5 Graph Homomorphism Inequalities

CHAPTER HIGHLIGHTS

- A suite of techniques for proving inequalities between subgraph densities
- The maximum/minimum triangle density in a graph of given edge density.
- How to apply Cauchy–Schwarz and Hölder inequalities
- Lagrangian method (another proof of Turán's theorem, and inequalities between clique densities)
- Entropy method (and applications to Sidorenko's conjecture)

In this chapter, we study inequalities between graph homomorphism densities. Here is a typical example.

Question 5.0.1 (Linear inequality between homomorphism densities)
Given fixed graphs $F_1, \dots, F_k$ and reals $c_1, \dots, c_k$, does

$$c_1 t(F_1, G) + c_2 t(F_2, G) + \cdots + c_k t(F_k, G) \geq 0. \tag{5.0.1}$$

hold for all graphs G? Recall $t(F, G) = \hom(F, G)/v(G)^{v(F)}$.

Although the left-hand side is a linear combination of various graph homomorphism densities in G, polynomial combinations can also be written this way, as $t(F_1, G)t(F_2, G) = t(F_1 \sqcup F_2, G)$ where $F_1 \sqcup F_2$ is the disjoint union of the two graphs.

More generally, we would like understand constrained optimization problems in terms of graph homomorphism density. Many problems in extremal graph theory can be cast in this framework. For example, Turán's theorem from Chapter 1 on the maximum edge density of a K_r-free graph can be phrased in terms of the optimization problem

$$\text{maximize } t(K_2, G) \quad \text{subject to } t(K_r, G) = 0.$$

Turán's theorem (Corollary 1.2.6) says that the answer is $1/(r-1)$, achieved by $G = K_{r-1}$. We will see another proof of Turán's theorem in later in this Chapter, in Section 5.4 using the method of Lagrangians.

Remark 5.0.2 (Undecidability). Perhaps surprisingly, Question 5.0.1 is *undecidable* (for the question to make sense, we need to restrict the coefficients to a countable sets, say the rationals), as shown by Hatami and Norine (2011). This means that there is no algorithm that always correctly decides whether a given inequality is true for all graphs (however, it does not prevent us from proving/disproving specific inequalities). This undecidability stands in stark contrast to the decidability of polynomial inequalities over the reals, which follows from a classic result of Tarski (1948) that the first order theory of real numbers is decidable (via quantifier elimination). This undecidability of graph homomorphism inequalities is related to **Matiyasevich's theorem** (1970) (also known as the Matiyasevich–Robinson–Davis–Putnam theorem) giving a negative solution to **Hilbert's 10th Problem**, showing that diophantine equations are undecidable (equivalently: polynomial inequalities over the integers are undecidable). In fact, the proof of the former proceeds by converting polynomial inequalities over the integers to inequalities between $t(F, G)$ for various F.

As in the case of diophantine equations, the undecidability of graph homomorphism inequalities should be positively viewed as evidence of the richness of this space of problems. There are still many open problems, such as Sidorenko's inequality that we will see shortly.

Remark 5.0.3 (Graphs vs. graphons)**.** In the space of graphons with respect to the cut norm, $W \mapsto t(F, W)$ is continuous (by the counting lemma, Theorem 4.5.1), and graphs are a dense subset (Theorem 4.2.8). It follows any inequality for continuous functions of $t(F, G)$ over various F's (e.g., linear combinations as in Question 5.0.1) holds for all graphs G if and only if they hold for all graphons W in place of G. Furthermore, due to the compactness of the space of graphons, the extremum of continuous functions of F-densities is always attained at some graphon. The graphon formulation of the results can be often succinct and attractive.

For example, consider the following extremal problem (already mentioned in Chapter 4), where $p \in [0, 1]$ is a given constant,

$$\text{minimize } t(C_4, G) \quad \text{subject to } t(K_2, G) \geq p.$$

The minimum (or rather infimum) p^4 is not attained by any single graph, but rather by a sequence of quasirandom graphs (see Section 3.1). However, if we enlarge the space from graphs G to graphons W, then the minimizer is attained, in this case by the constant graphon p.

Sidorenko's conjecture and forcing conjecture

There are many important open problems on graph homomorphism inequalities. A major conjecture in extremal combinatorics is Sidorenko's conjecture (1993) (an equivalent conjecture was given earlier by Erdős and Simonovits).

Definition 5.0.4 (Sidorenko graphs)
We say that a graph F is **Sidorenko** if for every graph G,

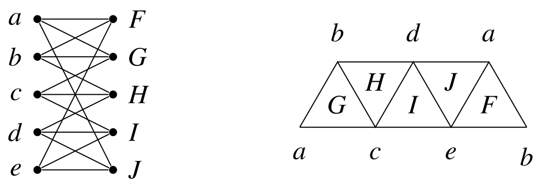
$$t(F, G) \geq t(K_2, G)^{e(F)}.$$

Conjecture 5.0.5 (Sidorenko's conjecture)
Every bipartite graph is Sidorenko.

In other words, the conjecture says that for a fixed bipartite graph F, the F-density in a graph of a given edge density is asymptotically minimized by a random graph. We will develop techniques in this chapter to prove several interesting special cases of Sidorenko's conjecture.

Every Sidorenko graph is necessarily bipartite. Indeed, given a non-bipartite F, we can take a non-empty bipartite G to get $t(F, G) = 0$ while $t(K_2, G) > 0$.

A notable open case of Sidorenko's conjecture is $F = K_{5,5} \setminus C_{10}$ (below left). This F is called the *Möbius graph* since it is the point-face incidence graph of a minimum simplicial decomposition of a Möbius strip (below right).



Sidorenko's conjecture has the equivalent graphon formulation: for every bipartite graph F and graphon W,

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

Note that equality occurs when $W \equiv p$, the constant graphon. One can think of Sidorenko's conjecture as a separate problem for each F, and asking to minimize $t(F, W)$ among graphons W with $\int W \geq p$. Whether the constant graphon is the unique minimizer is the subject of an even stronger conjecture known as the forcing conjecture.

Definition 5.0.6 (Forcing graphs)
We say that a graph F is **forcing** if every graphon W with $t(F, W) = t(K_2, W)^{e(F)}$ is a constant graphon (up to a set of measure zero).

By translating back and forth between graph limits and sequences of graphs, being forcing is equivalent to the quasirandomness condition. Thus any forcing graph can play the role of C_4 in Theorem 3.1.1. This is what led Chung, Graham, and Wilson to consider forcing graphs. In particular, C_4 is forcing.

Proposition 5.0.7 (Forcing and quasirandomness)
A graph F is forcing if and only if for every constant $p \in [0,1]$, every sequence of graphs $G = G_n$ with

$$t(K_2, G) = p + o(1) \qquad \text{and} \qquad t(F, G) = p^{e(F)} + o(1)$$

is quasirandom in the sense of Definition 3.1.2.

Exercise 5.0.8. Prove Proposition 5.0.7.

The forcing conjecture, below, states a complete characterization of forcing graphs (Skokan and Thoma 2004; Conlon, Fox, and Sudakov 2010).

Conjecture 5.0.9 (Forcing conjecture)
A graph is forcing if and only if it is bipartite and has at least one cycle.

Exercise 5.0.10. Prove the "only if" direction of the forcing conjecture.

Exercise 5.0.11. Prove that every forcing graph is Sidorenko.

Exercise 5.0.12 (Forcing and stability)**.** Show that a graph F is forcing if and only if for every $\epsilon > 0$, there exists $\delta > 0$ such that if a graph G satisfies $t(F, G) \le t(K_2, G)^{e(F)} + \delta$, then $\delta_\square(G, p) \le \epsilon$.

The following exercise shows that to prove a graph is Sidorenko, we do not lose anything by giving away a constant factor. The proof is a quick and neat application of the tensor power trick.

