2):367–375, 1947.
- [20] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Tight bounds on the coefficients of partition functions via stability. *arXiv preprint arXiv:1704.07784*, 2017.
- [21] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Extremes of the internal energy of the potts model on cubic graphs. *Random Structures & Algorithms*, to appear.
- [22] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Independent sets, matchings, and occupancy fractions. *Journal of the London Mathematical Society*, to appear.
- [23] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. On the average size of independent sets in triangle-free graphs. *Proceedings of the American Mathematical Society*, to appear.
- [24] G. Fiz Pontiveros, S. Griffiths, and R. Morris. The triangle-free process and $R(3, k)$. *arXiv preprint arXiv:1302.6279*, 2013.
- [25] S. Friedland, E. Krop, P. H. Lundow, and K. Markström. On the validations of the asymptotic matching conjectures. *Journal of Statistical Physics*, 133(3):513–533, 2008.
- [26] S. Friedland, E. Krop, and K. Markström. On the number of matchings in regular graphs. *the electronic journal of combinatorics*, 15(1):R110, 2008.
- [27] D. Galvin. Maximizing h -colorings of a regular graph. *Journal of Graph Theory*, 73(1):66–84, 2013.
- [28] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.
- [29] D. Galvin. Counting colorings of a regular graph. *Graphs and Combinatorics*, 31(3):629–638, 2015.
- [30] D. Galvin and P. Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.
- [31] T. C. Hales. A proof of the kepler conjecture. *Annals of mathematics*, 162(3):1065–1185, 2005.
- [32] O. J. Heilmann and E. H. Lieb. Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25(3):190–232, 1972.
- [33] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.
- [34] M. Jenssen, F. Joos, and W. Perkins. On the hard sphere model and sphere packings in high dimensions. *arXiv preprint arXiv:1707.00476*, 2017.
- [35] G. A. Kabatiansky and V. I. Levenshtein. On bounds for packings on a sphere and in space. *Problemy Peredachi Informatsii*, 14(1):3–25, 1978.
- [36] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(03):219–237, 2001.
- [37] J. H. Kim. The Ramsey number $R(3, t)$ has order of magnitude $t^2 / \log t$. *Random Structures & Algorithms*, 7(3):173–207, 1995.
- [38] M. Krivelevich, S. Litsyn, and A. Vardy. A lower bound on the density of sphere packings via graph theory. *International Mathematics Research Notices*, 2004(43):2271–2279, 2004.
- [39] L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Society Providence, 2012.
- [40] H. Löwen. Fun with hard spheres. In *Statistical physics and spatial statistics*, pages 295–331. Springer, 2000.
- [41] R. Peled and W. Samotij. Odd cutsets and the hard-core model on z^d . *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 50(3):975–998, 2014.
- [42] G. Perarnau and W. Perkins. Counting independent sets in cubic graphs of given girth. *arXiv preprint arXiv:1611.01474*, 2016.
- [43] W. Perkins. Birthday inequalities, repulsion, and hard spheres. *Proceedings of the American Mathematical Society*, 144(6):2635–2649, 2016.

- [44] R. B. Potts. Some generalized order-disorder transformations. *Mathematical proceedings of the Cambridge philosophical society*, 48(01):106–109, 1952.
- [45] J. Radhakrishnan. An entropy proof of Bregman’s theorem. *Journal of combinatorial theory, Series A*, 77(1):161–164, 1997.
- [46] A. A. Razborov. Flag algebras. *The Journal of Symbolic Logic*, 72(04):1239–1282, 2007.
- [47] C. A. Rogers. Existence theorems in the geometry of numbers. *Annals of Mathematics*, pages 994–1002, 1947.
- [48] L. Sernau. Graph operations and upper bounds on graph homomorphism counts. *arXiv preprint arXiv:1510.01833*, 2015.
- [49] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.
- [50] J. B. Shearer. On the independence number of sparse graphs. *Random Structures & Algorithms*, 7(3):269–271, 1995.
- [51] A. D. Sokal. Chromatic polynomials, Potts models and all that. *Physica A: Statistical Mechanics and its Applications*, 279(1):324–332, 2000.
- [52] J. Spencer. *Eighty Years of RamseyR(3, k) . . . and Counting!*, pages 27–39. Birkhäuser Boston, Boston, MA, 2011.
- [53] W. Staton. Some Ramsey-type numbers and the independence ratio. *Transactions of the American Mathematical Society*, 256:353–370, 1979.
- [54] J. Van den Berg and J. Steif. Percolation and the hard-core lattice gas model. *Stochastic Processes and their Applications*, 49(2):179–197, 1994.
- [55] S. Vance. Improved sphere packing lower bounds from hurwitz lattices. *Advances in Mathematics*, 227(5):2144–2156, 2011.
- [56] A. Venkatesh. A note on sphere packings in high dimension. *Int. Math. Res. Not.*, 2013(7):1628–1642, 2013.
- [57] M. S. Viazovska. The sphere packing problem in dimension 8. *Annals of Mathematics*, 185(3):991–1015, 2017.
- [58] D. J. Welsh and C. Merino. The Potts model and the Tutte polynomial. *Journal of Mathematical Physics*, 41(3):1127–1152, 2000.
- [59] F.-Y. Wu. The Potts model. *Reviews of modern physics*, 54(1):235, 1982.
- [60] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(02):315–320, 2010.
- [61] Y. Zhao. The bipartite swapping trick on graph homomorphisms. *SIAM Journal on Discrete Mathematics*, 25(2):660–680, 2011.
- [62] Y. Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *arXiv preprint arXiv:1610.09210*, 2016.

APPENDIX A. NOTATION

- $\theta(d)$: The maximum sphere packing density in \mathbb{R}^d .
- $\mathcal{T}_d(n)$: The d -dimensional torus of volume n .
- r_d : The radius of the ball of volume 1 in \mathbb{R}^d .
- $\mathcal{I}(G)$: The set of all independent sets of a graph G .
- $\mathcal{M}(G)$: The set of all matchings of a graph G .
- $i(G)$: The total number of independent sets of a graph G .
- $m(G)$: The total number of matchings of a graph G .
- $m_{\text{perf}}(G)$: The number of perfect matchings of G .
- $i_k(G)$: The number of independent sets of size exactly k in G .
- $m_k(G)$: The number of matchings of size exactly k in G .
- $Z_G(\lambda)$: The independence polynomial of a graph G or the partition function of the hard-core model.
- $Z_G^{\text{match}}(\lambda)$: The matching polynomial of a graph G or the partition function of the monomer-dimer model.

$Z_G^q(\beta)$: The q -color Potts model partition function on G at inverse temperature β .
 $\mathbb{E}_{G,\lambda}, \Pr_{G,\lambda}, \mathbf{I}, \mathbf{M}, \chi$: \mathbf{I} is a random independent set drawn from the hard-core model on G at fugacity λ . Likewise \mathbf{M} is a random matching drawn from the monomer-dimer model and χ a random q -coloring from the Potts model.
 $K_{d,d}$: The complete d -regular bipartite graph on $2d$ vertices.
 K_{d+1} : The complete graph on $d+1$ vertices.
 C_n : The cycle on n vertices.
 $\alpha(G)$: The independence number (size of largest independent set) of G .
 $\nu(G)$: The matching number (size of largest matching) of G .
 $\bar{\alpha}_G(\lambda)$: The hard-core occupancy fraction of a graph G at fugacity λ ; $\alpha_G(\lambda) = \frac{1}{|V(G)|} \mathbb{E}_{G,\lambda} |\mathbf{I}|$.
 $\bar{\alpha}_G^{\text{match}}(\lambda)$: The monomer-dimer occupancy fraction of a graph G at fugacity λ .
 $N_q(G)$: The number of proper q -colorings of G .

APPENDIX B. BACKGROUND MATERIAL

B.1. Statistical physics and probability.

B.2. Entropy.

Definition. Let X be a random variable taking values in a finite set Ω . Then the **entropy** of X is

$$H(X) = - \sum_{\omega \in \Omega} \mathbb{P}(X = \omega) \log \mathbb{P}(X = \omega).$$

It is immediate from the definition that entropy is maximized by the uniform distribution; that is,

$$H(X) \leq \log |\Omega|,$$

with equality if and only if X is uniformly distributed over Ω .

The *conditional entropy* of X given Y is

$$\begin{aligned} H(X|Y) &= \sum_y \mathbb{P}(Y = y) \cdot H(X|Y = y) \\ &= - \sum_{x,y} \Pr[X = x, Y = y] \log \frac{\Pr[X = x, Y = y]}{\Pr[Y = y]}. \end{aligned}$$

The *chain rule* for entropy states

$$H(X_1, \dots, X_n) = H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_1, \dots, X_{n-1}).$$

Conditioning can only reduce entropy:

$$H(X|Y) \leq H(X),$$

and so we have the subadditivity property of entropy:

$$H(X_1, \dots, X_n) \leq \sum_{i=1}^n H(X_i).$$

B.3. Linear programming. Let us review some basic facts about linear programming. (For much more on linear programming, see for example Boyd and Vandenberghe’s book [6].

Suppose we have the linear program in standard form with variables x, \dots, x_n :

$$\begin{aligned} & \text{maximize} && \sum_{i=1}^n c_i x_i \\ & \text{subject to} && x_i \geq 0 \quad \forall i \\ & && \sum_{i=1}^n A_{ij} x_i \leq b_j \quad \text{for } j = 1, \dots, m. \end{aligned}$$

This is the *primal LP*. The corresponding *dual LP* has variables $\Lambda_1, \dots, \Lambda_m$ for each constraint of the primal and constraints for each variable of the primal:

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^m b_j \Lambda_j \\ & \text{subject to} && \Lambda_j \geq 0 \quad \forall j \\ & && \sum_{j=1}^m A_{ij} \Lambda_j \geq c_i \quad \text{for } i = 1, \dots, n. \end{aligned}$$

Theorem (Strong duality theorem). *If the primal and dual linear programs have feasible solutions, then their objective values coincide.*

In particular, if we have a feasible primal solution that we believe is optimal, we can prove this by finding a feasible dual solution with the same objective value.

APPENDIX C. AN ALTERNATIVE PROOF OF THEOREM 12

Here we prove Theorem 12 with a different local view and linear program. In some sense this is a ‘level 0’ proof: the local view simply records the number of occupied neighbors of a random vertex, disregarding even the graph structure among the neighbors. The constraints in this proof are different in form than most of those given in the examples above.

Let G be a d -regular graph. Consider drawing an independent set I from G according to the hard-core model with fugacity λ , then choosing a vertex v uniformly at random from G . Let $\alpha_v = \Pr[v \in I]$, and $\alpha_N = \frac{1}{d} \sum_{u \sim v} \Pr[u \in I]$. Let Y be the (random) number of neighbors of v in the independent set. Let $p_k = \Pr[Y = k]$.

$$\begin{aligned} \alpha_v &= \frac{\lambda}{1 + \lambda} p_0 \\ \alpha_N &= \frac{1}{d} \mathbb{E}Y. \end{aligned}$$

This imposes a constraint on the distribution of Y :

$$\frac{\lambda}{1 + \lambda} p_0 = \frac{1}{d} \sum_{k=0}^d k p_k.$$

Now consider the event that k neighbors of v are occupied. We can choose any one of these k vertices and remove it, giving an independent set with $k - 1$ vertices in the neighborhood of v , and with weight $1/\lambda$ times the weight of the initial independent set. An independent set with $k - 1$ vertices in the neighborhood of v can be extended by adding one vertex in the neighborhood of v in at most $d - k + 1$ different ways. This gives the family of constraints

$$p_{k-1} \geq \frac{k}{d-k+1} \cdot \frac{p_k}{\lambda}$$

for $k = 2, \dots, d$.

Now we can optimize subject to these constraints.

$$\begin{aligned} & \text{maximize } \frac{\lambda}{1+\lambda} p_0 \\ & \text{subject to } \sum_{k=0}^d p_k = 1 \\ & \quad p_k \geq 0 \quad \forall k, \\ & \quad \frac{\lambda}{1+\lambda} p_0 = \frac{1}{d} \sum_{k=0}^d k p_k, \\ & \quad p_{k-1} \geq \frac{k}{d-k+1} \frac{p_k}{\lambda} \text{ for } k = 2, \dots, d. \end{aligned}$$

This is a linear program with variables p_0, \dots, p_d and $d + 1$ constraints (in addition to the non-negativity constraints).

We claim that in an optimal solution all of the inequality constraints (not including the non-negativity constraints) must be tight. Suppose that $p_{j-1} = \frac{j}{d-j+1} \frac{p_j}{\lambda}$ for all $j < k$ but $p_{k-1} = \epsilon + \frac{k}{d-k+1} \frac{p_k}{\lambda}$ for some $\epsilon > 0$. Then we show how to generate a new feasible solution p'_0, \dots, p'_d with a larger objective value. We let

$$\begin{aligned} p'_0 &= p_0 + \Delta_0 \\ p'_j &= p_j - \Delta_j \text{ for } j < k \\ p'_k &= p_k + \Delta_k \\ p'_j &= p_j \text{ for } j > k. \end{aligned}$$

To maintain feasibility we must have

$$(12) \quad \Delta_0 + \Delta_k = \sum_{j=1}^{k-1} \Delta_j$$

$$(13) \quad \frac{\lambda}{1+\lambda} \Delta_0 = \frac{k}{d} \Delta_k - \sum_{j=1}^{k-1} \frac{j}{d} \Delta_j.$$

As long as $\Delta_0 > 0$ then the objective function will increase. We will choose the Δ_j 's in such a way that the inequality constraints for $j < k$ remain tight. That is,

$$\Delta_{j-1} = \frac{j}{d-j+1} \frac{\Delta_j}{\lambda},$$

or

$$\Delta_j = \frac{\lambda(d-j+1)}{j} \Delta_{j-1}$$

for $2 \leq j \leq k-1$. In particular,

$$\Delta_j = \frac{\lambda^{j-1}}{d} \binom{d}{j} \Delta_1,$$

and so

$$A := \sum_{j=1}^{k-1} \Delta_j = \frac{\Delta_1}{\lambda d} \sum_{j=1}^{k-1} \binom{d}{j} \lambda^j$$

and

$$B := \sum_{j=1}^{k-1} \frac{j}{d} \Delta_j = \frac{\Delta_1}{\lambda d^2} \sum_{j=1}^{k-1} j \binom{d}{j} \lambda^j.$$

Solving the feasibility constraints (12) and (13) gives

$$\Delta_0 = \frac{1 + \lambda}{k(1 + \lambda) + d\lambda} (Ak - Bd),$$

so by choosing Δ_1 small enough as a function of ϵ so that the constraint $p'_{k-1} \geq \frac{k}{d-k+1} \frac{p'_k}{\lambda}$ is not violated, it is enough to show that $Ak > Bd$, or in other words,

$$k \sum_{j=1}^{k-1} \binom{d}{j} \lambda^j > \sum_{j=1}^{k-1} j \binom{d}{j} \lambda^j$$

which is clear.

Therefor we must satisfy the $d-1$ inequality constraints with equality. There are several ways to complete the proof from here. One simple way is to observe that in $K_{d,d}$ the inequality constraints hold with equality, and the constraint matrix is full rank, and so the distribution induced by $K_{d,d}$ must be the unique optimal solution.

Alternatively, we can compute directly

$$p_k = \frac{\lambda^{k-1}}{d} \binom{d}{k} p_1$$

for $k = 2, \dots, d$, and so

$$p_0 + \frac{p_1}{\lambda d} \sum_{k=1}^d \lambda^k \binom{d}{k} = 1$$

and

$$\frac{\lambda}{1 + \lambda} p_0 = \frac{p_1}{\lambda d^2} \sum_{k=1}^d k \lambda^k \binom{d}{k}.$$

Solving these two equations gives $p_0 = \frac{(1+\lambda)^d}{2(1+\lambda)^d - 1}$, and therefor

$$\alpha_{OPT} \leq \frac{\lambda(1 + \lambda)^{d-1}}{2(1 + \lambda)^d - 1},$$

which proves Theorem 12 since $\alpha_G(\lambda) \leq \alpha_{OPT}$ for any d -regular G .