2 Combinatorics and statistical physics

In this lecture we will look at combinatorics problems from the perspective of statistical physics and see the usefulness of some of statistical physics tools.

2.1 Terminology and perspectives

Some of the central questions in combinatorics can be summarized as follows.

1. **Extremal problems**: what is the largest structure of a given type?
2. **Counting**: how many structures of a given type are there?
3. **Typical structure**: what are the structural properties of a typical (uniformly random) structure of a given type?
4. **Stability**: are all (or most?) near-extremal structures of a given type close in some metric to an extremal structure?

Each of these questions has an analogue in statistical physics.

1. **Ground state**: which configuration σ minimizes $H(\sigma)$?
2. **Partition function**: compute or approximate the partition function Z.
3. **Gibbs measure**: what is the typical structure of a sample from the Gibbs measure μ?
4. **Order/disorder phase transition**: at positive temperature do samples from the Gibbs measure resemble (in some sense) a ground state?

These questions are essentially the same but using different language. On the other hand, the perspectives of combinatorics and statistical physics can be quite different: combinatorics is often concerned with optimizing over all graphs (or hypergraphs), while statistical physics is often focused on specific infinite graphs like $\mathbb{Z}^d$.

This difference in perspectives can often be very useful. Some important results in statistical physics would be very hard to prove without taking the worst-case combinatorics perspective (like the Lee–Yang [17] and Heilmann–Lieb [15] theorems). On the other hand, by focusing on specific graphs, statistical physicists can understand behavior of these models very precisely and provide intuition useful in other settings.

There are also algorithmic variants of these questions.

1. **Optimization**: find a minimizer or maximizer of a certain function over a set.
2. **Counting**: compute or estimate the number of objects in a given set.
3. **Sampling**: output a random object with distribution equal to or close to a given target distribution.

4. **Computational threshold**: does the complexity of a given optimization, approximate counting, or sampling problem change from tractable to intractable as a parameter changes?

In the next section, we will prove a couple of results about independent sets in graphs, taking a statistical physics approach to combinatorics problems.

2.2 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 an 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 2.1 (Kahn [16])**.** *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 follows. For any d-regular, bipartite G,

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

or, more convenient from our perspective,

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

Work of Galvin and Tetali [14] and Zhao [21] extended this result to all values of the independence polynomial and all d-regular graphs.

Theorem 2.2 (Kahn; Galvin-Tetali; Zhao)**.** *For all d-regular graphs G and all $\lambda > 0$,*

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

See Galvin's lecture notes on the entropy method [13] for an exposition of the proof of Theorem 2.1 and extensions. See also the recent work of Sah, Sawhney, Stoner, and Zhao [18, 19] for an extension to irregular graphs among other results.

The question of minimizing the number of (weighted) independent sets in a d-regular graph is somewhat simpler: the answer is the clique K_{d+1}, proved by Cutler and Radcliffe [9]; for a short proof see the exercises.

Among all d-regular graphs, the graph with the smallest scaled independence number is the clique K_{d+1}. If we impose the condition that G has no triangles, then it is not immediately clear which graph has the smallest independence number $\alpha(G)$.

Following Ajtai, and Komlós, and Szemerédi [1], Shearer proved the following.

Theorem 2.3 (Shearer [20]). *For any triangle-free graph G on n vertices of average degree at most d,*

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

As a consequence, Shearer obtained the current best upper bound on the Ramsey number $R(3, k)$.

Corollary 2.4 (Shearer [20]). *The Ramsey number $R(3, k)$ satisfies*

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

The random d-regular graph (conditioned on being triangle-free) satisfies

$$\alpha(G) = (1 + o_d(1)) \frac{2 \log d}{d} n \,,$$

and this is the smallest independence ratio known for a max degree d graph (in the large d limit), and so there is potentially a factor of 2 that could potentially be gained in Shearer's bound. This factor of 2 would immediately give a factor 2 improvement to the upper bound on $R(3, k)$. The best lower bound (proved very recently) is $\frac{1}{3} \frac{k^2}{\log k}$ [6].

In this lecture we will use statistical physics methods to prove a strengthening of Theorem 2.2 and a result closely resembling Theorem 2.3 but for the average size of an independent set rather than the maximum size.

To start we define the *occupancy fraction* of the hard-core model to be the expected fraction of vertices in the random independent set:

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

The first theorem states that $K_{d,d}$ maximizes the occupancy fraction over all d-regular graphs and all λ.

Theorem 2.5 (Davies, Jenssen, Perkins, Roberts [11]). *For all $d \geq 2$, all $\lambda \geq 0$, and all d-regular graphs G,*

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

This implies Theorem 2.2 as follows:

$$\begin{aligned} \frac{1}{n} \log Z_G(\lambda) &= \frac{1}{n} \log Z_G(0) + \frac{1}{n} \int_0^\lambda (\log Z_G(t))' \, dt \\ &= 0 + \int_0^\lambda \frac{\overline{\alpha}_G(t)}{t} \, dt \\ &\leq \int_0^\lambda \frac{\overline{\alpha}_{K_{d,d}}(t)}{t} \, dt \end{aligned}$$

$$= \frac{1}{2d} \log Z_{K_{d,d}}(\lambda)\,.$$

The next theorem states that the expected density of an independent set in a triangle-free graph G of max degree d is at least $(1 + o_d(1))\frac{\log d}{d}$, the same bound that Shearer achieves for the maximum density (though he proves this with the weaker condition of average degree d). This result also yields a lower bound on the number of independent sets in a triangle-free graph.

Theorem 2.6 (Davies, Jenssen, Perkins, Roberts [12])**.** *For all triangle-free graph G of maximum degree d,*

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

Moreover,

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

The respective constants 1 *and* 1/2 *are best possible and attained by the random d-regular graph.*

We start by proving Theorem 2.5 for triangle-free graphs along with Theorem 2.6. Consider the hard-core model on a d-regular, triangle-free G on n vertices.

We imagine the following two-part experiment: pick a random independent set $\mathbf{I}$ from the hard-core model on G and independently pick a uniformly random vertex $\mathbf{v}$ from $V(G)$. We then will record some local information about $\mathbf{I}$ from the perspective of v. In particular, we say v is *uncovered* with respect to an independent set I if $N(v) \cap I = \emptyset$, and we will record $\mathbf{Y}$, the number of uncovered neighbors of $\mathbf{v}$ with respect to the random independent set $\mathbf{I}$. The random variable $\mathbf{Y}$ takes integer values between 0 and d and its distribution depends on both G and λ.

We will also use the following two facts that follow from the spatial Markov property.

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 with our two-part experiment in mind, we write $\overline{\alpha}_G(\lambda)$ in two ways:

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

and

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

Recall $\mathbf{Y}$ is the number of uncovered neighbors of $\mathbf{v}$ with respect to $\mathbf{I}$. Now our two expressions for $\overline{\alpha}_G(\lambda)$ can be interpreted as expectations over $\mathbf{Y}$.

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

Thus the identity

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

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

Theorems 2.5 and 2.6 are optimization problems (maximization and minimization respectively) over the set of all d-regular graphs. Now we can *relax* these optimization problems: instead of maximizing or minimizating $\overline{\alpha}_G(\lambda)$ over all d-regular triangle-free 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 (1). In particular all d-regular triangle-free graphs induce a distribution $\mathbf{Y}$ satisfying these conditions, but there may be additional distributions that do not arise from graphs.

Consider the maximization problem first. By convexity we 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 (1), 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,

$$\overline{\alpha}_G(\lambda) \le \overline{\alpha}_{K_{d,d}}(\lambda)\,.$$

This proves Theorem 2.5 in the special case of triangle-free graphs.

What if we want to *minimize* $\mathbb{E}\mathbf{Y}$ subject to these constraints? In this case, by convexity, we should take $\mathbf{Y}$ to be constant: $\mathbf{Y} = y^*$ where $(1+\lambda)^{-y^*} = \frac{y^*}{d}$, or in other words, $y^*\cdot e^{y^*\log(1+\lambda)} = d$.

Formally, we can use Jensen's inequality:

$$\frac{1}{d}\mathbb{E}\mathbf{Y} = \mathbb{E}(1+\lambda)^{-\mathbf{Y}} \ge (1+\lambda)^{-\mathbb{E}\mathbf{Y}}$$

and so $\mathbb{E}\mathbf{Y} \ge y^*$ as above.

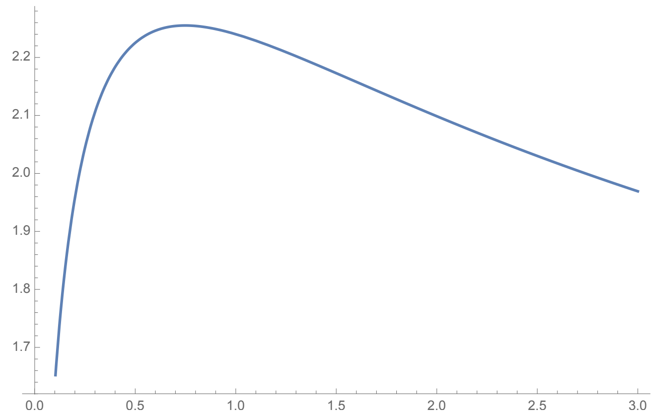

Figure 1: $\frac{\lambda}{1+\lambda} y^*$ as a function of λ with $d = 100$.

The solution is

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

where $W(\cdot)$ is the W-Lambert function (that is, $x = W(x) e^{W(x)}$). This gives

$$\overline{\alpha}_G(\lambda) \geq \frac{1}{d} \frac{\lambda}{1+\lambda} \frac{W(d \log(1+\lambda))}{\log(1+\lambda)} . \tag{2}$$

Now although $\overline{\alpha}_G(\lambda)$ is monotone increasing in λ, somewhat surprisingly the bound (2) is not monotone in λ (see Figure 1 for example).

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

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

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

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

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

and by monotonicity this extends to all larger λ.

To obtain the counting result we integrate the bound (2) for $\lambda = 0$ to 1 to obtain a lower bound on the partition function.

$$\begin{aligned} \frac{1}{n} \log i(G) = \frac{1}{n} \log Z_G(1) &= \int_0^1 \frac{\overline{\alpha}_G(t)}{t} \, dt \\ &\geq \int_0^1 \frac{1}{d} \frac{1}{1+t} \frac{W(d \log(1+t))}{\log(1+t)} \, dt \qquad \text{from (2)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{d}\int_0^{W(d\log 2)} 1 + u\,du \quad \text{using the substitution } u = W(d\log(1+t)) \\
&= \frac{1}{d}\left[W(d\log 2) + \frac{1}{2}W(d\log 2)^2\right] \\
&= \left(\frac{1}{2} + o_d(1)\right)\frac{\log^2 d}{d}\,.
\end{aligned}
$$

Using a similar argument to the proof of the $R(3,k)$ upper bound, we can use Theorem 2.6 to give a lower bound on the number of independent sets in a triangle-free graph without degree restrictions.

Corollary 2.7 ([12])**.** *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 2.6. 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 $2^d = e^{\left(\frac{1}{2}+o_d(1)\right)\frac{\log^2 d}{d}n}$, that is,

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

and so

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

□

This improves the bound of $e^{\left(\frac{\sqrt{\log 2}}{4}+o(1)\right)\sqrt{n}\log n}$ from [7].

In the next section we give the full proof of Theorem 2.5, dispensing with the triangle-free assumption.

2.2.1 Linear programming and occupancy fractions

We begin by reviewing the basics of linear programming, duality, and complementary slackness.

Linear programming review

Suppose we have the linear program in standard form with variables $x_1, \dots x_n$:

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

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:

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

Theorem 2.8 (Weak duality theorem)**.** *If $x_1, \dots, x_n$ and $\Lambda_1, \dots, \Lambda_m$ are feasible solutions to the primal and dual LP's respectively, then*

$$\sum_{i=1}^{n} c_i x_i \leq \sum_{j=1}^{m} b_j \Lambda_j \,.$$

In particular, the objective value of *any* feasible dual solution gives an upper bound on the optimum of the primal LP. The strong duality theorem says that an optimal upper bound can be found if both primal and dual are feasible.

Theorem 2.9 (Strong duality theorem)**.** *If the primal and dual linear programs are both feasible, 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.

The strong duality theorem implies that there are four possibilities for a primal/dual pair of LP's:

1. Both primal and dual are feasible and their optima coincide.
2. Both primal and dual are infeasible.
3. The primal is infeasible and the dual is unbounded.
4. The dual is infeasible and the primal is unbounded.

Very often it is useful to use *complementary slackness* to check for optimality. Say $x_1, \dots, x_n$ and $\Lambda_1, \dots, \Lambda_m$ are feasible solutions to the primal and dual LP's. We say complementary slackness holds for the pair of solutions if:

- For each $j = 1, \dots, m$ either $\Lambda_j = 0$ or the jth constraint of the primal holds with equality under the solution $x_1, \dots, x_n$ (or both are true).
- For each $i = 1, \dots, n$ either $x_i = 0$ or the ith constraint of the dual holds with equality under the solution $\Lambda_1, \dots, \Lambda_m$ (or both are true).

Theorem 2.10. *The feasible solutions $x_1, \dots, x_n$ and $\Lambda_1, \dots, \Lambda_m$ to the primal and dual LP's respectively are an optimal pair of solutions if and only if complementary slackness holds.*

For much more on linear programming, see for example Boyd and Vandenberghe's book [4].

Proof of Theorem 2.5

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 on at most d vertices over the randomness in our two-part experiment. Specifically we mean the neighbors of $\mathbf{v}$ not covered by vertices in $V(G) \setminus N(\mathbf{v})$; so we don't consider covering among the neighbors.

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

$$\overline{\alpha}_G(\lambda) = \frac{\lambda}{1+\lambda} \Pr_{G,\lambda}[\mathbf{v} \text{ uncovered}] = \frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] \tag{3}$$

$$\overline{\alpha}_G(\lambda) = \frac{1}{d} \mathbb{E}_{G,\lambda}[\mathbf{I} \cap N(\mathbf{v})] = \frac{\lambda}{d} \mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right], \tag{4}$$

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

$$\frac{\lambda}{1+\lambda} \mathbb{E}_{G,\lambda}\left[\frac{1}{Z_{\mathbf{H}}(\lambda)}\right] = \frac{\lambda}{d} \mathbb{E}_{G,\lambda}\left[\frac{Z'_{\mathbf{H}}(\lambda)}{Z_{\mathbf{H}}(\lambda)}\right]. \tag{5}$$

Now again we can relax our optimization problem from maximizing $\overline{\alpha}_G$ 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 (5).

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

$$\Pr_{K_{d,d}}(\mathbf{H} = \emptyset) = \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}$$

$$\Pr_{K_{d,d}}(\mathbf{H} = \overline{K_d}) = \frac{(1+\lambda)^d}{2(1+\lambda)^d - 1}\,.$$

To show that this distribution is the maximizer we will use linear programming duality.

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.

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

The first two constraints ensure that the variables $p(H)$ form a probability distribution; the last is constraint (5).

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:

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

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 $\Lambda_p, \Lambda_c \ge 0$ 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:

$$\frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1} + \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}\left[\frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)} - \frac{\lambda}{d}\frac{Z'_H(\lambda)}{Z_H(\lambda)}\right] \geq \frac{\lambda}{1+\lambda}\frac{1}{Z_H(\lambda)}\,. \tag{6}$$

Multiplying through by $Z_H(\lambda) \cdot (2(1+\lambda)^d - 1)$ and simplifying, (6) reduces to

$$\frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d - 1} \geq \frac{\lambda Z'_H(\lambda)}{Z_H(\lambda) - 1}\,, \tag{7}$$

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 (7) has a nice probabilistic interpretation: the RHS is simply $\mathbb{E}_{H,\lambda}\left[|\mathbf{I}|\big||\mathbf{I}| \geq 1\right]$, 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 isolated vertices, $\overline{K_d}$. Proving (7) is left for the exercises, and this completes the proof.

2.3 Further directions and open questions

Zhao has a nice survey on the area of extremal problems for regular graphs [22]. See also the paper of Csikvári [8].

Theorem 2.6 implies the upper bound on $R(3,k)$ 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).

Question 1. *Can we use Theorem 2.6 to improve the current upper bound on $R(3,k)$?*

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

Conjecture 2.11 ([12])**.** *For any triangle-free graph G, we have*

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

Conjecture 2.12 ([12])**.** *For any triangle-free graph G of minimum degree d, we have*

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

as $d \to \infty$.

Conjecture 2.11 would imply a factor 4/3 improvement on the current upper bound for $R(3,k)$, while Conjecture 2.12 would imply a factor 2 improvement.

Matchings and perfect matchings

A classic result that can be interpreted as an extremal problem for bounded degree graphs is Bregman's Theorem [5]. This theorem gives an upper bound on the permanent of a 0/1 matrix with prescribed row sums.

Recall that the permanent of an $n \times n$ matrix A, $\mathrm{per}(A)$, is

$$\mathrm{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \,,$$

where the sum is over permutations on n elements.

Theorem 2.13 (Bregman)**.** *Let A be an $n \times n$ matrix with $\{0,1\}$-valued entries and row sums $d_1, \dots d_n$. Then*

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

Bregman's theorem can be stated as an upper bound on the number of perfect matchings in a balanced bipartite graph with a given degree sequence on one side. Let $\mathrm{pm}(G)$ denote the number of perfect matchings of a graph G.

Corollary 2.14. *Suppose G is a bipartite graph on two parts of $n/2$ vertices each, with left degrees $d_1, \dots d_{n/2}$, then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{n/2} (d_i!)^{1/d_i}.$$

One can also ask about perfect matchings in not-necessarily bipartite graphs.

Theorem 2.15 (Kahn and Lovasz)**.** *Let G be a graph on $2n$ vertices with vertex degrees $d_1, \dots, d_{2n}$. Then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{2n} (d_i!)^{1/(2d_i)}.$$

In the case of a d-regular graph on n vertices, this means that

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

One can then ask which graph maximizes the number of total matchings, not just perfect matchings. Let $\mathcal{M}(G)$ be the set of all matchings of G. Let $m(G) = |\mathcal{M}(G)|$ be the number of matchings and $m_k(G)$ be the number of matchings of size k (so $m_{n/2}(G) = \mathrm{pm}(G)$). The matching polynomial, or the partition function of the monomer-dimer model, is

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

(Sometimes the matching polynomial is defined differently, as $\sum_{k \geq 0} (-1)^k \lambda^{n-2k} m_k(G)$ but these are equivalent up to scaling of the polynomial and the argument). The matching polynomial is the independence polynomial of the *line graph* of G: the graph on the edges of G with adjacency determined by incidence.

Theorem 2.16 (Davies, Jenssen, Perkins, Roberts [11]). *For any d-regular graph G, and any $\lambda > 0$,*

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

This is the analogue of Theorem 2.2 for matchings. In particular, taking $\lambda = 1$ we have $m(G)^{1/n} \le m(K_{d,d})^{1/2d}$ for all d-regular graphs G.

Question 2. *For a given d, λ, what is the minimum and minimizer of*

$$\frac{1}{|V(G)|}\log Z_G^{\text{match}}(\lambda)$$

over all d-regular graphs?

See discussion in [10, 3].

Question 3. *Is there a degree sequence version of Theorem 2.16, that is, a generalization of Bregman's theorem to the matching polynomial?*

Question 4. *Is there an entropy-based proof of Theorem 2.16? (Or simply an easier proof than that in [11]).*

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] M. Borbenyi and P. Csíkvári. Matchings in regular graphs: minimizing the partition functionn. *Transactions on Combinatorics*, 10(2):1–23, 2021.

[4] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[5] L. Bregman. Some properties of nonnegative matrices and their permanents. *Soviet Math. Dokl*, 14(4):945–949, 1973.

[6] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. A new lower bound for the ramsey numbers $R(3, k)$. *arXiv preprint arXiv:2505.13371*, 2025.

[7] J. Cooper, K. Dutta, and D. Mubayi. Counting independent sets in hypergraphs. *Combinatorics, Probability and Computing*, 23(4):539–550, 2014.

[8] P. Csikvári. Extremal regular graphs: the case of the infinite regular tree. *arXiv preprint arXiv:1612.01295*, 2016.

[9] 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, Jan. 2014.

[10] E. Davies, M. Jenssen, and W. Perkins. A proof of the upper matching conjecture for large graphs. *Journal of Combinatorial Theory, Series B*, 151:393–416, 2021.

[11] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Independent sets, matchings, and occupancy fractions. *Journal of the London Mathematical Society*, 96(1):47–66, 2017.

[12] 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*, 146:111–124, 2018.

[13] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.

[14] D. Galvin and P. Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.

[15] O. J. Heilmann and E. H. Lieb. Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25(3):190–232, 1972.

[16] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(03):219–237, 2001.

[17] T.-D. Lee and C.-N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Physical Review*, 87(3):410, 1952.

[18] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. The number of independent sets in an irregular graph. *Journal of Combinatorial Theory, Series B*, 138:172–195, 2019.

[19] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. A reverse Sidorenko inequality. *Inventiones mathematicae*, pages 1–47, 2020.

[20] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.

[21] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(02):315–320, 2010.

[22] Y. Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *The American Mathematical Monthly*, 124(9):827–843, 2017.