Exercise 5.0.13 (Tensor power trick)**.** Let F be a bipartite graph. Suppose there is some constant $c > 0$ such that

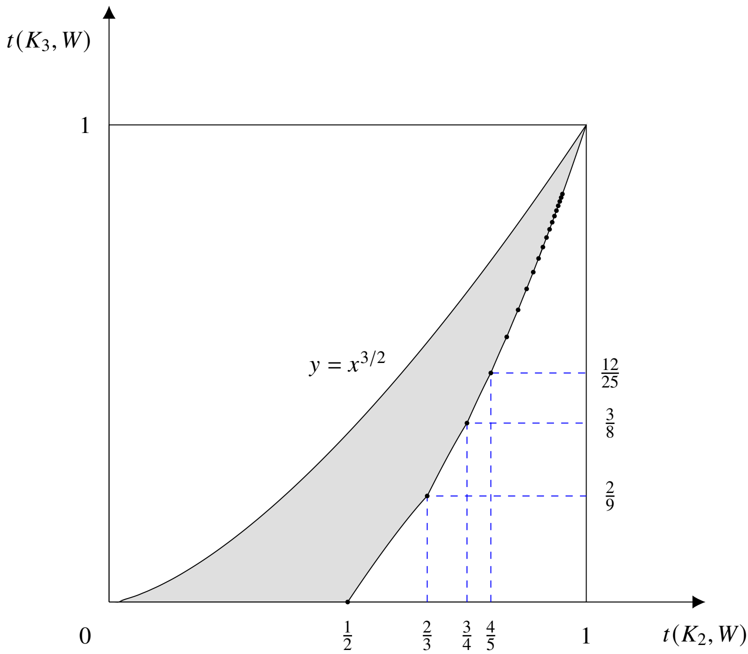
$$t(F, G) \ge c\, t(K_2, G)^{e(F)} \qquad \text{for all graphs } G.$$

Show that F is Sidorenko.

5.1 Edge vs. Triangle Densities

What are all the pairs of edge and triangles densities that can occur in a graph (or graphon)? Since the set of graphs is dense in the space of graphons, the closure of $\{(t(K_2, G), t(K_3, G)) : \text{graph } G\}$ is the

$$\textbf{edge-triangle region} := \{(t(K_2, W), t(K_3, W)) \;:\; \text{graphon } W\} \subset [0,1]^2. \qquad (5.1.1)$$



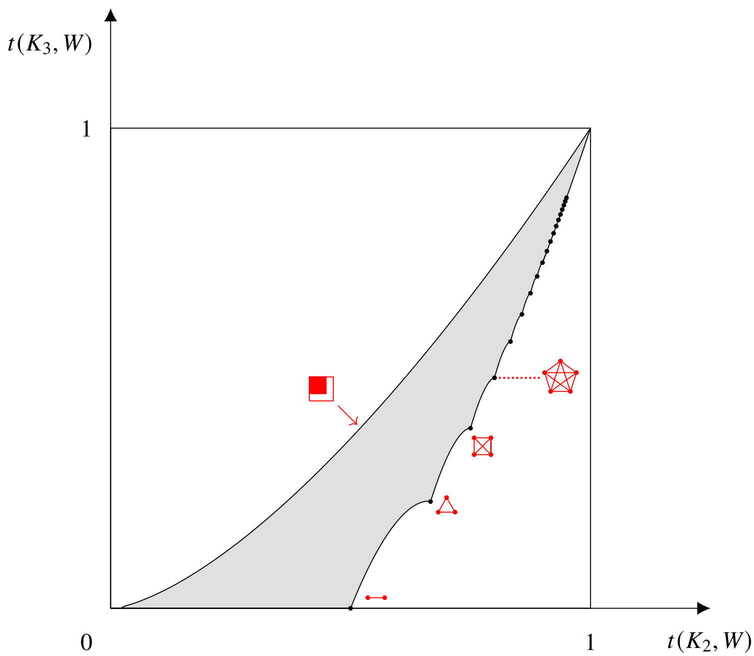


Figure 5.1.1: The top figure shows the edge-triangle region. This region is often depicted as in the bottom figure, which better highlights the concave scallops on the lower boundary but is a less accurate plot.

This is a closed subset of $[0,1]^2$, due to the compactness of the space of graphons. This set has been completely determined, and it is illustrated in Figure 5.1.1. We will discuss its features in this section.

The upper and lower boundaries of this region correspond to the answers of the following question.

Question 5.1.1 (Extremal triangle density given edge density)
Fix $p \in [0,1]$. What are the minimum and maximum possible $t(K_3, W)$ among all graphons with $t(K_2, W) = p$?

For a given $p \in [0,1]$, the set $\{t(K_3, W) : t(K_2, W) = p\}$ is a closed interval. Indeed, if W_0 achieves the minimum triangle density, and W_1 achieves the maximum, then their linear interpolation $W_t = (1-t)W_0 + tW_1$, ranging over $0 \le t \le 1$, must have triangle density continuously interpolating between those of W_0 and W_1, and therefore achieves every intermediate value.

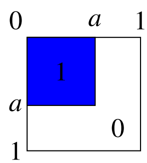
Maximum triangle density

The maximization part of Question 5.1.1 is easier. The answer is $p^{3/2}$.

Theorem 5.1.2 (Max triangle density)
For every graph G,
$$t(K_3, G) \le t(K_2, G)^{3/2}.$$

This inequality is asymptotically tight for G being a clique on a subset of vertices. The equivalent graphon inequality $t(K_3, W) \le t(K_2, W)^{3/2}$ attains equality for the clique graphon

$$W(x,y) = \begin{cases} 1 & \text{if } x, y \le a, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1.2}$$



For the above W, we have $t(K_3, G) = a^3$ while $t(K_2, G) = a^2$.

Proof. The quantities $\hom(K_3, G)$ and $\hom(K_2, G)$ count the number of closed walks in the graph of length 3 and 2, respectively. Let $\lambda_1 \ge \cdots \ge \lambda_n$ be the eigenvalues of the adjacency matrix A_G of G, then

$$\hom(K_3, G) = \operatorname{tr} A_G^3 = \sum_{i=1}^{k} \lambda_i^3 \qquad \text{and} \qquad \hom(K_2, G) = \operatorname{tr} A_G^2 = \sum_{i=1}^{k} \lambda_i^2$$

Then (see lemma below)

$$\hom(K_3, G) = \sum_{i=1}^{n} \lambda_i^3 \le \left(\sum_{i=1}^{n} \lambda_i^2\right)^{3/2} = \hom(K_2, G)^{3/2}.$$

After dividing by $v(G)^3$ on both sides, the result follows. □

Lemma 5.1.3 (A power sum inequality)
Let $t \ge 1$, and $a_1, \cdots, a_n \ge 0$. Then,

$$a_1^t + \cdots + a_n^t \le (a_1 + \cdots + a_n)^t.$$

Proof. Assume at least one a_i is positive, or else both sides equal to zero. Then

$$\frac{\text{LHS}}{\text{RHS}} = \sum_{i=1}^{n} \left(\frac{a_i}{a_1 + \cdots + a_n}\right)^t \le \sum_{i=1}^{n} \frac{a_i}{a_1 + \cdots + a_n} = 1. \qquad \square$$

Remark 5.1.4. We will see additional proofs of Theorem 5.1.2 not invoking eigenvalues later in Exercise 5.2.14 and in Section 5.3. Theorem 5.1.2 is an inequality in "physical space" (as opposed to going into the "frequency space" of the spectrum), and it is a good idea to think about how to prove it while staying in the physical space.

More generally, the clique graphon (5.1.2) also maximizes K_r-densities among all graphons of given edge density.

Theorem 5.1.5 (Maximum clique density)
For any graphon W and integer $k \ge 3$,

$$t(K_k, W) \le t(K_2, W)^{k/2}.$$

Proof. There exist integers $a, b \ge 0$ such that $k = 3a + 2b$ (e.g., take $a = 1$ if k is odd and $a = 0$ if k is even). Then $aK_3 + bK_2$ (a disjoint union of a triangles and b isolated edges) is a subgraph of K_k. So

$$t(K_k, W) \le t(aK_3 + bK_2, W) = t(K_3, W)^a t(K_2, W)^b \le t(K_2, W)^{3a/2+b} = t(K_2, W)^{k/2}. \quad \square$$

Remark 5.1.6 (Kruskal–Katona theorem). Thanks to a theorem of Kruskal (1963) and Katona (1968), the exact answer to the following non-asymptotic question is completely known:

What is the maximum number of copies of K_k's in an n-vertex graph with m edges?

When $m = \binom{a}{2}$ for some integer a, the optimal graph is a clique on a vertices. More generally, for any value of m, the optimal graph is obtained by adding edges in *colexicographic order*:

$$12, 13, 23, 14, 24, 34, 15, 25, 35, 45, \ldots$$

This is stronger than Theorem 5.1.5, which only gives an asymptotically tight answer as $n \to \infty$. The full Kruskal–Katona theorem also answers:

What is the maximum number of k-cliques in an r-graph with n vertices and m edges?

When $m = \binom{a}{r}$, the optimal r-graph is a clique on a vertices. (An asymptotic version of this statement can be proved using techniques in Section 5.3.) More generally, the optimal r-graph is obtained by adding the edges in colexicographic order. For example, for 3-graphs, the edges should be added in the following order:

$$123, 124, 134, 234, 125, 135, 235, 145, 245, 345, \ldots$$

Here $a_1 \ldots a_r < b_1 \ldots b_r$ in colexicographic order if $a_i < b_i$ at the last i where $a_i \neq b_i$ (i.e., dictionary order when read from right to left). Here we sort the elements of each r-tuple in increasing order.

The Kruskal–Katona theorem can be proved by a compression/shifting argument. The idea is to repeatedly modify the graph so that we eventually end up at the optimal graph. At each step, we "push" all the edges towards a clique along some "direction" in a way that does not reduce the number of k-cliques in the graph.

Minimum triangle density

Now we turn to the lower boundary of the edge-triangle region. What is the minimum triangle density in a graph of given edge density p?

For $p \leq 1/2$, we can have complete bipartite graphs of density $p + o(1)$, which are triangle-free. For $p > 1/2$, the triangle density must be positive due to Mantel's theorem (Theorem 1.1.1) and supersaturation (Theorem 1.3.4). It turns out that among graphs with edge density $p + o(1)$, the triangle density is asymptotically minimized by certain complete multipartite graphs, although this is not easy to prove.

For each positive integer k, we have

$$t(K_2, K_k) = 1 - \frac{1}{k} \qquad \text{and} \qquad t(K_3, K_k) = \left(1 - \frac{1}{k}\right)\left(1 - \frac{2}{k}\right).$$

As k ranges over all positive integers, these pairs form special points on the lower boundary of the edge-triangle region, as illustrated in Figure 5.1.1. (Recall that K_k is associated to the same graphon as a complete k-partite graph with equal parts.)

Now suppose the given edge density p lies strictly between $1 - 1/(k-1)$ and $1 - 1/k$ for some integer $k \geq 2$. To obtain the graphon with edge density p and minimum triangle density, we first start with K_k with all vertices having equal weight. And then shrink the relative weight of exactly one of the k vertices (while keeping the remaining $k-1$ vertices to have the same vertex weight). For example, the graphon illustrated below is obtained by starting with K_4 and shrinking the weight on one vertex.

	I_1	I_2	I_3	I_4
I_1	0	1	1	1
I_2	1	0	1	1
I_3	1	1	0	1
I_4	1	1	1	0

During this process, the total edge density (account for vertex weights) decreases continuously from $1-1/k$ to $1-1/(k-1)$. At some point, the edge density is equal to p. This vertex-weighted k-clique W turns out minimize triangle density among all graphons with edge density p.

The above claim is much more difficult to prove than the maximum triangle density result. This theorem, stated below, due to Razborov (2008), was proved using an involved Cauchy–Schwarz calculus that he coined *flag algebra*. We will say a bit more about this method in Section 5.2.

Theorem 5.1.7 (Minimum triangle density)

Fix $0 \le p \le 1$ and $k = \lceil 1/(1-p) \rceil$. The minimum of $t(K_3, W)$ among graphons W with $t(K_2, W) = p$ is attained by the stepfunction W associated to a k-clique with node weights $a_1, a_2, \cdots, a_k$ with sum equal to 1, $a_1 = \cdots = a_{k-1} \ge a_k$, and $t(K_2, W) = p$.

We will not prove this theorem in full here. See Lovász (2012, Section 16.3.2) for a presentation of the proof of Theorem 5.1.7. Later in this Chapter, we give lower bounds that match the edge-triangle region at the cliques. In particular, Theorem 5.4.4 will allow us to determine the convex hull of the region.

The graphon described in Theorem 5.1.7 turns out to be not unique unless $p = 1-1/k$ for some positive integer k. Indeed, suppose $1-1/(k-1) < p < 1-1/k$. Let $I_1, \dots, I_k$ be the partition of $[0,1]$ into the intervals corresponding to the vertices of the vertex-weighted k-clique, with $I_1, \dots, I_{k-1}$ all having equal length, and I_k strictly smaller length. We can replace the graphon on some $I_{k-1} \cup I_k$ by any triangle-free graphon without changing the edge density (why is this possible?).

	I_1	I_2	I_3	I_4
I_1	0	1	1	1
I_2	1	0	1	1
I_3	1	1	any triangle-free graphon	
I_4	1	1		

This operation does not change the edge-density or the triangle-density of the graphon (check!). The non-uniqueness of the minimizer hints at the difficulty of the result.

This completes our discussion of the edge-triangle region (Figure 5.1.1).

Theorem 5.1.7 was generalized from K_3 to K_4 (Nikiforov 2011), and then to all cliques K_r (Reiher 2016). The construction for the minimizing graphon is the same as for the triangle case.

Theorem 5.1.8 (Minimum clique density)
Fix $0 \le p \le 1$ and $k = \lceil 1/(1-p) \rceil$. The minimum of $t(K_r, W)$ among graphons W with $t(K_2, W) = p$ is attained by the stepfunction W associated to a k-clique with node weights $a_1, a_2, \cdots, a_k$ with sum equal to 1, $a_1 = \cdots = a_{k-1} \ge a_k$, and $t(K_2, W) = p$.

Exercise 5.1.9. Prove that C_6 is Sidorenko.

Hint: Write $\hom(C_6, G)$ and $\hom(K_2, G)$ in terms of the spectrum of G.

5.2 Cauchy–Schwarz

We will apply the Cauchy–Schwarz inequality in the following form: given real-valued functions f and g on the same space (always assuming the usual measurability assumptions without further comments), we have

$$\left(\int_X fg\right)^2 \le \left(\int_X f^2\right)\left(\int_X g^2\right).$$

It is one of the most versatile inequalities in combinatorics.

To better emphasize the variables being integrated, we write below the integral sign. The domain of integration (usually $[0,1]$ for each variable) is omitted to avoid clutter. We write

$$\int_{x,y,\ldots} f(x,y,\ldots) \qquad \text{for} \qquad \int f(x,y,\ldots)\,dxdy\cdots.$$

In practice, we will often apply the Cauchy–Schwarz inequality by changing the order of integration, and separating an integral into an outer integral and an inner integral.

A typical application of the Cauchy–Schwarz inequality is demonstrated in the following calculation (here one should think of x, y, z each as collections of variables):

$$\begin{aligned}\int_{x,y,z} f(x,y)g(x,z) &= \int_x \left(\int_y f(x,y)\right)\left(\int_z g(x,z)\right) \\ &\le \left(\int_x \left(\int_y f(x,y)\right)^2\right)^{1/2} \left(\int_x \left(\int_z g(x,z)\right)^2\right)^{1/2} \\ &= \left(\int_{x,y,y'} f(x,y)f(x,y')\right)^{1/2} \left(\int_{x,z,z'} g(x,z)g(x,z')\right)^{1/2}\end{aligned}$$

Note that in the final step, "expanding a square" has the effect of "duplicating a variable." It is useful to recognize expressions with duplicated variables that can be folded back into a square.

Let us warm up by proving that $K_{2,2}$ is Sidorenko. We actually already proved this statement in Proposition 3.1.14 in the context of the Chung–Graham–Wilson theorem on quasirandom graphs. We repeat the same calculations here to demonstrate the integral notation.

Theorem 5.2.1 ($K_{2,2}$ is Sidorenko)
$t(K_{2,2}, W) \ge t(K_2, W)^4$.

The theorem follows from the next two lemmas.

Lemma 5.2.2
$t(K_{1,2}, W) \ge t(K_2, W)^2$.

Proof.

$$\begin{aligned}t(K_{1,2}, W) = \int_{x,y,y'} W(x,y)W(x,y') &= \int_x \left(\int_y W(x,y)\right)^2 \\ &\ge \left(\int_{x,y} W(x,y)\right)^2 = t(K_2, W)^2.\end{aligned}$$ □

Lemma 5.2.3
$t(K_{2,2}, W) \ge t(K_{1,2}, W)^2$.

Proof.

$$\begin{aligned} t(K_{2,2}, W) &= \int_{x,y,z,z'} W(x,z)W(x,z')W(y,z)W(y,z') \\ &= \int_{x,y} \left(\int_z W(x,z)W(y,z) \right)^2 \geq \left(\int_{x,y,z} W(x,z)W(y,z) \right)^2 = t(K_{1,2}, W)^2. \end{aligned}$$

□

Proofs involving Cauchy–Schwarz are sometimes called "sum-of-square" proofs. The Cauchy–Schwarz inequality can be proved by writing the difference between the two sides as a sum of square quantity:

$$\left(\int f^2\right)\left(\int g^2\right) - \left(\int fg\right)^2 = \frac{1}{2}\int_{x,y} (f(x)g(y) - f(y)g(x))^2 .$$

Commonly, $g = 1$, in which case we can also write

$$\left(\int f^2\right) - \left(\int f\right)^2 = \int_x \left(f(x) - \int_y f(y)\right)^2 .$$

For example, We can write the proof of Lemma 5.2.3 as

$$t(K_{1,2}, W) - t(K_2, W)^2 \geq \int_x \left(\int_y W(x,y) - t(K_2, W) \right)^2 .$$

Exercise 5.2.4. Write $t(K_{2,2}, W) - t(K_2, W)^4$ as a single sum-of-squares expression.

The next inequality tells us that if we color the edges of K_n using two colors, then at least $1/4 + o(1)$ fraction of all triangles are monochromatic (Goodman 1959). Note that this $1/4$ constant is tight since it is obtained by a uniform random coloring. In the graphon formulation below, the graphons W and $1 - W$ correspond to edges of each color. We have equality for the constant $1/2$ graphon.

Theorem 5.2.5 (Triangle is common)

$$t(K_3, W) + t(K_3, 1 - W) \geq 1/4$$

Proof. Expanding, we have

$$\begin{aligned} t(K_3, 1 - W) &= \int (1 - W(x,y))(1 - W(x,z))(1 - W(y,z))\, dxdydz \\ &= 1 - 3t(K_2, W) + 3t(K_{1,2}, W) - t(K_3, W). \end{aligned}$$

So

$$\begin{aligned} t(K_3, W) + t(K_3, 1 - W) &= 1 - 3t(K_2, W) + 3t(K_{1,2}, W) \\ &\geq 1 - 3t(K_2, W) + 3t(K_2, W)^2 \\ &= \frac{1}{4} + 3\left(t(K_2, W) - \frac{1}{2}\right)^2 \geq \frac{1}{4}. \end{aligned}$$ □

Which graphs, other than triangles, have the above property? We do not know the full answer.

Definition 5.2.6 (Common graphs)
We say that a graph F is **common** if for all graphons W,

$$t(F, W) + t(F, 1 - W) \geq 2^{-e(F)+1}.$$

In other words, the left-hand side is minimized by the constant $1/2$ graphon.

Although it was initially conjectured that all graphs are common, this turns out to be false. In particular, K_t is not common for all $t \geq 4$ (Thomason 1989).

Proposition 5.2.7
Every Sidorenko graph is common.

Proof. Suppose F were Sidorenko. Let $p = t(K_2, W)$. Then $t(F, W) \geq p^{e(F)}$ and $t(F, 1 - W) \geq t(K_2, 1 - W)^{e(F)} = (1 - p)^{e(F)}$. Adding up and using convexity,

$$t(F, W) + t(F, 1 - W) \geq p^{e(F)} + (1 - p)^{e(F)} \geq 2^{-e(F)+1}.$$ □

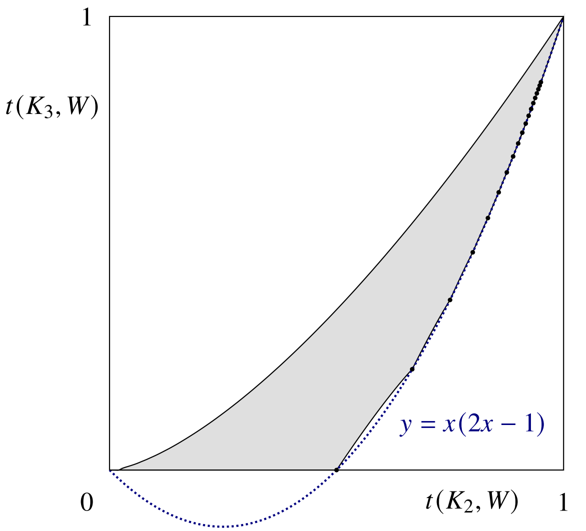
The converse is false. The triangle is common but not Sidorenko (recall that every Sidorenko graph is bipartite).

We also have the following lower bound on the minimum triangle density given edge density (Goodman 1959).

Theorem 5.2.8 (Lower bound on triangle density)

$$t(K_3, W) \geq t(K_2, W)(2t(K_2, W) - 1).$$

Below is plot of Goodman's bound against the true edge triangle region from Figure 5.1.1. The inequality is tight whenever $W = K_n$, in which case $t(K_2, W) = 1 - 1/n$ and $t(K_3, W) = \binom{n}{3}/n^3 = (1 - 1/n)(1 - 2/n)$. In particular, Goodman's bound implies that $t(K_3, W) > 0$ whenever $t(K_2, W) > 1/2$, which we saw from Mantel's theorem.



Proof. Since $0 \le W \le 1$, we have $(1 - W(x,z))(1 - W(y,z)) \ge 0$, and so

$$W(x,z)W(y,z) \ge W(x,z) + W(y,z) - 1.$$

Thus

$$\begin{aligned} t(K_3, G) &= \int_{x,y,z} W(x,y)W(x,z)W(y,z) \\ &\ge \int_{x,y,z} W(x,y)(W(x,z) + W(y,z) - 1) \\ &= 2t(K_{1,2}, W) - t(K_2, W) \\ &\ge 2t(K_2, W)^2 - t(K_2, W). \end{aligned}$$

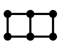
□

Finally, let us demonstrate an application of the Cauchy–Schwarz inequality in the following form, for nonnegative functions f and g:

$$\left(\int f^2 g\right)\left(\int g\right) \ge \left(\int fg\right).$$

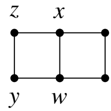
Recall that a graph F is Sidorenko if $t(F, W) \ge t(K_2, W)^{e(F)}$ for all graphons W (Definition 5.0.4).

Theorem 5.2.9

 is Sidorenko.

Proof. The idea is the "fold" the above graph F in half along the middle using the Cauchy–Schwarz inequality. Using w and x to indicate the two vertices in the middle, we have

$$t(F,W) = \int_{w,x,y,z} \left(\int W(w,y)W(y,z)W(z,x) \right)^2 W(w,x).$$



So

$$\begin{aligned} t(F,W)t(K_2,W) &\geq \left(\int_{w,x,y,z} \int W(w,y)W(y,z)W(z,x)W(w,x) \right)^2 \\ &= t(C_4,W)^2 \geq t(K_2,W)^8, \end{aligned}$$

with the last step due to Theorem 5.2.1. Therefore $t(F,W) \geq t(K_2,W)^7$ and hence F is Sidorenko. $\square$

Remark 5.2.10 (Flag algebra). The above examples were all simple enough to be found by hand. As mentioned earlier, every application of the Cauchy–Schwarz inequality can be rewritten in the form of a sum of a squares. One could actually search for these sum-of-squares proofs more systematically using a computer program. This idea, first introduced by Razborov (2007), can be combined with other sophisticated methods to determine the lower boundary of the edge-triangle region (Razborov 2008). Razborov coined the term **flag algebra** to describe a formalization of such calculations. The technique is also sometimes called **graph algebra**, **Cauchy–Schwarz calculus**, **sum-of-squares proof**.

Conceptually, the idea is that we are looking for all the ways to obtain nonnegative linear combinations of squared expressions. In a typical application, one is asked to solve an extremal problem of the form

$$\begin{array}{ll} \text{Minimize} & t(F_0,W) \\ \text{Subject to} & t(F_1,W) = q_1, \quad \ldots, \quad t(F_\ell,W) = q_\ell, \\ & W \text{ a graphon.} \end{array}$$

The technique is very flexible. The objectives and constraints could be any linear combinations of densities. It could be maximization instead of minimization. Extensions of the techniques can handle wider classes of extremal problems, such as for hypergraphs, directed graphs, edge-colored graphs, permutations, and more.

Let us illustrate the technique. The nonnegativity of squares implies inequalities such as

$$\int_{x,y,z} W(x,y)W(x,z) \left(\int_{u,w} (aW(x,u)W(y,u) - bW(x,w)W(w,u)W(u,z) + c) \right)^2 \geq 0.$$

Here $a, b, c \in \mathbb{R}$ are constants (to be chosen). We can expand the above expression, e.g.,

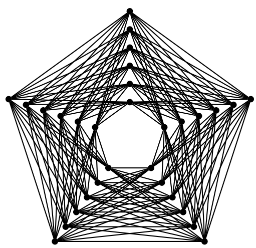
$$\text{replacing} \quad \left(\int_{u,w} G_{x,y,z}(u,w)\right)^2 \quad \text{by} \quad \int_{u,w,u',w'} G_{x,y,z}(u,w)G_{x,y,z}(u',w'),$$

we obtain a nonnegative linear combination of $t(F, W)$ over various F with undetermined real coefficients.

The idea is to now consider all such nonnegative expressions (in practice, on a computer, we consider a large but finite set of such inequalities). Then we try to optimize the previously undetermined real coefficients (a, b, c above). By adding together an optimized nonnegative linear combination of all such inequalities, and combining with the given constraints, we aim to obtain an inequality $t(F_0, W) \geq \alpha$ for some real α. This would prove a bound on the minimization problem stated earlier. We can find such coefficient and nonnegative combinations efficiently using a **semidefinite program (SDP)** solver. If we also happen to have an example of W satisfying the constraints and matching the bound, i.e., $t(F_0, W) = \alpha$, then we have solved the extremal problem.

The flag algebra method, with computer assistance, has successfully solved many interesting extremal problems in graph theory. For example, a conjecture of Erdős (1984) on the maximum pentagon density in a triangle-free graph was solved using flag algebra methods; the extremal construction is a blow-up of a 5-cycle (Grzesik 2012; Hatami, Hladký, Kráľ, Norine, and Razborov 2013).

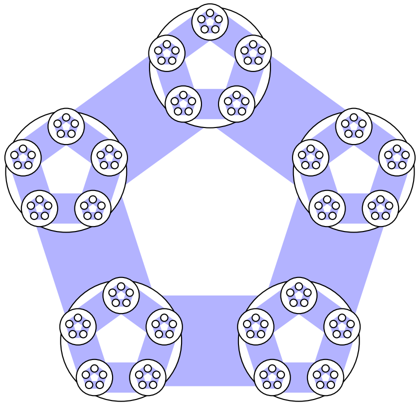
Theorem 5.2.11 (Maximum number 5-cycles in a triangle-free graph)
Every n-vertex triangle-free graph has at most $(n/5)^5$ cycles of length 5.



Let us mention another nice result obtained using the flag algebra method.

What is the maximum possible number of induced copies of a given graph H among all n-vertex graphs? (Pippenger and Golumbic 1975)

The optimal limiting density (as a fraction of $\binom{n}{v(H)}$, as $n \to \infty$) is called the **inducibility** of graph H. They conjectured that for every $k \geq 5$, the inducibility of a k-cycle is $k!/(k^k - k)$, obtained by an *iterated blow-up* of a k-cycle ($k = 5$ illustrated below; in the limit the should be infinitely many fractal-like iterations).



The conjecture for 5-cycles was proved by using flag algebra methods combined with additional "stability" methods (Balogh, Hu, Lidický, and Pfender 2016). The constant factor in the following theorem is tight.

Theorem 5.2.12 (Inducibility of the 5-cycle)
Every n-vertex graph has at most $n^5/(5^5 - 5)$ induced 5-cycles.

Although the flag algebra method has successfully solved several extremal problems, in many interesting cases, the method does not give a tight bound. Nevertheless, for many open extremal problems, such as the tetrahedron hypergraph Turán problem, the best known bound comes from this approach.

Remark 5.2.13 (Incompleteness)**.** Can every true linear inequality for graph homomorphism densities be proved via Cauchy–Schwarz/sum-of-squares?

Before giving the answer, we first discuss classical results about real polynomials. Suppose $p(x_1, \dots, x_n)$ is a real polynomial such that $p(x_1, \dots, x_n) \geq 0$ for all $x_1, \dots, x_n \in \mathbb{R}$. Can such a nonnegative polynomial always be written as a sum of squares? Hilbert (1888; 1893) proved that the answer is yes for $n \leq 2$ and no in general for $n \geq 3$. The first explicit counterexample was given by Motzkin (1967):

$$p(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$$

is always nonnegative due to the AM-GM inequality, but it cannot be written as a nonnegative sum of squares. Solving Hilbert's 17th problem, Artin (1927) proved that every $p(x_1, \dots, x_n) \geq 0$ can be written as a sum of squares of *rational* functions, i.e., there is some nonzero polynomial q such that pq^2 can be written as a sum of squares of polynomials. For the earlier example,

$$p(x, y) = \frac{x^2y^2(x^2 + y^2 + 1)(x^2 + y^2 - 2)^2 + (x^2 - y^2)^2}{(x^2 + y^2)^2}.$$

Turning back to inequalities between graph homomorphism densities, if $f(W) = \sum_i c_i t(F_i, W)$ is nonnegative for every graphon W, can f always be written as a nonnegative sum of squares of rational functions in $t(F, W)$? In other words, can every true inequality can be proved using a finite number of Cauchy–Schwarz inequalities (i.e., via vanilla flag algebra calculations).

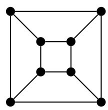
It turns out that the answer is no (Hatami and Norine 2011). Indeed, if there were always a sum-of-squares proof, then we could obtain an algorithm for deciding whether $f(W) \geq 0$ (with rational coefficients, say) holds for all graphons W, thereby contradicting the undecidability of the problem (Remark 5.0.2). Consider the algorithm that enumerates over all possible forms of sum-of-squares expressions (with undetermined coefficients that can then be solved for) and in parallel enumerates over all graphs G and checks whether $f(G) \geq 0$. If every true inequality had a sum-of-squares proof, then this algorithm would always terminate and tell us whether $f(W) \geq 0$ for all graphons W.

Exercise 5.2.14 (Another proof of maximum triangle density)**.** Let $W\colon [0,1]^2 \to \mathbb{R}$ be a symmetric measurable function. Write W^2 for the function taking value $W^2(x, y) = W(x, y)^2$.

1. Show that $t(C_4, W) \leq t(K_2, W^2)^2$.
2. Show that $t(K_3, W) \leq t(K_2, W^2)^{1/2} t(C_4, W)$.

Combining the two inequalities we deduce $t(K_3, W) \leq t(K_2, W^2)^{3/2}$, which is somewhat stronger than Theorem 5.1.2. We will see another proof below in Corollary 5.3.8.

Exercise 5.2.15. Prove that the skeleton of the 3-cube (below) is Sidorenko.



Exercise 5.2.16. Prove that K_4^- is common, where K_4^- is K_4 with one edge removed.

Exercise 5.2.17. Prove that every path is Sidorenko, by extending the proof of Theorem 5.3.4.

Exercise 5.2.18 (A lower bound on clique density)**.** Show that for every positive integer $r \geq 3$, and graphon W, writing $p = t(K_2, W)$,

$$t(K_r, W) \geq p(2p-1)(3p-2)\cdots((r-1)p-(r-2))\,.$$

Note that this inequality is tight when W is the associated graphon of a clique.

Exercise 5.2.19 (Triangle vs. diamond). Prove there is a function $f\colon [0,1] \to [0,1]$ with $f(x) \geq x^2$ and $\lim_{x\to 0} f(x)/x^2 = \infty$ such that

$$t(K_4^-, W) \geq f(t(K_3, W))$$

for all graphons W. Here K_4^- is K_4 with one edge removed.

Hint: Apply the triangle removal lemma

5.3 Hölder

Hölder's inequality is a generalization of the Cauchy–Schwarz inequality. It says that given $p_1, \ldots, p_k \geq 1$ with $1/p_1 + \cdots + 1/p_k = 1$, and real-valued functions $f_1, \ldots, f_k$ on a common space, we have

$$\int f_1 f_2 \cdots f_k \leq \|f_1\|_{p_1} \cdots \|f_k\|_{p_k},$$

where the ***p*-norm** of a function f is defined by

$$\|f\|_p := \left(\int |f|^p\right)^{1/p}.$$

In practice, the case $p_1 = \cdots = p_k = k$ of Hölder's inequality is used often.

We can apply Hölder's inequality to show that $K_{s,t}$ is Sidorenko. The proof is essentially verbatim to the proof of Theorem 5.2.1 that $t(K_{2,2}, W) \geq t(K_2, W)^4$ from the previous section, except that we now apply Hölder's inequality instead of the Cauchy–Schwarz inequality. We outline the steps below and leave the details as an exercise.

Theorem 5.3.1 (Complete bipartite graphs are Sidorenko)
$t(K_{s,t}, W) \geq t(K_2, W)^{st}$.

Lemma 5.3.2
$t(K_{s,1}, W) \geq t(K_2, W)^t$.

Lemma 5.3.3
$t(K_{s,t}, W) \geq t(K_{s,1}, W)^t$.

Sidorenko's conjecture for 3-edge path

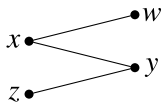
It is already quite a non-trivial fact that all paths are Sidorenko (Mulholland and Smith 1959; Atkinson, Watterson, and Moran 1960; Blakley and Roy 1965). You are encouraged to try it yourself before looking at the next proof.

Theorem 5.3.4

The 3-edge path is Sidorenko.

Let us give two short proofs that both appeared as answers to a MathOverflow question `https://mathoverflow.net/q/189222`. Later in Section 5.5 we will see another proof using the entropy method.

The first proof is a special case of a more general technique by Sidorenko (1991).



First proof that the 3-edge path is Sidorenko. Let P_4 be the 3-edge path. Let W be a graphon. Let $g(x) = \int_y W(x, y)$, representing the "degree" of vertex x. We have

$$t(P_4, W) = \int_{w,x,y,z} W(x, w)W(x, y)W(z, y) = \int_{x,y,z} g(x)W(x, y)W(z, y).$$

By relabeling, we can also write it as

$$t(P_4, W) = \int_{x,y,z} W(x, y)W(z, y)g(z).$$

Applying the Cauchy–Schwarz inequality twice, following by Hölder's inequality,

$$\begin{aligned} t(P_4, W) &= \left(\int_{x,y,z} g(x)W(x, y)W(z, y)\right)\left(\int_{x,y,z} g(x)W(x, y)W(z, y)\right) \\ &\geq \int_{x,y,z} \sqrt{g(x)}W(x, y)W(z, y)\sqrt{g(z)} \\ &= \int_y \left(\int_x \sqrt{g(x)}W(x, y)\right)^2 \\ &\geq \left(\int_{x,y} \sqrt{g(x)}W(x, y)\right)^2 \\ &= \left(\int_x g(x)^{3/2}\right)^2 \geq \left(\int_x g(x)\right)^3 = \left(\int_{x,y} W(x, y)\right)^3. \end{aligned}$$

□

The second proof is due to Lee (2019).

Second proof that the 3-edge path is Sidorenko. Define $g(x) = \int_y W(x, y)$ as earlier. We have

$$t(P_4, W) = \int_{w,x,y,z} W(x, w)W(x, y)W(z, y) = \int_{x,y} g(x)W(x, y)g(y).$$

Note that

$$\int_{x,y} \frac{W(x,y)}{g(x)} = \int_x \frac{g(x)}{g(x)} = 1.$$

Similarly we have

$$\int_{x,y} \frac{W(x,y)}{g(y)} = 1.$$

So by Hölder's inequality

$$\begin{aligned} t(P_4, W) &= \left(\int_{x,y} g(x)W(x,y)g(y) \right) \left(\int_{x,y} \frac{W(x,y)}{g(x)} \right) \left(\int_{x,y} \frac{W(x,y)}{g(y)} \right) \\ &\geq \left(\int_{x,y} W(x,y) \right)^3 . \end{aligned}$$

□

A generalization of Hölder's inequality

Now we discuss a powerful variant of Hölder's inequality due to Finner (1992), which is related more generally to Brascamp–Lieb inequalities. Here is a representative example.

Theorem 5.3.5 (Generalized Hölder inequality for a triangle)
Let X, Y, Z be measure spaces. Let $f : X \times Y \to \mathbb{R}$, $g : X \times Z \to \mathbb{R}$, and $h : Y \times Z \to \mathbb{R}$ be measurable functions (assuming integrability whenever needed). Then

$$\int_{x,y,z} f(x,y)g(x,z)h(y,z) \leq \|f\|_2 \|g\|_2 \|h\|_2 .$$

Note that a straightforward application of Hölder's inequality, when X, Y, Z are probability spaces (so that $\int_{x,y,z} f(x,y) = \int_{x,y} f(x,y)$) would yield

$$\int_{x,y,z} f(x,y)g(x,z)h(y,z) \leq \|f\|_3 \|g\|_3 \|h\|_3$$

which is implied by Theorem 5.3.5. Indeed, in a probability space, $\|f\|_p$ is nondecreasing as a function of p, which follows as a simple corollary of Hölder's inequality.

Proof of Theorem 5.3.5. We apply the Cauchy–Schwarz inequality three times. First to the integral over x (this affects f and g while leaving h intact):

$$\int_{x,y,z} f(x,y)g(x,z)h(y,z) \leq \int_{y,z} \left(\int_x f(x,y)^2 \right)^{1/2} \left(\int_x g(x,z)^2 \right)^{1/2} h(y,z).$$

Next, we apply the Cauchy–Schwarz inequality to the variable y (this affects f and h while leaving g intact). Continuing the above inequality,

$$\leq \int_z \left(\int_{x,y} f(x,y)^2\right)^{1/2} \left(\int_x g(x,z)^2\right)^{1/2} \left(\int_y h(y,z)^2\right)^{1/2}.$$

Finally, we apply the Cauchy–Schwarz inequality to the variable z (this affects g and h while leaving x intact). Continuing the above inequality,

$$\leq \left(\int_{x,y} f(x,y)^2\right)^{1/2} \left(\int_{x,z} g(x,z)^2\right)^{1/2} \left(\int_{y,z} h(y,z)^2\right)^{1/2}.$$

This completes the proof of Theorem 5.3.5. □

Remark 5.3.6 (Projection inequalities). What is the maximum volume of a body $K \subset \mathbb{R}^3$ whose projection on each coordinate plane is at most 1? A unit cube has volume 1, but is this the largest possible?

Letting $|\cdot|$ denote both volume and area (depending on the dimension) and $\pi_{xy}(K)$ denote the project of K onto the xy-plane, and likewise with the other planes. Using $1_K(x,y,z) \leq f(x,y)g(x,z)h(y,z)$, Theorem 5.3.5 implies

$$|K|^2 \leq \left|\pi_{xy}(K)\right| \left|\pi_{xz}(K)\right| \left|\pi_{yz}(K)\right|. \tag{5.3.1}$$

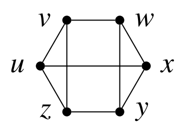
This shows that if all three projections have volume at most 1, then $|K| \leq 1$.

The inequality (5.3.1), which holds more generally in higher dimensions, is due to Loomis and Whitney (1949). It has important applications in combinatorics. A powerful generalization known as **Shearer's entropy inequality** will be discussed in Section 5.5.

Now let us state a more general form of Theorem 5.3.5, which can be proved using the same techniques. The key point of the inequality in Theorem 5.3.5 is that each variable (i.e., x, y, and z) is contained in exactly 2 of the factors (i.e., $f(x,y)$, $g(x,z)$, and $h(y,z)$). Everything works the same way as long as each variable is contained in exactly k factors, as long as we use L^k norms on the right-hand side.

For example,

$$\int_{u,v,w,x,y,z} f_1(u,v) f_2(v,w) f_3(w,z) f_4(x,y) \cdot f_5(y,z) f_6(z,u) f_7(u,x) f_8(u,z) f_9(w,y) \leq \prod_{i=1}^{9} \|f_i\|_3.$$



Here the factors in the integral correspond to edges of a 3-regular graph shown. In particular, every variable lies in exactly 3 factors.

More generally, each function f_i can take as input any number of variables, as long as every variable appears in exactly k functions. For example

$$\int_{w,x,y,z} f(w,x,y)g(w,y,z)h(x,z) \le \|f\|_2\|g\|_2\|h\|_2.$$

The inequality is stated more generally below. Given $x = (x_1, \ldots, x_m) \in X_1 \times \cdots \times X_m$ and $I \subset [m]$, we write $\pi_I(x) = (x_i)_{i\in I} \in \prod_{i\in I} X_i$ for the projection onto the coordinate subspace of I.

Theorem 5.3.7 (Generalized Hölder inequality)

Let $X_1, \ldots, X_m$ be measure spaces. Let $A_1, \ldots, A_\ell \subset [m]$ such that each element of $[m]$ appears in exactly k different A_i's. For each $i \in [m]$, let $f_i \colon \prod_{j\in A_i} X_j \to \mathbb{R}$. Then

$$\int_{X_1\times\cdots\times X_\ell} f_1(\pi_{A_1}(x)) \cdots f_\ell(\pi_{A_\ell}(x))\, dx \le \|f_1\|_k \cdots \|f_\ell\|_k\,.$$

Furthermore, if every X_i is a probability space, then we can relax the hypothesis to "each element of $[m]$ appears in *at most* k different A_i's.

The version of Theorem 5.3.7 with each X_i being a probability space is useful for graphons.

Corollary 5.3.8 (Upper bound on F-density)

For any graph F with maximum degree at most k, and graphon W,

$$t(F,W) \le \|W\|_k^{e(F)}\,.$$

In particular, since

$$\|W\|_k^k = \int W^k \le t(K_2, W),$$

the inequality implies that

$$t(F,W) \le t(K_2,W)^{e(F)/k}.$$

This implies the upper bound on clique densities (Theorems 5.1.2 and 5.1.5). The stronger statement of Corollary 5.3.8 with the L^k norm of W on the right-hand side has no direct interpretations for subgraph densities, but it is important for certain applications such as to understanding large deviation rates in random graphs (Lubetzky and Zhao 2017).

More generally, using different L^p norms for different factors in Hölder's inequality, we have the following statement (Finner 1992).

Theorem 5.3.9 (Generalized Hölder inequality)

Let $X_1, \dots, X_m$ be measure spaces. For each $i \in [\ell]$, let $p_i \ge 1$, let $A_i \subset [m]$, and $f_i \colon \prod_{j \in A_i} X_j \to \mathbb{R}$. If either

1. $\sum_{i : j \in A_i} 1/p_i = 1$ for each $j \in [m]$,
   OR
2. each X_i is a probability space and $\sum_{i : j \in A_i} 1/p_i \le 1$ for each $j \in [m]$,

then

$$\int_{X_1 \times \cdots \times X_\ell} f_1(\pi_{A_1}(x)) \cdots f_\ell(\pi_{A_\ell}(x))\, dx \le \|f_1\|_{p_1} \cdots \|f_\ell\|_{p_\ell} .$$

The proof proceeds by applying Hölder's inequality k times in succession, once for each variable $x_i \in X_i$, nearly identically to the proof of Theorem 5.3.5.

An application of generalized Hölder inequality

Now we turn to another graph inequality that where the above generalization of Hölder's inequality plays a key role.

Question 5.3.10

Fix d. Among d-regular graphs, which graph G maximizes $i(G)^{1/v(G)}$, where $i(G)$ denotes the number of independent sets of G.

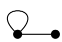
The answer turns out to be $G = K_{d,d}$. We can also take G to be a disjoint union of copies of $K_{d,d}$'s, and this would not change $i(G)^{1/v(G)}$. This result, stated below, was shown by Kahn (2001) for bipartite regular graphs G, and later extended by Zhao (2010) to all regular graphs G.

Theorem 5.3.11 (Maximum number of independent sets in a regular graph)

For every n-vertex d-regular graph G,

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

The set of independent sets of G is in bijection with the set of graph homomorphisms from G to the following graph:



Indeed, a map between their vertex sets form a graph homomorphism if and only if the vertices of G that map to the non-looped vertex is an independent set of G.

Let us first prove Theorem 5.3.11 for bipartite regular G. The following more general inequality was shown by Galvin and Tetali (2004). It implies the bipartite case of Theorem 5.3.11 by the above discussion.

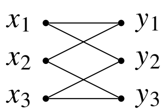
Theorem 5.3.12 (Upper bound on the number of H-colorings)
For every n-vertex d-regular bipartite graph G, and any graph H (allowing looped vertices on H)
$$\hom(G, H) \le \hom(K_{d,d}, H)^{n/(2d)}.$$

This is equivalent to the following statement.

Theorem 5.3.13
For any d-regular bipartite graph F,
$$t(F, W) \le t(K_{d,d}, W)^{e(F)/d^2}$$

Let us prove this theorem in the case $F = C_6$ to illustrate the technique more concretely. The general proof is basically the same. Let
$$f(x_1, x_2) = \int_y W(x_1, y)W(x_2, y).$$

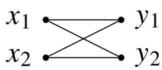
This function should be thought of the codegree of vertices x_1 and x_2. Then, grouping the factors in the integral according to their right-endpoint, we have



$$\begin{aligned}
t(C_6, W) &= \int_{x_1,x_2,x_3,y_1,y_2,y_3} W(x_1, y_1)W(x_2, y_1)W(x_1, y_2)W(x_3, y_2)W(x_2, y_3)W(x_2, y_3) \\
&= \int_{x_1,x_2,x_3} \left(\int_{y_1} W(x_1, y_1)W(x_2, y_1)\right)\left(\int_{y_2} W(x_1, y_2)W(x_3, y_2)\right) \\
&\qquad\qquad \cdot \left(\int_{y_3} W(x_2, y_3)W(x_2, y_3)\right) \\
&= \int_{x_1,x_2,x_3} f(x_1, x_2)f(x_1, x_3)f(x_2, x_3) \\
&\le \|f\|_2^3 \qquad \text{[by generalized Hölder, Theorem 5.3.7]}
\end{aligned}$$

On the other hand, we have

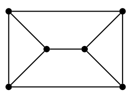
$$\begin{aligned}
\|f\|_2^2 &= \int_{x_1,x_2} f(x_1,x_2)^2 \\
&= \int_{x_1,x_2} \left(\int_{y_1} W(x_1,y_1)W(x_2,y_1)\right)\left(\int_{y_2} W(x_1,y_2)W(x_2,y_2)\right) \\
&= \int_{x_1,x_2,y_1,y_2} W(x_1,y_1)W(x_2,y_1)W(x_1,y_2)W(x_2,y_2) \\
&= t(C_4, W).
\end{aligned}$$



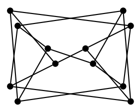
This proves Theorem 5.3.13 in the case $F = C_6$. The theorem in general can be proved via a similar calculation and left to the readers as an exercise.

Remark 5.3.14. Kahn (2001) first proved the bipartite case of Theorem 5.3.11 using Shearer's entropy inequality, which we will see in Section 5.5. His technique was extended by Galvin and Tetali (2004) to prove Theorem 5.3.12. The proof using generalized Hölder's inequality presented here was given by Lubetzky and Zhao (2017).

So far we proved Theorem 5.3.11 for bipartite regular graphs. To prove it for all regular graphs, we apply the following inequality by Zhao (2010). Here $G \times K_2$ (tensor product) is the bipartite double cover of G. An example is illustrated below:



G



$G \times K_2$

The vertex set of $G \times K_2$ is $V(G) \times \{0,1\}$. Its vertices are labeled v_i with $v \in V(G)$ and $i \in \{0,1\}$. Its edges are u_0v_1 for all $uv \in E(G)$. Note that $G \times K_2$ is always a bipartite graph.

Theorem 5.3.15 (Bipartite double cover for independent sets)
For every graph G,

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

Assuming Theorem 5.3.15, we can now prove Theorem 5.3.11 by reducing the statement to the bipartite case, which we proved earlier. Indeed, for every d-regular graph G,

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

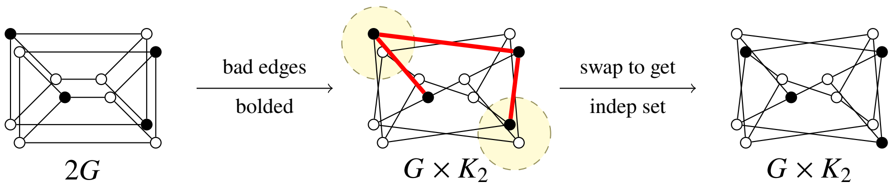
where the last step follows from applying Theorem 5.3.11 to the bipartite graph $G \times K_2$.

Proof of Theorem 5.3.15. 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.) $\square$

Remark 5.3.16 (Reverse Sidorenko)**.** Does Theorem 5.3.12 generalize to all regular graphs G like Theorem 5.3.11? Unfortunately, no. For example, when $H =$  consists of two isolated loops, $\hom(G, H) = 2^{c(G)}$, with $c(G)$ being the number of connected components of G. So $\hom(G, H)^{1/v(G)}$ is minimized among d-regular graphs G for $G = K_{d+1}$, which is the connected d-regular graph with the fewest vertices.

Theorem 5.3.12 actually extends to every triangle-free regular graph G. Furthermore, for every non-triangle-free regular graph G, there is some graph H for which the inequality in Theorem 5.3.12 fails.

There are several families interesting graphs H where Theorem 5.3.12 is known to extend to all regular bipartite G. Notably, this is true for $H = K_q$, which is significant since $\hom(G, K_q)$ is the number of proper q-colorings of G.

There are also generalizations of the above to non-regular graphs. For example, for a graph G without isolated vertices, letting d_u denote the degree of $u \in V(G)$, we have

$$i(G) \le \prod_{uv \in E(G)} i(K_{d_u, d_v})^{1/(d_u d_v)}.$$

And similarly for the number of proper q-colorings. In fact, the results mentioned in this remark about regular graphs are proved by induction on vertices of G, and thus require considering the larger family of not necessarily regular graphs G.

The results discussed in this remark are due to Sah, Sawhney, Stoner, and Zhao (2019; 2020). They introduced the term *reverse Sidorenko inequalities* to describe these inequalities $t(F, W)^{1/e(F)} \le t(K_{d,d}, W)^{1/d^2}$, which mirror the inequality $t(F, W)^{1/e(F)} \ge t(K_2, W)$ in Sidorenko's conjecture. Also see the earlier survey by Zhao (2017) for discussions of related results and open problems.

We already know through the quasirandom graph equivalences (Theorem 3.1.1) that C_4 is forcing. The following exercise generalizes this fact.

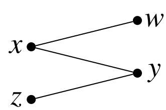
Exercise 5.3.17. Prove that $K_{s,t}$ is forcing whenever $s, t \ge 2$.

Exercise 5.3.18. Let F be a bipartite graph with vertex bipartition $A \cup B$ such that every vertex in B has degree d. Let d_u denote the degree of u in F. Prove that for every graphon W,

$$t(F, W) \le \prod_{uv \in E(G)} t(K_{d_u, d_v}, W)^{1/(d_u d_v)}.$$

Exercise 5.3.19 (Sidorenko for 3-edge path with vertex weights)**.** Let $W\colon [0,1]^2 \to [0, \infty)$ be a measurable function (not necessarily symmetric). Let $p, q, r, s\colon [0,1] \to [0, \infty)$ be measurable functions. Prove that

$$\int_{w,x,y,z} p(w)q(x)r(y)s(z)W(x,w)W(x,y)W(z,y)$$
$$\ge \left(\int_{x,y} (p(w)q(x)r(y)s(z))^{1/3} W(x,w) \right)^3.$$



Exercise 5.3.20. For a graph G, let $f_q(G)$ denote the number of maps $V(G) \to \{0, 1, \ldots, q\}$ such that $f(u) + f(v) \le q$ for every $uv \in E(G)$. Prove that for every n-vertex d-regular graph G (not necessarily bipartite),

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

5.4 Lagrangian

Here is another proof of Turán's theorem due to Motzkin and Straus (1965). It can be viewed as a continuous/analytic analogue of the Zykov symmetrization proof of Turán's theorem from Section 1.2 (the third proof there).

Theorem 5.4.1 (Turán theorem)

The number of edges in an n-vertex K_{r+1}-free graph is at most

$$\left(1-\frac{1}{r}\right)\frac{n^2}{2}.$$

Proof. Let G be a K_{r+1}-free graph on vertex set $[n]$. Consider the function

$$f(x_1,\dots,x_n)=\sum_{ij\in E(G)} x_i x_j.$$

We want to show that

$$f\left(\frac{1}{n},\dots,\frac{1}{n}\right)\le\frac{1}{2}\left(1-\frac{1}{r}\right).$$

In fact, we will show that

$$\max_{\substack{x_1,\dots,x_n\ge 0\\ x_1+\cdots+x_n=1}} f(x_1,\dots,x_n)\le\frac{1}{2}\left(1-\frac{1}{r}\right).$$

By compactness, the maximum is achieved at some $x=(x_1,\dots,x_n)$. Let us choose such a maximizing vector with the minimum support size (i.e., the number of nonzero coordinates).

Suppose $ij\notin E(G)$ for some pair of distinct $x_i,x_j>0$. If we replace (x_i,x_j) by (s,x_i+x_j-s), then f changes linearly in s (since x_ix_j does not come up as a summand in f), and since f is already maximized at x, it must not actually change with s. So we can replace (x_i,x_j) by $(x_i+x_j,0)$, which keeps f the same while decreasing the number of nonzero coordinates of x.

Thus the support of x is a clique in G. By labeling vertices, say that $x_1,\dots,x_k>0$ and $x_{k+1}=x_{k+2}=\cdots=x_n=0$. Since G is K_{r+1}-free, this clique has size $k\le r$. So

$$f(x)=\sum_{1\le i<j\le k} x_ix_j\le\frac{1}{2}\left(1-\frac{1}{k}\right)\left(\sum_{i=1}^k x_i\right)^2=\frac{1}{2}\left(1-\frac{1}{k}\right)\le\frac{1}{2}\left(1-\frac{1}{r}\right). \qquad\square$$

Remark 5.4.2 (Hypergraph Lagrangians). The **Lagrangian** of a hypergraph H with vertex set $[n]$ is defined to be

$$\lambda(H):=\max_{\substack{x_1,\dots,x_n\ge 0\\ x_1+\cdots+x_n=1}} f(x_1,\dots,x_n),\qquad \text{where } f(x_1,\dots,x_n)=\sum_{e\in E(H)}\prod_{i\in e} x_i.$$

It is a useful tool for certain hypergraph Turán problems. The above proof of Turán's theorem shows that for every graph G, $\lambda(G) = (1 - 1/\omega(G))/2$, where $\omega(G)$ is the size of the largest clique in G. A maximizing x has coordinate $1/\omega(G)$ on vertices of the clique and zero elsewhere.

As an alternate but equivalent perspective, the above proof can rephrased in terms of maximizing the edge density among K_{r+1}-free vertex-weighted graphs (vertex weights are given by the vector x above). The proof shifts weights between non-adjacent vertices while not decreasing the edge density, and this process preserves K_{r+1}-freeness.

The next theorem shows that to check whether a *linear* inequality in clique densities in G holds, it suffices to check it for G being cliques (Bollobás 1976; Schelp and Thomason 1998).

We first need the following lemma about the extrema of a symmetric polynomial over a simplex.

Lemma 5.4.3 (Extreme point of a symmetric polynomial)
Let $f(x_1, \dots, x_n)$ be a symmetric polynomial with real coefficients. Suppose $x = (x_1, \dots, x_n)$ minimizes $f(x)$ among all vectors $x \in \mathbb{R}^n$ with $x_1, \dots, x_n \geq 0$ and $x_1 + \cdots + x_n = 1$, and furthermore x has minimum support size among all such minimizers. Then, up to permuting the coordinates of x, there is some $1 \leq k \leq n$ so that

$$x_1 = \cdots = x_k = 1/k \quad \text{and} \quad x_{k+1} = \cdots = x_n = 0.$$

Proof. Suppose $x_1, \dots, x_k > 0$ and $x_{k+1} = \cdots = x_n = 0$ with $k \geq 2$. Fixing $x_3, \dots, x_n$, we see that as a function of (x_1, x_2), f has the form

$$Ax_1x_2 + Bx_1 + Bx_2 + C$$

where A, B, C depend on $x_3, \dots, x_n$. Notably the coefficients of x_1 and x_2 agree due since f is a symmetric polynomial. Holding $x_1 + x_2$ fixed, f has the form

$$Ax_1x_2 + C'.$$

If $A \geq 0$, then holding $x_1 + x_2$ fixed, we can set either x_1 or x_2 to be zero while not increasing f, which contradicts the hypothesis that the minimizing x has minimum support size. So $A < 0$, so that with $x_1 + x_2$ held fixed, $Ax_1x_2 + C'$ is minimized uniquely at $x_1 = x_2$. Thus $x_1 = x_2$. