10 Entropy

> My greatest concern was what to call it. I thought of calling it “information,” but the word was overly used, so I decided to call it “uncertainty.” When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, “You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.”
>
> Claude Shannon, 1971

In this chapter, we look at some neat and powerful applications of entropy to combinatorics. For a standard introduction to information theory, see the textbook by Cover and Thomas.

10.1 Basic properties

We define the (binary) entropy of a discrete random variable as follows.

Definition 10.1.1

Given a discrete random variable X taking values in S, with $p_s := \mathbb{P}(X = s)$, its ***entropy*** (or ***binary entropy*** to emphasis the base-2 logarithm) is defined to be

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

(by convention if $p_s = 0$ then the corresponding summand is set to zero).

Remark **10.1.2** (Base of the logarithm)**.** It is also fine to use another base for the logarithm, e.g., the natural log, as long as we are consistent throughout. There is some combinatorial preference for base-2 due to its interpretation as counts bits. For certain results, such as Pinsker’s inequality (which we will unfortunately not cover here), the choice of the base does matter.

***Remark* 10.1.3 (Information theoretic interpretation).** Intuitively, $H(X)$ measures the amount of "surprise" in the randomness of X. It can also be interpreted as the amount of information learned by seeing the random variable X. A more rigorous interpretation of this intuition is given by the **Shannon source coding theorem**, which, informally, says that the minimum number of bits needed to encode n iid copies of X is $nH(X) + o(n)$.

Here are some basic properties. Throughout we only consider discrete random variables.

The proofs are all routine calculations. It will useful to understand the information theoretic interpretations of these properties.

Lemma 10.1.4 (Uniform bound)

$$H(X) \le \log_2 |\operatorname{support}(X)|,$$

with equality if and only if X is uniformly distributed.

Proof. Let function $f(x) = -x \log_2 x$ is concave for $x \in [0,1]$. Let $S = \operatorname{support}(X)$. Then

$$H(X) = \sum_{s \in S} f(p_s) \le |S|\, f\left(\frac{1}{|S|} \sum_{s \in S} p_s\right) = |S|\, f\left(\frac{1}{|S|}\right) = \log_2 |S|\,. \qquad \square$$

We write $H(X,Y)$ for the entropy of the joint random variables (X,Y). In other words, letting $Z = (X,Y)$,

$$\boldsymbol{H(X,Y)} := H(Z) = \sum_{(x,y)} -\mathbb{P}(X = x, Y = y) \log_2 \mathbb{P}(X = x, Y = y).$$

We can similarly write $H(X_1, \dots, X_n)$ for joint entropy.

Lemma 10.1.5 (Independence)

If X and Y are independent random variables, then

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

Proof.

$$\begin{aligned} H(X,Y) &= \sum_{(x,y)} -\mathbb{P}(X=x, Y=y)\log_2 \mathbb{P}(X=x,Y=y) \\ &= \sum_{(x,y)} -p_x p_y \log_2(p_x p_y) \\ &= \sum_{(x,y)} -p_x p_y (\log_2 p_x + \log_2 p_y) \\ &= \sum_x -p_x \log_2 p_x + \sum_y -p_y \log_2 p_y = H(X) + H(Y). \end{aligned}$$ □

Definition 10.1.6 (Conditional entropy)
Given jointly distributed random variables X and Y, define

$$\begin{aligned} \boldsymbol{H(X|Y)} &:= \mathbb{E}_y[H(X|Y=y)] \\ &= \sum_y \mathbb{P}(Y=y) H(X|Y=y) \\ &= \sum_y \mathbb{P}(Y=y) \sum_x -\mathbb{P}(X=x|Y=y)\log_2 \mathbb{P}(X=x|Y=y) \end{aligned}$$

(each line unpacks the previous line. In the summations, x and y range over the supports of X and Y respectively).

Intuitively, the conditional entropy $H(X|Y)$ measures the amount of additional information in X not contained in Y. This is intuition is also captured by the next lemma.

Some important special cases:

- If $X = Y$, or $X = f(Y)$, then $H(X|Y) = 0$.
- If X and Y are independent, then $H(X|Y) = H(X)$
- If X and Y are conditionally independent on Z, then $H(X,Y|Z) = H(X|Z) + H(Y|Z)$ and $H(X|Y,Z) = H(X|Z)$.

Lemma 10.1.7 (Chain rule)

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

Proof. Writing $p(x,y) = \mathbb{P}(X=x, Y=y)$, etc., we have by Bayes's rule

$$p(x|y)p(y) = p(x,y),$$

and so

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

Lemma 10.1.8 (Subadditivity)
$H(X,Y) \le H(X) + H(Y)$, and more generally,

$$H(X_1, \dots, X_n) \le H(X_1) + \cdots + H(X_n).$$

Proof. Let $f(t) = \log_2(1/t)$, which is convex. Then

$$\begin{aligned} H(X) + H(Y) - H(X,Y) &= \sum_{x,y} \left(-p(x,y)\log_2 p(x) - p(x,y)\log_2 p(y) + p(x,y)\log_2 p(x,y)\right) \\ &= \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y) f\left(\frac{p(x)p(y)}{p(x,y)}\right) \\ &\ge f\left(\sum_{x,y} p(x,y) \frac{p(x)p(y)}{p(x,y)}\right) = f(1) = 0 \end{aligned}$$

More generally, by iterating the above inequality for two random variables, we have

$$\begin{aligned} H(X_1, \dots, X_n) &\le H(X_1, \dots, X_{n-1}) + H(X_n) \\ &\le H(X_1, \dots, X_{n-2}) + H(X_{n-1}) + H(X_n) \\ &\le \cdots \le H(X_1) + \cdots + H(X_n). \end{aligned}$$ □

Remark **10.1.9** (Mutual information)**.** The nonnegative quantity

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

is called ***mutual information***. Intuitively, it measures the amount of common information between X and Y.

Lemma 10.1.10 (Dropping conditioning)
$H(X|Y) \le H(X)$ and more generally,

$$H(X|Y,Z) \le H(X|Z).$$

Proof. By chain rule and subadditivity, we have

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

The inequality conditioning on Z follows since the above implies that

$$H(X|Y, Z = z) \ge H(X|Z = z)$$

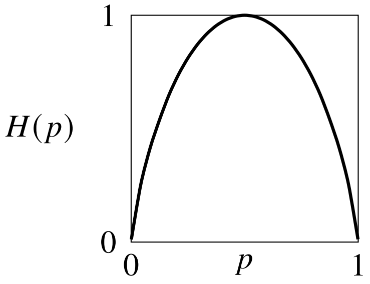
holds for every z, and taking expectation of z yields $H(X|Y,Z) \le H(X|Z)$. □

Remark* 10.1.11.** A related theorem is the ***data processing inequality: $H(X|f(Y)) \ge H(X|Y)$ for any function f. More generally, f can be random. In other words, if $X \to Y \to Z$ is a Markov chain, then $H(X|Z) \ge H(X|Y)$ (exercise: prove this).

Here are some simple applications of entropy to **tail bounds**.

Let us denote the entropy of a Bernoulli random variable by

$$H(p) := H(\text{Bernoulli}(p)) = -p \log_2 p - (1-p)\log_2(1-p).$$



(This notation $H(\cdot)$ is standard but unfortunately ambiguous: $H(X)$ versus $H(p)$. It is usually clear from context which is meant.)

Theorem 10.1.12
If $0 < k \le n/2$, then

$$\sum_{0\le i\le k} \binom{n}{i} \le 2^{H(k/n)n} = \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k}.$$

This bound can be established using our proof technique for Chernoff bound by applying Markov's inequality to the moment generating function:

$$\sum_{0\le i\le k}\binom{n}{i} \le \frac{(1+x)^n}{x^k} \qquad \text{for all } x\in[0,1].$$

The infimum of the RHS over $x \in [0,1]$ is precisely $2^{H(k/n)n}$.

Now let us give a purely information theoretic proof to get some practice with entropy.

Proof. Let $(X_1,\dots,X_n)\in\{0,1\}^n$ be chosen uniformly *conditioned* on $X_1+\cdots+X_n \le k$. Then

$$\log_2 \sum_{0\le i\le k}\binom{n}{i} = H(X_1,\dots,X_n) \le H(X_1)+\cdots+H(X_n).$$

Each X_i is a Bernoulli with probability $\mathbb{P}(X_i=1)$. Note that conditioned on $X_1+\cdots+X_n=m$, one has $\mathbb{P}(X_i=1)=m/n$. Varying over $m\le k\le n/2$, we find $\mathbb{P}(X_i=1)\le k/n$, so $H(X_i)\le H(k/n)$. Hence

$$\log_2 \sum_{0\le i\le k}\binom{n}{i} \le H(k/n)n. \qquad \square$$

***Remark* 10.1.13.** One can extend the above proof to bound the tail of Binomial(n,p) for any p. The result can be expressed in terms of the *relative entropy* (also known as the *Kullback–Leibler divergence* between two Bernoulli random variables). More concretely, for $X\sim$ Binomial(n,p), one has

$$\frac{\log \mathbb{P}(X\le nq)}{n} \le -q\log\frac{q}{p} - (1-q)\log\frac{1-q}{1-p} \qquad \text{for all } 0\le q\le p,$$

and

$$\frac{\log \mathbb{P}(X\ge nq)}{n} \le -q\log\frac{q}{p} - (1-q)\log\frac{1-q}{1-p} \qquad \text{for all } p\le q\le 1.$$

10.2 Permanent, perfect matchings, and Steiner triple systems

Permanent

We define the ***permanent*** of an $n\times n$ matrix A by

$$\operatorname{per} A := \sum_{\sigma\in S_n}\prod_{i=1}^{n} a_{i,\sigma(i)}.$$

The formula for the permanent is simply that of the determinant without the sign factor:

$$\det A := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma_i}.$$

We'll consider $\{0,1\}$-valued matrices. If A is the bipartite adjacency matrix of a bipartite graph, then

$$\operatorname{per} A = \text{ the number of perfect matchings.}$$

The following theorem gives an upper bound on the number of perfect matchings of a bipartite graph with a given degree distribution. It was conjectured by Minc (1963) and proved by Brégman (1973).

Theorem 10.2.1 (Brégman–Minc inequality)
Let $A = (a_{ij}) \in \{0,1\}^{n\times n}$, whose i-th row has sum d_i. Then

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

Note that equality is attained when A consists diagonal blocks of 1's (corresponding to perfect matchings in a bipartite graph of the form $K_{d_1,d_1} \sqcup \cdots \sqcup K_{d_t,d_t}$).

Let σ be a uniform random permutation of $[n]$ conditioned on $a_{i\sigma_i} = 1$ for all $i \in [n]$. Then

$$\log_2 \operatorname{per} A = H(\sigma) = H(\sigma_1,\dots,\sigma_n) = H(\sigma_1) + H(\sigma_2|\sigma_1) + \cdots + H(\sigma_n|\sigma_1,\dots,\sigma_{n-1}).$$

We have

$$H(\sigma_i|\sigma_1,\dots,\sigma_{i-1}) \le H(\sigma_i) \le \log_2 |\operatorname{support} \sigma_i| = \log_2 d_i,$$

but this step would be too lossy. In fact, what we just did amounts to a naive worst case counting argument.

The key new idea is to **reveal the chosen entries in a uniform random order.**

Proof. (Radhakrishnan 1997) Let σ be as earlier. Consider a permutation of τ representing an ordering of the rows of the matrix. Say that *i appears before j* if $\tau_i < \tau_j$.

Let $N_i = N_i(\sigma, \tau)$ be the number of ones on row i that does not lie in the same column as some entry (j, σ_j) that comes before i. (Intuitively, N_i is the number of "greedily available" choices for σ_i before it is revealed.)

For any τ, the chain rule gives

$$H(\sigma) = \sum_{i=1}^{n} H\left(\sigma_i \,\middle|\, \sigma_j : j \text{ comes before } i\right),$$

and the uniform bound gives

$$H\left(\sigma_i \,\middle|\, \sigma_j : j \text{ comes before } i\right) \le \mathbb{E}_\sigma \log_2 N_i.$$

Let τ vary uniformly over all permutations. Then,

$$H(\sigma) \le \sum_{i=1}^{n} \mathbb{E}_{\sigma,\tau} \log_2 N_i.$$

For any fixed σ, as τ varies uniformly over all permutations of $[n]$, N_i varies uniformly over $[d_i]$. (Why?) Thus

$$\mathbb{E}_\tau \log_2 N_i = \frac{\log_2 1 + \cdots + \log_2 d_i}{d_i} = \frac{\log_2(d_i!)}{d_i}.$$

Taking expectation over σ and summing over i yields

$$\log_2 \operatorname{per} A = H(\sigma) \le \sum_{i=1}^{n} \mathbb{E}_{\sigma,\tau} \log_2 N_i \le \sum_{i=1}^{n} \frac{\log_2(d_i!)}{d_i}. \qquad \square$$

Corollary 10.2.2 (Kahn and Lovász)

Let G be a graph. Let d_v denote the degree of v. Then the number $\operatorname{pm}(G)$ of perfect matchings of G satisfies

$$\operatorname{pm}(G) \le \prod_{v\in V(G)} (d_v!)^{1/(2d_v)} = \prod_{v\in V(G)} \operatorname{pm}(K_{d_v,d_v})^{1/(2d_v)}.$$

Proof. (Alon and Friedland 2008) Brégman's theorem implies the statement for bipartite graphs G (by considering a bipartition on $G \sqcup G$). For the extension of non-bipartite G, one can proceed via a combinatorial argument that $\operatorname{pm}(G \sqcup G) \le \operatorname{pm}(G \times K_2)$, which is left as an exercise. □

The maximum number of Hamilton paths in a tournament

Question 10.2.3

What is the maximum possible number of directed Hamilton paths in an n-vertex tournament?

Earlier we saw that a uniformly random tournament has $n!/2^{n-1}$ Hamilton paths in expectation, and hence there is some tournament with at least this many Hamilton paths. This result, due to Szele, is the earliest application of the probabilistic method.

Using Brégman's theorem, Alon proved a nearly matching upper bound.

Theorem 10.2.4 (Alon 1990)

Every n-vertex tournament has at most $O(n^{3/2} \cdot n!/2^n)$ Hamilton paths.

***Remark* 10.2.5.** The upper bound has been improved to $O(n^{3/2-\gamma} n!/2^n)$ for some small constant $\gamma > 0$ (Friedgut and Kahn 2005), while the lower bound $n!/2^{n-1}$ has been improved by a constant factor (Adler, Alon, and Ross 2001, Wormald 2004). It remains open to close this $n^{O(1)}$ factor gap.

We first prove an upper bound on the number of Hamilton cycles.

Theorem 10.2.6 (Alon 1990)

Every n-vertex tournament has at most $O(\sqrt{n} \cdot n!/2^n)$ Hamilton cycles.

Proof. Let A be an $n \times n$ matrix whose (i, j) entry is 1 if $i \to j$ is an edge of the tournament and 0 otherwise. Let d_i be the sum of the i-th row. Then per A counts the number of 1-factors (spanning disjoint unions of directed cycles) of the tournament. So by Brégman's theorem, we have

$$\text{number of Hamilton cycles} \le \operatorname{per} A \le \prod_{i=1}^{n} (d_i!)^{1/d_1}.$$

One can check (omitted) that the function $g(x) = (x!)^{1/x}$ is log-concave, i.e, $g(n)g(n+2) \ge g(n+1)^2$ for all $n \ge 0$. Thus, by a smoothing argument, among sequences $(d_1, \dots, d_n)$ with sum $\binom{n}{2}$, the RHS above is maximized when all the d_i's are within 1 of each other, which, by Stirling's formula, gives $O(\sqrt{n} \cdot n!/2^n)$. □

Theorem 10.2.4 then follows by applying the above bound with the following lemma.

Lemma 10.2.7

Given an n-vertex tournament with P Hamilton paths, one can add a new vertex to obtain a $(n+1)$-vertex tournament with at least $P/4$ Hamilton cycles.

Proof. Add a new vertex and orient its incident edges uniformly at random. For every Hamilton path in the n-vertex tournament, there is probability $1/4$ that it can be closed up into a Hamilton cycle through the new vertex. The claim then follows by linearity of expectation. □

Steiner triple systems

Definition 10.2.8 (Steiner triple system)

A ***Steiner triple system (STS)*** of order n is a 3-uniform hypergraph on n vertices where every pair of vertices is contained in exactly one triple.

Equivalently: an STS is a decomposition of a complete graph K_n into edge-disjoint triangles.

Example: the Fano plane is an STS of order 7.

It is a classic result that an STS of order n exists if and only if $n \equiv 1$ or $3 \bmod 6$. It is not hard to see that this is necessary, since if an STS of order n exsits, then $\binom{n}{2}$ should be divisible by 3, and $n-1$ should be divisible by 2. Keevash (2014+) obtained a significant breakthrough proving the existence of more general designs.

Question 10.2.9

How many STS are there on n labeled vertices?

We shall prove the following result.

Theorem 10.2.10 (Upper bound on the number of STS — Linial and Luria 2013)

The number of Steiner triple systems on n labeled vertices is at most

$$\left(\frac{n}{e^2+o(1)}\right)^{n^2}.$$

***Remark* 10.2.11.** Keevash (2018) proved a matching lower bound when $n \equiv 1, 3 \pmod 6$.

Proof. As in the earlier proof, the idea is to reveal the triples in a random order.

Let X denote a uniformly chosen STS on n vertices. We wish to upper bound $H(X)$.

We encode X as a tuple $(X_{ij})_{i<j} \in [n]^{\binom{n}{2}}$ where X_{ij} is the label of the unique vertex that forms a triple with i and j in the STS. Here whenever we write ij we mean the unordered pair $\{i, j\}$, i.e., an edge of K_n.

Let $y = (y_{ij})_{i<j} \in [0,1]^{\binom{n}{2}}$, and we order the edges of K_n in decreasing y_{ij}:

$$kl \prec ij \qquad \text{if} \qquad y_{kl} > y_{ij}.$$

By the chain rule,

$$H(X) = \sum_{ij} H\left(X_{ij} \mid X_{kl} : kl \prec ij\right).$$

Let

$N_{ij} = N_{ij}(X, y) =$ the number of possibilities for X_{ij} after revealing X_{kl} for all $kl \prec ij$.

By the uniform bound, we have

$$H(X) \le \sum_{ij} \mathbb{E}_X \log_2 N_{ij}.$$

Now let $y = (y_{ij})_{i<j} \in [0,1]^{\binom{n}{2}}$ be chosen uniformly at random. We have

$$H(X) \le \sum_{ij} \mathbb{E}_X \mathbb{E}_y \log_2 N_{ij}.$$

Write $y_{-ij} \in [0,1]^{\binom{n}{2}-1}$ to mean y with the ij-coordinate removed. Let us bound $\mathbb{E}_{y_{-ij}} \log_2 N_{ij}$ as a function of y_{ij}.

We define *ij shows up first in its triple* to be the event that $ij \prec ik, jk$ where $k = X_{ij}$. We have, for any fixed X,

$$\mathbb{P}_{y_{-ij}}(ij \text{ shows up first in its triple}) = \mathbb{P}_{y_{-ij}}(ij \prec ik, jk) = \mathbb{P}_{y_{-ij}}(y_{ij} > y_{ik}, y_{jk}) = y_{ij}^2.$$

If ij does not show up first in its triple, then X_{ij} has exactly one possibility (namely k) by the time it gets revealed, and so $N_{ij} = 1$ and $\log_2 N_{ij} = 0$. Thus

$$\begin{aligned}\mathbb{E}_{y_{-ij}} \log_2 N_{ij} &= y_{ij}^2 \mathbb{E}_{y_{-ij}}\left[\log_2 N_{ij} \mid ij \text{ shows up first in its triple}\right] \\ &\le y_{ij}^2 \log_2 \mathbb{E}_{y_{-ij}}\left[N_{ij} \mid ij \text{ shows up first in its triple}\right].\end{aligned}$$

Now we use linearity of expectations (over y_{-ij} with fixed X). For each $s \in [n] \setminus \{i, j, k\}$, if s is available as a possibility for X_{ij} by the time X_{ij} is revealed, then none of the six edges of K_n consisting of the two triangle isX_{ij} and jsX_{js} may occur before

X_{ij}; the latter event occurs with probability y_{ij}^6. So

$$\mathbb{E}_{y_{-ij}}\left[N_{ij} \mid ij \text{ shows up first in its triple}\right] \le 1 + (n-3)y_{ij}^6.$$

Thus

$$\mathbb{E}_y \log_2 N_{ij} \le \int_0^1 y_{ij}^2 \log_2(1+(n-3)y_{ij}^3)\, dy_{ij} = \frac{1}{3}\int_0^1 \log_2(1+(n-3)t^2)\, dt.$$

This integral actually has a closed-form antiderivative (e.g., check Mathematica/Wolfram Alpha), but it suffices for us to obtain the asymptotics. We have

$$\int_0^1 \log_2\left(\frac{1}{n-3}+t^2\right) dt \to \int_0^1 \log_2(t^2)\, dt = -2\log_2 e$$

as $n \to \infty$ by the monotone convergence theorem. Thus

$$\mathbb{E}_y \log_2 N_{ij} \le \frac{\log_2(n/e^2)+o(1)}{3}.$$

It follows therefore that the log-number of STS on n vertices is

$$H(X) \le \sum_{ij} \mathbb{E}_X \mathbb{E}_y \log_2 N_{ij} \le \binom{n}{2}\left(\frac{\log_2(n/e^2)+o(1)}{3}\right) = \frac{n^2}{6}\log_2\left(\frac{n}{e^2+o(1)}\right). \quad \square$$

***Remark* 10.2.12 (Guessing the formula).** Here is perhaps how we might have guessed the formula for the number of STSs. Suppose we select $\frac{1}{3}\binom{n}{2}$ triangles in K_n independently at random. What is the probability that every edge is contained in exactly one triangle? Each edge is contained one triangle on expectation, and so by the Poisson approximation, the probability that a single fixed edge is contained in exactly one triangle is $1/e + o(1)$. Now let us pretend as if all the edges behave independently (!) — the probability that every edge is contained in exactly one triangle is $(1/e+o(1))^{\binom{n}{2}}$. This would then lead us to guessing that the number of STSs being

$$\frac{1}{\left(\frac{1}{3}\binom{n}{2}\right)!}\binom{n}{3}^{\frac{1}{3}\binom{n}{2}}\left(\frac{1}{e}+o(1)\right)^{\binom{n}{2}} = \left(\left(\frac{n^2}{6e}\right)^{-n^2/6}\left(\frac{n^3}{6}\right)^{n^2/6}\left(\frac{1}{e}\right)^{n^2/2}\right)^{1+o(1)} = \left(\frac{n}{e^2+o(1)}\right)^{n^2/3}.$$

Here is another heuristic for getting the formula, and this time this method can actually be turned into a proof of matching lower bound on the number of STSs, though with a lot of work (Keevash 2018). Suppose we remove triangles from K_n one at a time. After k triangles have been removed, the number of edges remaining is $\binom{n}{2} - 3k$. Let us pretend that the remaining edges were randomly distributed. Then the number of triangles should be about

$$\binom{n}{3}\left(1-\frac{3k}{\binom{n}{2}}\right)^3 \sim \frac{36}{n^3}\left(\frac{1}{3}\binom{n}{2}-k\right)^3$$

If we multiply the above quantity over $0 \le k < \frac{1}{3}\binom{n}{2}$, and then divide by $\left(\frac{1}{3}\binom{n}{2}\right)!$ to account for the ordering of the triangles, we get

$$\frac{\left(\frac{36}{n^3}\right)^{n^2/6}\left(\frac{1}{3}\binom{n}{2}\right)!^3}{\left(\frac{1}{3}\binom{n}{2}\right)!} \approx \left(\frac{n}{e^2+o(1)}\right)^{n^2/3}.$$

10.3 Sidorenko's inequality

Given graphs F and G, a ***graph homomorphism*** from F to G is a map $\phi\colon V(F) \to V(G)$ of vertices that sends edges to edges, i.e., $\phi(u)\phi(v) \in E(G)$ for all $uv \in E(F)$.

Let

$$\hom(F,G) = \text{ the number of graph homomorphisms from } F \text{ to } G.$$

Define the ***homomorphism density*** (the ***H-density in G***) by

$$\begin{aligned} t(F,H) &= \frac{\hom(F,G)}{v(G)^{v(F)}} \\ &= \mathbb{P}(\text{a uniform random map } V(F) \to V(G) \text{ is a graph homomorphism } F \to G) \end{aligned}$$

In this section, we are interested in the regime of fixed F and large G, in which case almost all maps $V(F) \to V(G)$ are injective, so that there is not much difference between homomorphisms and subgraphs. More precisely,

$$\hom(F,G) = \operatorname{aut}(F)(\#\text{copies of } F \text{ in } G \text{ as a subgraph}) + O_F(v(G)^{v(F)-1}).$$

where $\operatorname{aut}(F)$ is the number of automorphisms of F.

Inequalities between graph homomorphism densities is a central topic in extremal graph theory. For example, see Chapter 5 of my book *Graph Theory and Additive Combinatorics*. Much of the rest of this chapter is adapted from §5.5 of the book.

Question 10.3.1

Given a fixed graph F and constant $p \in [0,1]$, what is the minimum possible F-density in a graph with edge density at least p?

The F-density in the random graph $G(n, p)$ is $p^{e(F)} + o(1)$. Here p is fixed and $n \to \infty$.

Can one do better?

If F is non-bipartite, then the complete bipartite graph $K_{n/2,n/2}$ has F-density zero. (The problem of minimizing F-density is still interesting and not easy; it has been solved for cliques.)

Sidorenko's conjecture (1993) (also proposed by Erdős and Simonovits (1983)) says for any fixed bipartite F, the random graph asymptotically minimizes F-density. This is an important and well-known conjecture in extremal graph theory.

Conjecture 10.3.2 (Sidorenko)
For every bipartite graph F, and any graph G,

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

The conjecture is known to hold for a large family of graphs F.

The entropy approach to Sidorenko's conjecture was first introduced by Li and Szegedy (2011) and later further developed in subsequent works. Here we illustrate the entropy approach to Sidorenko's conjecture with several examples.

We will construct a probability distribution μ on $\operatorname{Hom}(F, G)$, the set of all graph homomorphisms $F \to G$. Unlike earlier applications of entropy, here we are trying to prove a lower bound on $\hom(F, G)$ instead of an upper bound. So instead of taking μ to be a uniform distribution (which automatically has entropy $\log_2 \hom(F, G)$), we actually take μ to be carefully constructed distribution, and apply the upper bound

$$H(\mu) \leq \log_2 |\operatorname{support} \mu| = \log_2 \hom(F, G).$$

We are trying to show that

$$\frac{\hom(F, G)}{v(G)^{v(F)}} \geq \left(\frac{2e(G)}{v(G)^2}\right)^{e(F)}.$$

So we would like to find a probability distribution μ on $\operatorname{Hom}(F, G)$ satisfying

$$H(\mu) \geq e(F) \log_2(2e(G)) - (2e(F) - v(F)) \log_2 v(G). \tag{10.1}$$

Theorem 10.3.3 (Blakey and Roy 1965)
Sidorenko's conjecture holds if F is a three-edge path.

Proof. We choose randomly a walk $XYZW$ in G as follows:

- XY is a uniform random edge of G (by this we mean first choosing an edge of G uniformly at random, and then let X be a uniformly chosen endpoint of this edge, and then Y the other endpoint);
- Z is a uniform random neighbor of Y;
- W is a uniform random neighbor of Z.

Key observation: YZ is distributed as a uniform random edge of G, and likewise with ZW

Indeed, conditioned on the choice of Y, the vertices X and Z are both independent and uniform neighbors of Y, so XY and YZ are uniformly distributed.

Also, the conditional independence observation implies that

$$H(Z|X,Y) = H(Z|Y) \qquad \text{and} \qquad H(W|X,Y,Z) = H(W|Z)$$

and futhermore both quantities are equal to $H(Y|X)$ since XY, YZ, ZW are each distributed as a uniform random edge.

Thus

$$\begin{aligned} H(X,Y,Z,W) &= H(X) + H(Y|X) + H(Z|X,Y) + H(W|X,Y,Z) && \text{[chain rule]} \\ &= H(X) + H(Y|X) + H(Z|Y) + H(W|Z) && \text{[cond indep]} \\ &= H(X) + 3H(Y|X) \\ &= 3H(X,Y) - 2H(X) && \text{[chain rule]} \\ &\geq 3\log_2(2e(G)) - 2\log_2 v(G) \end{aligned}$$

In the final step we used $H(X,Y) = \log_2(2e(G))$ since XY is uniformly distributed among edges, and $H(X) \leq \log_2 |\text{support}(X)| = \log_2 v(G)$. This proves (10.1) and hence the theorem for a path of 4 vertices. (As long as the final expression has the "right form" and none of the steps are lossy, the proof should work out.) □

***Remark* 10.3.4.** See this MathOverflow discussion for the history as well as alternate proofs.

The above proof easily generalizes to all trees. We omit the details.

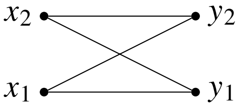
Theorem 10.3.5

Sidorenko's conjecture holds if F is a tree.

Theorem 10.3.6

Sidorenko's conjecture holds for all complete bipartite graphs.

Proof. Following the same framework as earlier, let us demonstrate the result for $F = K_{2,2}$. The same proof extends to all $K_{s,t}$.



We will pick a random tuple $(X_1, X_2, Y_1, Y_2) \in V(G)^4$ with $X_iY_j \in E(G)$ for all i, j as follows.

- X_1Y_1 is a uniform random edge;
- Y_2 is a uniform random neighbor of X_1;
- X_2 is a conditionally independent copy of X_1 given (Y_1, Y_2).

The last point deserves more attention. Note that we are *not* simply uniformly randomly choosing a common neighbor of Y_1 and Y_2 as one might naively attempt. Instead, one can think of the first two steps as generating a distribution for (X_1, Y_1, Y_2)—according to this distribution, we first generate (Y_1, Y_2) according to its marginal, and then produce two conditionally independent copies of X_1 (the second copy is X_2).

As in the previous proof (applied to a 2-edge path), we see that

$$H(X_1, Y_1, Y_2) = 2H(X_1, Y_1) - H(X_1) \geq 2\log_2(2e(G)) - \log_2 v(G).$$

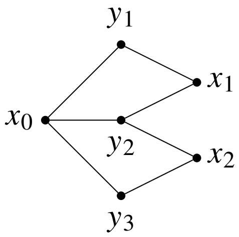
So we have

$$\begin{aligned}
&H(X_1, X_2, Y_1, Y_2) \\
&= H(Y_1, Y_2) + H(X_1, X_2|Y_1, Y_2) && \text{[chain rule]} \\
&= H(Y_1, Y_2) + 2H(X_1|Y_1, Y_2) && \text{[conditional independence]} \\
&= 2H(X_1, Y_1, Y_2) - H(Y_1, Y_2) && \text{[chain rule]} \\
&\geq 2(2\log_2(2e(G)) - \log_2 v(G)) - 2\log_2 v(G). && \text{[prev. ineq. and uniform bound]} \\
&= 4\log(2e(G)) - 4\log_2 v(G).
\end{aligned}$$

So we have verified (10.1) for $K_{2,2}$. □

Theorem 10.3.7 (Conlon, Fox, Sudakov 2010)
Sidorenko's conjecture holds for a bipartite graph that has a vertex adjacent to all vertices in the other part.

Proof. Let us illustrate the proof for the following graph. The proof extends to the general case.



Let us choose a random tuple $(X_0, X_1, X_2, Y_1, Y_2, Y_3) \in V(G)^6$ as follows:

- X_0Y_1 is a uniform random edge;
- Y_2 and Y_3 are independent uniform random neighbors of X_0;
- X_1 is a conditionally independent copy of X_0 given (Y_1, Y_2);
- X_2 is a conditionally independent copy of X_0 given (Y_2, Y_3).

(as well as other symmetric versions.) Some important properties of this distribution:

- X_0, X_1, X_2 are conditionally independent given (Y_1, Y_2, Y_3);
- X_1 and (X_0, Y_3, X_2) are conditionally independent given (Y_1, Y_2);
- The distribution of (X_0, Y_1, Y_2) is identical to the distribution of (X_1, Y_1, Y_2).

We have

$$\begin{aligned}
&H(X_0, X_1, X_2, Y_1, Y_2, Y_3) \\
&= H(X_0, X_1, X_2 | Y_1, Y_2, Y_3) + H(Y_1, Y_2, Y_3) && \text{[chain rule]} \\
&= H(X_0 | Y_1, Y_2, Y_3) + H(X_1 | Y_1, Y_2, Y_3) + H(X_2 | Y_1, Y_2, Y_3) + H(Y_1, Y_2, Y_3) && \text{[cond indep]} \\
&= H(X_0 | Y_1, Y_2, Y_3) + H(X_1 | Y_1, Y_2) + H(X_2 | Y_2, Y_3) + H(Y_1, Y_2, Y_3) && \text{[cond indep]} \\
&= H(X_0, Y_1, Y_2, Y_3) + H(X_1, Y_1, Y_2) + H(X_2, Y_2, Y_3) - H(Y_1, Y_2) - H(Y_2, Y_3). && \text{[chain rule]}
\end{aligned}$$

The proof of Theorem 10.3.3 actually lower bounds the first three terms:

$$\begin{aligned}
H(X_0, Y_1, Y_2, Y_3) &\geq 3 \log_2(2e(G)) - 2 \log_2 v(G) \\
H(X_1, Y_1, Y_2) &\geq 2 \log_2(2e(G)) - \log_2 v(G) \\
H(X_2, Y_2, Y_3) &\geq 2 \log_2(2e(G)) - \log_2 v(G).
\end{aligned}$$

We can apply the uniform support bound on the remaining terms.

$$H(Y_1, Y_2) = H(Y_2, Y_3) \le 2\log_2 v(G).$$

Putting everything together, we have

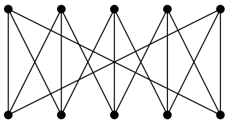
$$H(X_0, X_1, X_2, Y_1, Y_2, Y_3) \ge 7\log_2(2e(G)) - 8\log_2 v(G),$$

thereby verifying (10.1). □

To check that you understand the above proof, where did we use the assumption that F has a vertex complete to the other part?

Many other graphs can be proved by extending this method.

Remark* 10.3.8** (Möbius graph). An important open case (and the smallest in some sense) of Sidorenko conjecture is when F is the following graph, known as the ***Möbius graph. It is $K_{5,5}$ with a 10-cycle removed. The name comes from it being the face-vertex incidence graph of the simplicial complex structure of the Möbius strip, built by gluing a strip of five triangles.

Möbius graph $= K_{5,5} \setminus C_{10} =$ 

10.4 Shearer's lemma

Shearer's entropy lemma extends the subadditivity property of entropy. Before stating it in full generality, let us first see the simplest instance of Shearer's lemma.

Theorem 10.4.1 (Shearer's lemma, special case)

$$2H(X,Y,Z) \le H(X,Y) + H(X,Z) + H(Y,Z)$$

Proof. Using the chain rule and conditioning dropping, we have

$$\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) \\
H(X,Z) &= H(X) + H(Z|X) &&\ge H(X) + H(Z|X,Z) \\
H(Y,Z) &= H(Y) + H(Z|Y) &&\ge H(Y|X) + H(Z|X,Y)
\end{aligned}$$

Applying conditioning dropping, we see that their sum is at at least

$$2H(X) + 2H(Y|X) + 2H(Z|X,Y) = 2H(X,Y,Z).$$ □

Question 10.4.2
What is the maximum volume of a body in $\mathbb{R}^3$ that has area at most 1 when projected to each of the three coordinate planes?

The cube $[0,1]^3$ satisfies the above property and has area 1. It turns out that this is the maximum.

To prove this claim, first let us use Shearer's inequality to prove a discrete version.

Theorem 10.4.3
Let $S \subseteq \mathbb{R}^3$ be a finite set, and $\pi_{xy}(S)$ be its projection on the xy-plane, etc. Then

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

Proof. Let (X, Y, Z) be a uniform random point of S. Then

$$\begin{aligned} 2\log_2 |S| = 2H(X,Y,Z) &\leq H(X,Y) + H(X,Z) + H(Y,Z) \\ &\leq \log_2 \pi_{xy}(S) + \log_2 \pi_{xz}(S) + \log_2 \pi_{yz}(S). \end{aligned}$$ □

By approximating a body using cubes, we can deduce the following corollary.

Corollary 10.4.4
Let S be a body in $\mathbb{R}^3$. Then

$$\operatorname{vol}(S)^2 \leq \operatorname{area}(\pi_{xy}(S)) \operatorname{area}(\pi_{xz}(S)) \operatorname{area}(\pi_{yz}(S)).$$

Let us now state the general form of Shearer's lemma. (Chung, Graham, Frankl, and Shearer 1986)

Theorem 10.4.5 (Shearer's lemma)
Let $A_1, \ldots, A_s \subseteq [n]$ where each $i \in [n]$ appears in at least k sets A_j's. Writing $X_A := (X_i)_{i \in A}$,

$$kH(X_1, \ldots, X_n) \leq \sum_{j \in [s]} H(X_{A_j}).$$

The proof of the general form of Shearer's lemma is a straightforward adaptation of the proof of the special case earlier.

Like earlier, we can deduce an inequality about sizes of projections. (Loomis and Whitney 1949)

Corollary 10.4.6 (Loomis–Whitney inequality)

Writing π_i for the projection from $\mathbb{R}^n$ onto the hyperplane $x_i = 0$, we have for every $S \subseteq \mathbb{R}^n$,

$$|S|^{n-1} \le \prod_{i=1}^{n} |\pi_i(S)|$$

Corollary 10.4.7

Let $A_1, \dots, A_s \subseteq \Omega$ where each $i \in \Omega$ appears in at least k sets A_j. Then for every family $\mathcal{F}$ of subsets of Ω,

$$|\mathcal{F}|^k \le \prod_{j \in [s]} |\mathcal{F}|_{A_j}|$$

where $\mathcal{F}|_A := \{F \cap A : F \in \mathcal{F}\}$.

Proof. Each subset of Ω corresponds to a vector $(X_1, \dots, X_n) \in \{0,1\}^n$. Let $(X_1, \dots, X_n)$ be a random vector corresponding to a uniform element of $\mathcal{F}$. Then

$$k \log_2 |\mathcal{F}| = kH(X_1, \dots, X_n) \le \sum_{j \in [s]} H(X_{A_j}) = \log_2 \left|\mathcal{F}|_{A_j}\right|. \qquad \square$$

Triangle-intersecting families

We say that a set $\mathcal{G}$ of labeled graphs on the same vertex set is ***triangle-intersecting*** if $G \cap G'$ contains a triangle for every $G, G' \in \mathcal{G}$.

Question 10.4.8

What is the largest triangle-intersecting family of graphs on n labeled vertices?

The set of all graphs that contain a fixed triangle is triangle-intersecting, and they form a $1/8$ fraction of all graphs.

An easy upper bound: the edges form an intersecting family, so a triangle-intersecting family must be at most $1/2$ fraction of all graphs.

The next theorem improves this upper bound to $< 1/4$. It is also in this paper that Shearer's lemma was introduced.

Theorem 10.4.9 (Chung, Graham, Frankl, and Shearer 1986)

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

Proof. Let $\mathcal{G}$ be a triangle-intersecting family of graphs on vertex set $[n]$ (viewed as a collection of subsets of edges of K_n)

For $S \subseteq [n]$ with $|S| = \lfloor n/2 \rfloor$, let $A_S = \binom{S}{2} \cup \binom{[n]\setminus S}{2}$ (i.e., A_S is the union of the clique on S and the clique on the complement of S). Let

$$r = |A_S| = \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} \le \frac{1}{2}\binom{n}{2}.$$

For every S, every triangle has an edge in A_S, and thus $\mathcal{G}$ restricted to A_S must be an intersecting family. Hence

$$\left|\mathcal{G}|_{A_S}\right| \le 2^{|A_S|-1} = 2^{r-1}.$$

Each edge of K_n appears in at least

$$k = \frac{r}{\binom{n}{2}}\binom{n}{\lfloor n/2 \rfloor}$$

different A_S with $|S| = \lfloor n/2 \rfloor$ (by symmetry and averaging). Applying Corollary 10.4.7, we find that

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

Therefore

$$|\mathcal{G}| \le 2^{\binom{n}{2} - \frac{\binom{n}{2}}{r}} < 2^{\binom{n}{2}-2}. \qquad \square$$

***Remark* 10.4.10.** A tight upper bound of $2^{\binom{n}{2}-3}$ (matching the construction of taking all graphs containing a fixed triangle) was conjectured by Simonovits and Sós (1976) and proved by Ellis, Filmus, and Friedgut (2012) using Fourier analytic methods. Berger and Zhao (2023) gave a tight solution for K_4-intersecting families. The general conjecture for K_r-intersecting families is open.

The number of independent sets in a regular bipartite graph

Question 10.4.11

Fix d. Which d-regular graph on a given number of vertices has the most number of independent sets? Alternatively, which graph G maximizes $i(G)^{1/v(G)}$?

(Note that the number of independent sets is multiplicative: $i(G_1 \sqcup G_2) = i(G_1)i(G_2)$.)

Alon and Kahn conjectured that for graphs on n vertices, when n is a multiple of $2d$, a disjoint union of $K_{d,d}$'s maximizes the number of independent sets.

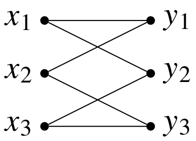
Alon (1991) proved an approximate version of this conjecture. Kahn (2001) proved it assuming the graph is bipartite. Zhao (2010) proved it in general.

Theorem 10.4.12 (Kahn, Zhao)
Let G be an n-vertex d-regular graph. Then

$$i(G) \le i(K_{d,d})^{n/(2d)} = (2^{d+1} - 1)^{n/(2d)}$$

where $i(G)$ is the number of independent sets of G.

Proof assuming G is bipartite. (Kahn) Let us first illustrate the proof for

$G =$ 

Among all independent sets of G, choose one uniformly at random, and let $(X_1, X_2, X_3, Y_1, Y_2, Y_3) \in \{0,1\}^6$ be its indicator vector. Then

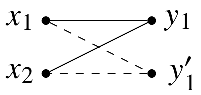
$$\begin{aligned} 2\log_2 i(G) &= 2H(X_1, X_2, X_3, Y_1, Y_2, Y_3) \\ &= 2H(X_1, X_2, X_3) + 2H(Y_1, Y_2, Y_3 | X_1, X_2, X_3) && \text{[chain rule]} \\ &\le H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\ &\qquad + 2H(Y_1|X_1, X_2, X_3) + 2H(Y_2|X_1, X_2, X_3) + 2H(Y_3|X_1, X_2, X_3) && \text{[Shearer]} \\ &= H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\ &\qquad + 2H(Y_1|X_1, X_2) + 2H(Y_2|X_1, X_3) + 2H(Y_3|X_2, X_3) && \text{[cond indep]} \end{aligned}$$

Here we are using that (a) Y_1, Y_2, Y_3 are conditionally independent given (X_1, X_2, X_3) and (b) Y_1 and (X_3, Y_2, Y_3) are conditionally independent given (X_1, X_2). A more general statement is that if $S \subseteq V(G)$, then the restrictions to the different connected components of $G - S$ are conditionally independent given X_S.

It remains to prove that

$$H(X_1, X_2) + 2H(Y_1|X_1, X_2) \le \log_2 i(K_{2,2})$$

and two other analogous inequalities. Let Y_1' be conditionally independent copy of Y_1 given (X_1, X_2). Then (X_1, X_2, Y_1, Y_1') is the indictor vector of an independent set of $K_{2,2}$ (though not necessarily chosen uniformly).



Thus we have

$$\begin{aligned} H(X_1, X_2) + 2H(Y_1|X_1, X_2) &= H(X_1, X_2) + H(Y_1|X_1, X_2) + H(Y_1'|X_1, X_2) \\ &= H(X_1, X_2, Y_1, Y_1') && \text{[chain rule]} \\ &\le \log_2 i(G) && \text{[uniform bound]} \end{aligned}$$

This concludes the proof for $G = K_{2,2}$, which works for all bipartite G. Here are the details.

Let $V = A \cup B$ be the vertex bipartition of G. Let $X = (X_v)_{v \in V}$ be the indicator function of an independent set chosen uniformly at random. Write $X_S := (X_v)_{v \in S}$. We have

$$\begin{aligned} d \log_2 i(G) = dH(X) &= dH(X_A) + dH(X_B|X_A) && \text{[chain rule]} \\ &\le \sum_{b \in B} H(X_{N(b)}) + d \sum_{b \in B} H(X_b|X_A) && \text{[Shearer]} \\ &\le \sum_{b \in B} H(X_{N(b)}) + d \sum_{b \in B} H(X_b|X_{N(b)}) && \text{[drop conditioning]} \end{aligned}$$

For each $b \in B$, we have

$$\begin{aligned} H(X_{N(b)}) + dH(X_b|X_{N(b)}) &= H(X_{N(b)}) + H(X_b^{(1)}, \dots, X_b^{(d)}|X_{N(b)}) \\ &= H(X_b^{(1)}, \dots, X_b^{(d)}, X_{N(b)}) \\ &\le \log_2 i(K_{d,d}) \end{aligned}$$

where $X_b^{(1)}, \dots, X_b^{(d)}$ are conditionally independent copies of X_b given $X_{N(b)}$. Summing over all b yields the result. □

Now we give the argument from Zhao (2010) that removes the bipartite hypothesis. The following combinatorial argument reduces the problem for non-bipartite G to that of bipartite G.

Starting from a graph G, we construct its ***bipartite double cover*** $G \times K_2$ (see Figure 10.1), which has vertex set $V(G) \times \{0, 1\}$. The vertices of $G \times K_2$ are labeled v_i for $v \in V(G)$ and $i \in \{0, 1\}$. Its edges are $u_0 v_1$ for all $uv \in E(G)$. Note that $G \times K_2$ is always a bipartite graph.

Lemma 10.4.13

Let G be any graph (not necessarily regular). Then

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

Once we have the lemma, Theorem 10.4.12 then reduces to the bipartite case, which we already proved. Indeed, for a d-regular G, since $G \times K_2$ is bipartite, the bipartite

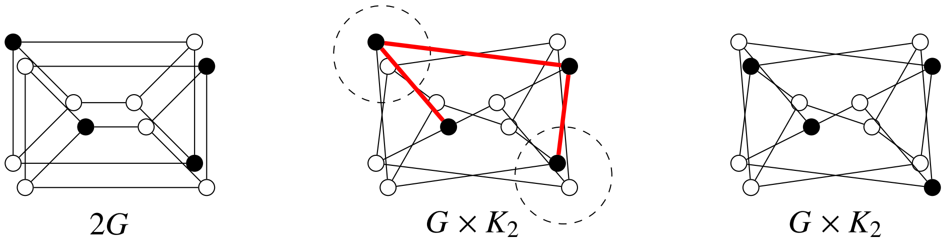


Figure 10.1: The bipartite swapping trick in the proof of Lemma 10.4.13: swapping the circled pairs of vertices (denoted A in the proof) fixes the bad edges (red and bolded), transforming an independent set of $2G$ into an independent set of $G \times K_2$.

case of the theorem gives

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

Proof of Lemma 10.4.13. Let $2G$ denote a disjoint union of two copies of G. Label its vertices by v_i with $v \in V$ and $i \in \{0,1\}$ so that its edges are $u_i v_i$ with $uv \in E(G)$ and $i \in \{0,1\}$. We will give an injection $\phi \colon I(2G) \to I(G \times K_2)$. Recall that $I(G)$ is the set of independent sets of G. The injection would imply $i(G)^2 = i(2G) \le i(G \times K_2)$ as desired.

Fix an arbitrary order on all subsets of $V(G)$. Let S be an independent set of $2G$. Let

$$E_{\text{bad}}(S) := \{uv \in E(G) : u_0, v_1 \in S\}.$$

Note that $E_{\text{bad}}(S)$ is a bipartite subgraph of G, since each edge of E_{bad} has exactly one endpoint in $\{v \in V(G) : v_0 \in S\}$ but not both (or else S would not be independent). Let A denote the first subset (in the previously fixed ordering) of $V(G)$ such that all edges in $E_{\text{bad}}(S)$ have one vertex in A and the other outside A. Define $\phi(S)$ to be the subset of $V(G) \times \{0,1\}$ obtained by "swapping" the pairs in A, i.e., for all $v \in A$, $v_i \in \phi(S)$ if and only if $v_{1-i} \in S$ for each $i \in \{0,1\}$, and for all $v \notin A$, $v_i \in \phi(S)$ if and only if $v_i \in S$ for each $i \in \{0,1\}$. It is not hard to verify that $\phi(S)$ is an independent set in $G \times K_2$. The swapping procedure fixes the "bad" edges.

It remains to verify that ϕ is an injection. For every $S \in I(2G)$, once we know $T = \phi(S)$, we can recover S by first setting

$$E'_{\text{bad}}(T) = \{uv \in E(G) : u_i, v_i \in T \text{ for some } i \in \{0,1\}\},$$

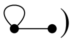
so that $E_{\text{bad}}(S) = E'_{\text{bad}}(T)$, and then finding A as earlier and swapping the pairs of A back. (Remark: it follows that $T \in I(G \times K_2)$ lies in the image of ϕ if and only if $E'_{\text{bad}}(T)$ is bipartite.) □

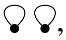
The entropy proof of the bipartite case of Theorem 10.4.12 extends to graph homomorphisms, yielding the following result.

Theorem 10.4.14 (Galvin and Tetali 2004)
Let G be an n-vertex d-regular bipartite graph. Let H be any graph allowing loops. Then

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

Some important special cases:

- $\hom(G,$  $) = i(G)$, the number of independent sets of G;
- $\hom(G, K_q)$ = the number of proper q-colorings of G.

The bipartite hypothesis in Theorem 10.4.14 cannot be always be removed. For example, if $H =$ , then $\log_2 \hom(G, H)$ is the number of connected components of G, so that the maximizers of $\log_2 \hom(G, H)/v(G)$ are disjoint unions of K_{d+1}'s.

For $H = K_q$, corresponding to the proper q-colorings, the bipartite hypothesis was recently removed.

Theorem 10.4.15 (Sah, Sawhney, Stoner, and Zhao 2020)
Let G be an n-vertex d-regular graph. Then

$$c_q(G) \le c_q(K_{d,d})^{n/(2d)}$$

where $c_q(G)$ is the number of q-colorings of G.

Furthermore, it was also shown in the same paper that in Theorem 10.4.14, the bipartite hypothesis on G can be weakened to triangle-free. Furthermore triangle-free is the weakest possible hypothesis on G so that the claim is true for all H.

For more discussion and open problems on this topic, see the survey by Zhao (2017).

Exercises

The problems in this section should be solved using entropy arguments or results derived from entropy arguments.

1. *Submodularity.* Prove that $H(X, Y, Z) + H(X) \le H(X, Y) + H(X, Z)$.
2. Let $\mathcal{F}$ be a collection of subsets of $[n]$. Let p_i denote the fraction of $\mathcal{F}$ that

contains i. Prove that

$$|\mathcal{F}| \le \prod_{i=1}^{n} p_i^{-p_i}(1-p_i)^{-(1-p_i)}.$$

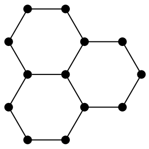
3. ⋆ *Uniquely decodable codes.* Let $[r]^*$ denote the set of all finite strings of elements in $[r]$. Let A be a finite subset of $[r]^*$ and suppose no two distinct concatenations of sequences in A can produce the same string. Let $|a|$ denote the length of $a \in A$. Prove that

$$\sum_{a \in A} r^{-|a|} \le 1.$$

4. *Sudoku.* A $n^2 \times n^2$ *Sudoku square* (the usual Sudoku corresponds to $n = 3$) is an $n^2 \times n^2$ array with entries from $[n^2]$ so that each row, each column, and, after partitioning the square into $n \times n$ blocks, each of these n^2 blocks consist of a permutation of $[n^2]$. Prove that the number of $n^2 \times n^2$ Sudoku squares is at most

$$\left(\frac{n^2}{e^3 + o(1)}\right)^{n^4}.$$

5. Prove Sidorenko's conjecture for the following graph.



6. ⋆ *Triangles versus vees in a directed graph.* Let V be a finite set, $E \subseteq V \times V$, and

$$\triangle = \left|\left\{(x,y,z) \in V^3 : (x,y),(y,z),(z,x) \in E\right\}\right|$$

(i.e., cyclic triangles; note the direction of edges) and

$$\wedge = \left|\left\{(x,y,z) \in V^3 : (x,y),(x,z) \in E\right\}\right|.$$

Prove that $\triangle \le \wedge$.

7. ⋆ *Box theorem.* Prove that for every compact set $A \subseteq \mathbb{R}^d$, there exists an axis-aligned box $B \subseteq \mathbb{R}^d$ with

$$\operatorname{vol} A = \operatorname{vol} B \qquad \text{and} \qquad \operatorname{vol} \pi_I(A) \ge \operatorname{vol} \pi_I(B) \quad \text{for all } I \subseteq [n].$$

Here π_I denotes the orthogonal projection onto the I-coordinate subspace.

(For the purpose of the homework, you only need to establish the case when A is a union of grid cubes. It is optional to give the limiting argument for compact A.)

8. Let $\mathcal{G}$ be a family of graphs on vertices labeled by $[2n]$ such that the intersection of every pair of graphs in $\mathcal{G}$ contains a perfect matching. Prove that $|\mathcal{G}| \leq 2^{\binom{2n}{2}-n}$.

9. *Loomis–Whitney for sumsets.* Let A, B, C be finite subsets of some abelian group. Writing $A + B := \{a + b : a \in A, b \in B\}$, etc., prove that

$$|A+B+C|^2 \leq |A+B|\,|A+C|\,|B+C|\,.$$

10. ⋆ *Shearer for sums.* Let X, Y, Z be independent random integers. Prove that

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