10 Entropy methods

10.1 Information entropy

We're going to shift away from concentration results now. This next concept was essentially invented by Shannon, and we'll focus on its combinatorial applications.

Definition 10.1

Let X be a discrete random variable taking values in some set S. Then the **entropy** of X is

$$H(X) = \sum_{s \in S} -p_s \log_2 p_s,$$

where $p_s = \Pr(X = s)$.

Intuitively, entropy is supposed to measure the amount of randomness or information in the random variable X.

Because we're doing combinatorics, we'll work with base-2 logarithms - this is really more of a convention than anything else, and all logs in this section mean base 2.

Example 10.2

The entropy of a Bernoulli variable $\text{Ber}(p)$ is just $-p \log_2 p - (1-p)\log_2(1-p)$, which has a maximum of 1 at $p = \frac{1}{2}$.

Basically, this tracks how "surprised" we are when we hear a sample from the distribution. This idea essentially comes from trying to encode messages efficiently: for example, if a coin only comes up heads 1% of the time, encoding it as a binary string directly is not the most efficient way.

Lemma 10.3

$H(X) \le \log_2 |\text{range}(X)|$.

Proof. This is convexity of the function $x \to x \log_2 x$. □

Equality holds when we have the uniform distribution: then $H(X)$ tells us the number of binary bits needed to specify which choice of X we pick out.

Denote by $H(X, Y)$ the entropy of the joint random variable $Z = (X, Y)$, where X and Y are not necessarily independent. This means we have

$$H(X, Y) = \sum_{(x,y)} -\Pr(X = x, Y = y) \log_2 \Pr(X = x, Y = y).$$

Lemma 10.4 (Subadditivity)

Given any two random variables X, Y, $H(X, Y) \le H(X) + H(Y)$.

Proof. Expanding $H(X) + H(Y) - H(X, Y)$ out, this gives

$$H(X) + H(Y) - H(X, Y) = \sum_{x,y} \left(-p(x, y) \log_2 p(x) - p(x, y) \log_2 p(y) + p(x, y) \log_2 p(x, y)\right) \quad = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}.$$

Let $f(t) = t \log t$, which is a convex function. Then by Jensen's, we can bound this as

$$= \sum_{x,y} p(x)p(y) f\left(\frac{p(x,y)}{p(x)p(y)}\right) \geq f(1) = 0.$$

□

Basically, there's at least as much information in X and Y individually as when we put them together.

$$H(X) + H(Y) - H(X,Y) = I(X,Y)$$

is called the **mutual information**, and it's always nonnegative.

In particular, if X and Y are independent, then $H(X,Y) = H(X) + H(Y)$. In this case, the amount of information in our variable X is just the sum of the individual parts.

Corollary 10.5

For any random variables $X_1, \cdots, X_n$,

$$H(X_1, \cdots, X_n) \leq H(X_1) + \cdots + H(X_n).$$

There's also a notion of "conditional entropy:" let E be an event with positive probability, and then we have

$$H(X|E) = \sum_x - \Pr(X = x|E) \log_2 \Pr(X = x|E).$$

What's really important to us, though, is when we condition on a second random variable: if X and Y are jointly distributed, we define

$$H(X|Y) = \mathbb{E}_y \left[H(X|Y = y) \right].$$

Essentially, this is how much new information we get given a certain piece of information about Y.

Lemma 10.6 (Chain rule)

For any random variables X, Y, $H(X|Y) = H(X,Y) - H(Y)$.

Proof.

$$\begin{aligned} H(X|Y) &= \mathbb{E}_y \left[H(X|Y = y) \right] \\ &= \sum_y \Pr(Y = y) H(X|Y = y) = \sum_y -p(y) \sum_x p(x|y) \log_2 p(x|y) \\ &= \sum_{x,y} -p(x,y) \log_2 p(x,y) + \sum_{x,y} p(x,y) \log_2 (y) \\ &= \sum_{x,y} -p(x,y) \log_2 p(x,y) + \sum_y p(y) \log_2 (y), \end{aligned}$$

where the first equality follows from Bayes' rule and the last because $\sum_x p(x,y) = p(y)$.

□

In other words, the conditional entropy is just the total entropy minus what we "already knew about Y." In particular, if $X = Y$, or if $X = f(Y)$ (so we know X given Y), the conditional entropy is 0. On the other hand, if X and Y are independent, the conditional entropy is just $H(X)$.

Lemma 10.7 (Dropping conditioning)
For any random variables X, Y, Z, $H(X|Y) \leq H(X)$ and $H(X|Y, Z) \leq H(X|Z)$.

Proof. These follow from the chain rule (Lemma 10.6) and subadditivity (Lemma 10.4). For example,

$$H(X|Y) = H(X, Y) - H(Y) \leq H(X).$$

□

10.2 Various direct applications

Let's start to see how this can be useful! Entropy's use primarily comes up in tail bounds. Here's a philosophy: we want to show an upper bound on some quantity, so we start by taking the log of both sides. The left side is the log of some quantity, so take a uniform probability distribution on the things we want to count: we now have an entropy.

Theorem 10.8
Let $\mathcal{F}$ be a collection of subsets of $[n]$, and let p_i be the fraction of subsets in $\mathcal{F}$ that contain the element i. Then

$$\log_2 |\mathcal{F}| \leq \sum_{i=1}^{n} H(p_i),$$

where $H(p)$ is the binary entropy of the Bernoulli variable

$$H(p) = -p \log_2 p - (1 - p) \log_2(1 - p).$$

Proof. Let $X = (X_1, \cdots, X_n)$ be the characteristic vector for a uniform random element $F \in \mathcal{F}$: this means that X_i is 1 if $i \in F$ and 0 otherwise. The entries aren't necessarily independent here, so we can play with this with entropy: $\log_2 |\mathcal{F}|$ is just $H(X)$, because we have a uniform distribution (this is the equality case of Lemma 10.3).

By subadditivity, this is at most $H(X_1) + \cdots + H(X_n)$. Each X_i is a Bernoulli random variable with probability p_i, which is what we want. □

Theorem 10.9
Let $k \leq \frac{n}{2}$. Then

$$\sum_{0 \leq i \leq k} \binom{n}{i} \leq 2^{H\left(\frac{k}{n}\right)n}.$$

Proof. Let $X = (X_1, \cdots, X_n) \in \{0, 1\}^n$ be the uniform random vector conditioned on $X_1 + \cdots + X_n \leq k$. By Lemma 10.3, the logarithm of the left hand side is $H(X)$, and by subadditivity, this is at most $H(X_1) + \cdots + H(X_n)$. Conditioning on the sum of the X_i being exactly m, each X_i is a Bernoulli variable with probability $\frac{m}{n}$. Since the sum is always at most k, we can say that X_i is Bernoulli with probability at most $\frac{k}{n}$. Since this is less than $\frac{1}{2}$ by assumption and the entropy of a Bernoulli increases until $p = 1/2$, we have that $H(X_i) \leq H\left(\frac{k}{n}\right)$.

Now there are n copies of this term, and rearranging gives the result. □

We get a similar result if we don't pick everything with probability $\frac{1}{2}$ but instead with probability p: then we get a relative entropy called the Kullback-Leibler divergence.

Theorem 10.10 (This was problem 32 from our problem set)
Let $S_1, \cdots, S_k$ be subsets of $[n]$, and suppose that for every pair of distinct subsets $A, B \subseteq [n]$, there exists an i such that

$$|S_i \cap A| \neq |S_i \cap B|.$$

Then $k \geq (2 - o(1))\frac{n}{\log_2 n}$.

This is called a coin weighing problem, because we can imagine that we have two types of coins, where one is a little heavier than the other. We can then weigh k times, and we want to be able to tell how many counterfeit coins we have. Well, if there always exists an i that distinguishes them, then we know exactly which coins we want. It turns out we need at least $\approx \frac{2n}{\log_2 n}$ weighings to do the job.

The main idea here is that there's some information that we're gaining on each comparison S_i: can we get enough to deduce the set of coins?

Proof. Let X be a uniform random subset of $[n]$. Since there are 2^n different possibilities that are uniformly weighted, the entropy of X is just n. Observe that X contains the same information as the sizes of all $|X \cap S_i|$ for $1 \leq i \leq k$: in particular, this is an injective map, since no two subsets have the same set of intersections. By subadditivity,

$$H(X) = H(|X \cap S_1|, \cdots, |X \cap S_k|) \leq H(|X \cap S_1|) + \cdots + H(|X \cap S_k|).$$

Because X is a uniform subset of 1 through n, $|X \cap S_i|$ is binomial with distribution $\text{Bin}\left(|S_i|, \frac{1}{2}\right)$. The entropy of such a binomial distribution is bounded by $\log_2 |S_i|$, and $|S_i| \leq n$, so this gives

$$n = H(X) \leq k \log_2 n,$$

which is enough to give everything except for the factor of 2.

However, note that the binomial distribution is not uniform: it's highly concentrated, and thus we should have much less entropy than a uniform distribution! Heuristically, we know that the binomial distribution is concentrated in a $\sqrt{|S_i|}$-interval, so the entropy should be essentially related to $\log_2(|S_i|)$. This turns out to be true if we work out the calculations, and that gives us

$$H\left(\text{Bin}\left(|S_i|, \frac{1}{2}\right)\right) \leq \left(\frac{1}{2} + o(1)\right) \log_2 m$$

and now rearranging gives the result that we want. □

As a sidenote, the actual entropy of the Binomial distribution is $\frac{1}{2}\log_2 m + O(1)$.

10.3 Bregman's theorem

Definition 10.11

The **permanent** of an $n \times n$ matrix is

$$\text{per } A = \sum_{\sigma \in S_n} \prod a_{i,\sigma_i}.$$

In contrast, the **determinant** is similar but includes a sign:

$$\det A = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod a_{i,\sigma_i}.$$

These are very different quantities - in particular, the determinant is believed to be much easier to calculate.

Let's only consider matrices $A \subset \{0,1\}^{n \times n}$: any such matrix can be encoded by a bipartite graph with n row nodes and n column nodes, where row i and column j are connected if and only if there's a 1 in the corresponding entry.

Lemma 10.12

The permanent of a matrix $A \subset \{0,1\}^{n \times n}$ is equal to the number of perfect matchings in the corresponding bipartite graph.

Proof. The permanent expands over all permutations, and we count a permutation if and only if every edge we try to use exists (giving us a product of 1). □

So here's a natural question to ask: if we have some degree distribution (for example, d-regular), what is the maximum number of perfect matchings that are possible? One possible extremal graph is a union of complete bipartite graphs: in the d-regular case, the number of perfect matchings is just $d!$ to some power. Is this the best we can do in general?

Theorem 10.13 (Bregman)

Given a matrix $A \in \{0,1\}^{n \times n}$ whose i^{th} row sums to d_i for all i,

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

Note that a disjoint union of complete bipartite graphs $K_{s,s}$ gives the equality case.

Proof by Radhakrishnan. Let σ be a uniform permutation of $[n]$, conditioned on all A_{i,σ_i} being 1 in our matrix. In other words, we are picking a uniform random perfect matching! By Lemma 10.12, $H(\sigma) = \log_2(\text{per} A)$.

Attempt 1 (Subadditivity): σ has n different coordinates, one for the entry σ_i picked in each row, so each coordinate is a random variable. As we've done in the previous examples, we can try to apply subadditivity here, bounding the entropy for the i^{th} coordinate by $H(\sigma_i)$. If there are d_i 1s in that row, we can say that $H(\sigma_i) \le \log_2 d_i$ (we may not have equality because we don't have a uniform distribution on the σ_is). Unfortunately, this is not enough! Because the σ_i are not chosen independently (e.g. picking something in the first row affects the others), applying subadditivity directly costs us a lot.

Attempt 2 (Randomization + chain rule): Instead, let's reveal the rows in a uniform random order and then apply the chain rule. If τ is a uniform permutation in S_n, we now have

$$H(\sigma) = H(\sigma_{\tau_1}) + H(\sigma_{\tau_2}|\sigma_{\tau_1}) + \cdots + H(\sigma_{\tau_n}|\sigma_{\tau_1}, \cdots, \sigma_{\tau_{n-1}}).$$

Take expectations on both sides. We know that the left hand side is independent of the ordering - at the end of the process, we still see all the rows, so the information we get is the same. Thus,

$$H(\sigma) = \mathbb{E}[H(\sigma_{\tau_1})] + \mathbb{E}[H(\sigma_{\tau_2}|\sigma_{\tau_1})] + \cdots + \mathbb{E}[H(\sigma_{\tau_n}|\sigma_{\tau_1}, \cdots, \sigma_{\tau_{n-1}})].$$

What's the contribution of the i^{th} row of our original matrix to this sum? If the row appears in the k^{th} term of the sum, then it contributes $\mathbb{E}[H(\sigma_i|\cdots)]$, where $\cdots$ represents a uniform subset of $k-1$ other rows. Then,

$$\mathbb{E}[H(\sigma_i|\cdots)] \leq \mathbb{E}[\log_2(\text{number of available entries in row } i|\cdots)].$$

Since we only care about the ordering of the d_i rows whose entries conflict with the d_i 1s in row i, and each ordering is equally likely,

$$= \frac{1}{d_i}(\log_2 1 + \log_2 2 + \cdots + \log_2 d_i) = \frac{1}{d_i}\log_2(d_i!).$$

Plugging this back into the sum,

$$\mathbb{E}[H(\sigma)] \leq \sum_{i=1}^{n} \frac{1}{d_i}\log_2(d_i!),$$

and exponentiating both sides yields the result. □

10.4 A useful entropy lemma

Lemma 10.14 (Shearer's lemma (special))
For any random variables X, Y, Z,

$$2H(X,Y,Z) \leq H(X,Y) + H(X,Z) + H(Y,Z).$$

Proof. By the chain rule,

$$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y).$$

Now, add up the following:

$$H(X,Y) = H(X) + H(Y|X)$$
$$H(X,Z) = H(X) + H(Z|X)$$
$$H(Y,Z) = H(Y) + H(Z|Y).$$

Dropping conditioning on $H(X,Y,Z)$ yields the result. □

What are some applications of this?

Corollary 10.15
Given a finite set $S \subset \mathbb{R}^3$, consider the orthogonal projections $\pi_{xy}(S)$ onto the xy-plane (and similarly for the xz and yz-planes). We have

$$|S|^2 \leq \pi_{xy}(S)\pi_{xz}(S)\pi_{yz}(S).$$

Equality holds for a Cartesian box.

Proof. Let (X, Y, Z) be a uniform point in S. Then $\log_2 |S|$ is the entropy $H(X, Y, Z)$, and by Shearer, this entropy is at most

$$2 \log_2 |S| \leq H(X, Y) + H(X, Z) + H(Y, Z).$$

The shadow distribution doesn't need to be uniform, but we can upper bound its entropy with that of the uniform distribution:

$$\leq \log_2 \pi_{xy}(S) + \log_2 \pi_{xz}(S) + \log_2 \pi_{yz}(S).$$

Taking 2 to the power of both sides gives the result. □

Remark. *We can actually get the same result for a volume $S \subset \mathbb{R}^3$: the volume of S squared is at most the areas of the projections onto the planes. This can be proved by approximating S as a union of grid boxes!*

Let's now look at Shearer's inequality in its general form.

Theorem 10.16 (Shearer's lemma, general)
Let $A_1, \cdots, A_s \subseteq [n]$, where each $i \in [n]$ appears in at least k different A_js. Let $X_1, \cdots, X_n$ be random variables and define the joint random variables

$$X_{A_j} = (X_i)_{i \in A_j}.$$

Then

$$kH(X_1, \cdots, X_n) \leq H(X_{A_1}) + \cdots + H(X_{A_S}).$$

In the special case (Lemma 10.14), A_1, A_2, and A_3 are just the two-element subsets of $(1, 2, 3)$. The proof of the general case is the same! Let's establish a corollary analogous to Corollary 10.15:

Corollary 10.17
Let $A_1, \cdots, A_s \subseteq \Omega$, where each $i \in \Omega$ appears in at least k different A_j's. Then for every family $\mathcal{F}$ of subsets of Ω,

$$|\mathcal{F}|^k \leq \prod_{j=1}^{S} |\mathcal{F}|_{A_j}|,$$

where the notation $\mathcal{F}|_{A_j}$ means $\mathcal{F}$ restricted to the elements of A_j: $\{S \cap A : S \in \mathcal{F}\}$.

Proof. Let $(X_1, \cdots, X_n) \in \{0, 1\}^n$ be the indicator vector of a uniform random $F \in \mathcal{F}$. Then

$$k \log_2 |\mathcal{F}| = kH(X_1, \cdots, X_n) \leq \sum_{j=1}^{S} H(X_{A_j}).$$

Again, we can upper bound by the uniform entropy:

$$k \log_2 |\mathcal{F}| \leq \sum_{j=1}^{s} \log_2 |\mathcal{F}|_{A_j}|,$$

and exponentiate both sides to get the desired result. □

Let's use this for a combinatorial application: in particular, what was the problem that inspired this inequality?

Problem 10.18 (Easy)

What is the largest intersecting family of subsets of m elements, where "intersecting family" means every pair has a nonempty intersection?

The answer is 2^{m-1}: we can just pick every subset that contains the element 1. This is maximal, because any set A and its complement $[m] \setminus A$ can't both appear. If we look back to the beginning of class, the original problem restricted us to only k-element subsets; without this restriction, the problem is easy.

Problem 10.19

What is the largest set of graphs on n labeled vertices so that every pair has a common triangle?

We can get $\frac{1}{8}$ of the total: fix a triangle, and pick all graphs containing that fixed triangle. We also know that it's less than $\frac{1}{2}$ of the total, because we can't pick both a graph and its complement.

Theorem 10.20 (Chung-Frankl-Graham-Shearer)

Every triangle-intersecting family of graphs on n labeled vertices has at most $2^{\binom{n}{2}-2}$ elements.

Proof. Let $\mathcal{G}$ be a triangle-intersecting family on n vertices. Notice that if we restrict ourselves to half our graph and look at the shadow on the two cliques, we must still have an edge-intersecting family, because what's left is a complete bipartite graph.

More concretely, let $m = \binom{n}{2}$. Pick a subset $S \subseteq [n]$ with $|S| = \left\lfloor \frac{n}{2} \right\rfloor$, and let A_S be the union of cliques on S and $[n] \setminus S$. $K_n \setminus A$ is triangle free, so $\mathcal{G}|_A$ must be intersecting. This means that $|\mathcal{G}|_{A_S}| \le 2^{|A_S|-1}$ by the logic above: now if we look at all possible Ss, each edge of K_n appears in k different A_Ss, where $k = \frac{r}{m}\binom{n}{\lfloor n/2 \rfloor}$, $r = |A_S|$.

Now by Shearer's lemma,

$$|\mathcal{G}|^k \le \prod |\mathcal{G}|_{A_S}| = \left(2^{r-1}\right)^{\binom{n}{\lfloor n/2 \rfloor}}.$$

This simplifies to $|\mathcal{G}| \le 2^{m - \frac{m}{r}}$, where $\frac{m}{r}$ is the inverse of the edge density, and since $\frac{m}{r} \ge 2$ for all n, this yields the desired result. □

What's the truth, though?

Theorem 10.21 (Ellis-Filmus-Friedgut, 2012)

Every triangle-intersecting family of graphs on n labeled vertices has at most $\frac{1}{8} \cdot 2^{\binom{n}{2}}$ elements.

The proof of this more refined result uses Fourier analysis!

10.5 Entropy in graph theory

Problem 10.22

Among all d-regular graphs G, how can we maximize the quantity

$$i(G)^{1/v(G)},$$

where $i(G)$ is the number of independent sets and $v(G)$ is the number of vertices of G?

It turns out that this quantity is maximized for a disjoint union of copies of $K_{d,d}$. Let's start by doing this in a special case:

Theorem 10.23 (Kahn)
For a bipartite n-vertex d-regular graph G,

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

Equality holds if and only if G is the disjoint union of copies of $K_{d,d}$.

Proof. Pick a bipartition of $V(G) = A \cup B$, and let $X = (X_v)_{v \in V(G)}$ be the indicator vector for an independent set of G chosen uniformly at random. (In other words, pick a random independent set, and put a 1 for each vertex in the set and 0 everywhere else.) Then the entropy of this variable is just $H(X) = \log_2(i(G))$.

How can we upper bound this? X is not necessarily uniform or independent on the vertices, but we can still write

$$\log_2(i(G)) = H(X) = H(X_A) + H(X_B|X_A).$$

Observe that because the graph is d-regular and bipartite, each vertex in A lies in the neighbor sets of d vertices in B. Therefore, we can simplify the first term using Theorem 10.16 and also bound the second term by subadditivity:

$$H(X) \le \frac{1}{d}\sum_{b \in B} H(X_{N(b)}) + \sum_{b \in B} H(X_b|X_A)$$

Dropping conditioning on the second term (forgetting about the non-neighbors),

$$H(X) \le \frac{1}{d}\sum_{b \in B} H(X_{N(b)}) + \sum_{b \in B} H(X_b|X_{N(b)}).$$

Fix a $b \in B$. We upper bound the expression

$$H(X_{N(b)}) + dH(X_b|X_{N(b)}).$$

We want to relate this to the entropy of $i(K_{d,d})$ somehow: we will do so by replacing X_b with d identical independent variables $X_b^{(1)}, \dots, X_b^{(d)}$ that have the same distribution given $X_{N(b)}$ as the original X_b. Then,

$$\begin{aligned} H(X_{N(b)}) + dH(X_b|X_{N(b)}) &= H(X_{N(b)}) + H\left(X_b^{(1)}|X_{N(b)}\right) + \cdots + H\left(X_b^{(d)}|X_{N(b)}\right) \\ &= H(X_{N(b)}) + H\left(X_b^{(1)}, \cdots, X_b^{(d)}|X_{N(b)}\right) \\ &= H(X_{N(b)}, X_b^{(1)}, \cdots, X_b^{(d)}), \end{aligned}$$

where the last equality follows from the chain rule. The key observation is that the joint random variable $Y = (X_{N(b)}, X_b^{(1)}, \cdots, X_b^{(d)})$ is the indicator variable of some random independent set of $K_{d,d}$: $X_{N(b)}$ corresponds to the d vertices on the left side and the d variables $X_b^{(i)}$ correspond to d different vertices on the right side! The values that Y takes correspond to independent sets, because the original X_b (and thus none of the copies) is never 1 if there's a 1 in any coordinate of $X_{N(b)}$.

This distribution of Y may not be uniform, but we can still upper bound its entropy by the entropy of the uniform distribution over independent sets of $K_{d,d}$, which is (by Lemma 10.3 as always) $\log_2(i(K_{d,d}))$.

Our graph G is d-regular, so the two pieces of the bipartition have size $\frac{n}{2}$. Because the above bound holds for every $b \in B$,

$$\log_2 i(G) \le \frac{n}{2d}\log_2(i(K_{d,d}))$$

as desired. □

In this proof, we used almost nothing about independent sets, and that motivates us to generalize this result.

Definition 10.24

A **graph homomorphism** $G \to H$ is a map of the vertex set $V(G) \to V(H)$ such that every edge $uv \in G$ is mapped to an edge $\phi(u)\phi(v)$ in H.

Example 10.25

Here are two examples of graph homomorphisms:

- **Independent sets:** Let H be the graph on two vertices $\{0, 1\}$ with an edge between 0 and 1 and a self-loop on 0. Then, a map $\phi\colon (V(G)) \to H$ induces a homomorphism if and only if $\phi^{-1}(1)$ forms an independent set.
- **q-colorings:** Let $H = K_q$. The proper q-colorings of a graph G correspond to homomorphisms from G to H: color each vertex in G mapping to i with the color i.

Theorem 10.26 (Galvin-Tetai)

Let G be an n-vertex, d-regular bipartite graph, and let H be any (possibly looped) graph. Let $\mathrm{Hom}(G, H)$ to be the set of homomorphisms from G to H: then

$$|\mathrm{Hom}(G, H)| \le |\mathrm{Hom}(K_{d,d}, H)|^{n/2d}.$$

The proof of this result is identical to the proof of Theorem 10.23.

Corollary 10.27

Let G be an n-vertex d-regular bipartite graph, and let $q \in \mathbb{N}$. Let $c_q(G)$ denote the number of proper q-colorings of G: then

$$c_q(G) \le c_q(K_{d,d})^{v(G)}.$$

Proof. Let X be the vector of colors of a uniformly random coloring of G, and the rest follows as above. □

Is it possible to prove an analog of Theorem 10.23 for general (not necessarily bipartite) graphs? The answer is yes!

Theorem 10.28

For a n-vertex d-regular graph G,

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

Equality holds if (and only if) G is the disjoint union of copies of $K_{d,d}$.

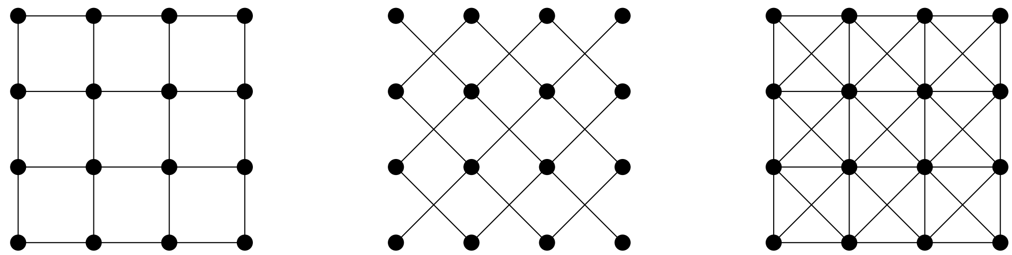
Proof. We will reduce to the bipartite case.

Lemma 10.29 (Zhao)

For all G,

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

Remark. *There's lots of ways to denote a graph product. Given two paths G and H on 4 vertices, there's three main ways to construct a graph product of those paths:*



These, naturally, should be denoted $G \square H$, $G \times H$, and $G \boxtimes H$, respectively.

Proof of lemma. We will construct an injection from $\mathcal{I}(G \sqcup G)$, the collection of independent sets in two disjoint copies of G, to $\mathcal{I}(G \times K_2)$. Think of $G \sqcup G$ as two copies of G, one above the other, and $G \times K_2$ as the same thing but with parallel edges replaced with crosses.

Let's say we have some independent set $S \in \mathcal{I}(G \sqcup G)$: if we take those same vertices in $G \times K_2$, we might not have an independent set, because there are some bad edges: treating $G \sqcup G$ as two layers 0 and 1,

$$E_{\text{bad}} = \{uv \in E(G) : (u,0),(v,1) \in S\}.$$

All edges in E_{bad} correspond to an edge in $G \times K_2$ with one endpoint in $\{u \in V(G) : (u,0) \in S\}$, which is the set of vertices of S (our not-quite independent set) in the top layer. Fix some ordering of the subsets of $V(G)$ (for example, lexicographical, and take Q to be the first subset (in our ordering) of $V(G)$ such that each bad edge in E_{bad} has exactly one endpoint in Q. In other words, we're finding some canonical subset that "shows" our bipartition.

Now swap each pair of $V(G \times K_2)$ in Q (in other words, replace $(v,0)$ with $(v,1)$ and vice versa): we can check that this gives us an independent set in $G \times K_2$. In addition, this mapping is injective: find the edges of E that correspond to E_{bad}, and then we can find Q and reverse all of the swaps that we did. □

The graph $G \times K_2$ is d-regular and bipartite with $2n$ vertices, so we can apply Theorem 10.23. This gives an inequality

$$i(G) \le i(G \times K_2)^{\frac{1}{2}} \le i(K_{d,d})^{n/2d},$$

and we're done. □

This means that for independent sets, we can drop the bipartite hypothesis: can we do the same in general for graph homomorphisms? The answer is no!

Example 10.30

Take H to be two disjoint loops. Any graph homomorphism into H sends each connected component to one of the two vertices of H, so the graph with the most graph homomorphisms into H is not a union of copies of $K_{d,d}$ but rather a union of cliques K_{d+1}, since we're just trying to maximize the number of connected components.

The above bipartite swapping trick does not work for some variants of the problem, such as the number of q-colorings instead of the number of independent sets. Recently, the problem for the number of proper colorings was settled using a different method by Sah, Sawhney, Stoner, and Zhao.

Also, we can reduce "bipartite" to "triangle-free" in the graph homomorphism theorem. On the flip side, for any G with triangles, there exists a graph H for which the theorem is not true! However, we don't have good conjectures on classifications of the graph H.

10.6 More on graph homomorphisms: Sidorenko's conjecture

Definition 10.31

Let $t(H, G)$ denote the number of homomorphisms from G to H, divided by the total number of vertex maps $|V(G)|^{|V(H)|}$.

In other words, this is the probability that a uniform random vertex map induces a graph homomorphism.

Conjecture 10.32 (Sidorenko)

If H is a bipartite graph, then for all G, the homomorphism density

$$t(H, G) \geq t(K_2, G)^{e(H)},$$

where $t(K_2, G)$ is the edge density.

Rephrased, this can be phrased another way: among all graphs G with a fixed edge density, which G has the minimum number of copies of H? Sidorenko's conjecture says (informally) that this is a "random" G. This is still an open problem, but let's look at a specific case.

Theorem 10.33

Let G be a graph with n vertices and m edges, and let P_4 be a three-edge path. Then

$$\hom(P_4, G) \geq n^3 \left(\frac{2m}{n^2}\right)^3 = \frac{8m^3}{n^2}.$$

Proof. We'll use the entropy method, but the proof will look slightly different from the techniques that have been used so far. We're trying to lower-bound our quantity this time, so we don't necessarily want to start with a uniform distribution.

Basically, our goal is to construct a probability distribution on the set of homomorphisms $\text{Hom}(P_4, G)$ with entropy at least $\log_2\left(\frac{(2m)^3}{n^2}\right)$. Then by the uniform inequality, we can find that the entropy of the uniform distribution, which is $\log_2$ of the number of homomorphisms, is at least that quantity. Note that a homomorphism is just a 4-vertex path.

Construct X, Y, Z, W to be a 4-vertex walk on G in the following way: let XY be a unfirom edge of the graph, Z be a uniform neighbor of Y (allowing X), and W be a uniform neighbor of Z. The entropy of this distribution is, by the chain rule,

$$H(X, Y, Z, W) = H(X) + H(Y|X) + H(Z|X, Y) + H(W|X, Y, Z).$$

Note that if XY is a uniform edge, YZ and ZW are also uniformly distributed. This is because the vertex probability distribution of X is proportional to $d(v)$: specifically,

$$\Pr(X = v) = \frac{d(v)}{2m}.$$

This is true for Y as well, and now the distribution of Z as a uniform neighbor of Y is the same as the distribution of X as a uniform neighbor of Y: $Z|Y \sim X|Y$. So YZ is uniform, and so is ZW by the same argument. That means

$$H(X, Y, Z, W) = H(X) + H(Y|X) + H(Z|Y) + H(W|Z) = H(X) + 3H(Y|X),$$

and this is (by definition)

$$\sum_v \frac{-d(v)}{2m} \log_2 \frac{d(v)}{2m} + 3 \sum_v \frac{d(v)}{2m} \log_2 d(v),$$

where $\log_2 d(v)$ is $H(Y|X=v)$. Expanding and applying convexity, this is

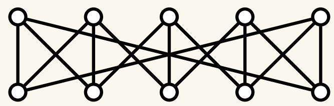
$$= \log_2(2m) + 2\sum_v \frac{d(v)}{2m} \log_2 d(v) \geq \log_2(2m) + \frac{2n}{2m} \cdot \frac{2m}{n} \log_2 \frac{2m}{n},$$

and rearranging gives an entropy of $\log_2 \frac{(2m)^3}{2n}$. □

It turns out that this proof works for every tree. For what kinds of graphs is Sidorenko's conjecture harder to resolve?

Fact 10.34

The smallest open case of Sidorenko's conjecture is the following Mobius graph: it's $K_{5,5}$ minus a Hamiltonian cycle.



It turns out this is the incidence graph of the smallest simplicial complex of the Mobius strip. One side is the set of vertices, and the other is the set of faces.

Notably, the Mobius graph doesn't fit the conditions of the following theorem, which resolves Sidorenko's conjecture for certain graphs:

Theorem 10.35 (Conlon-Fox-Sudakov)

Sidorenko's conjecture holds for a graph H if there exists a bipartition $H = A \sqcup B$ such that there exists a vertex $a \in A$ with $N(a) = B$.

There are also ways to interpret Sidorenko's conjecture beyond graph theory! It turns out Sidorenko's conjecture where H is a three-edge path (Theorem 10.33) is equivalent to the following inequality:

Proposition 10.36

Given a function $f : [0,1]^2 \to [0,\infty]$,

$$\int_{[0,1]^4} f(x,y)f(y,z)f(z,w)dxdydzdw \geq \left(\int_{[0,1]^2} f(x,y)dxdy\right)^3.$$

As a grad student, Professor Zhao posted on Math Overflow a few years ago asking for a Cauchy-Schwarz proof of this. A week ago, Sidorenko actually answered it!

Sidorenko. Think of $g(x) = \int f(x,y)dy$ as representing the "degree of x." Then the left hand side becomes

$$\int f(x,y)f(z,y)g(z)dxdydz$$

but we can also rewrite the graph as a path $u \to x \to y \to z$, so the left hand side is also

$$\int g(x)f(x,y)f(z,y)dxdydz.$$

Applying Cauchy-Schwarz,

$$\text{LHS} \geq \int g(x)^{1/2} f(x,y) f(z,y) g(z)^{1/2}$$

and since this integral is symmetric with respect to x and z, we can write this as

$$= \int \left(\int g(x)^{1/2} f(x,y) dx \right)^2 dy \geq \left(\int g(x)^{1/2} f(x,y) dx dy \right)^2$$

by Cauchy-Schwarz, and now we can integrate out

$$\left(\int g(x)^{3/2} dx \right)^2 \geq \left(\int g(x) dx \right)^3 = \left(\int f(x,y) dx dy \right)^3,$$

and we're done. $\square$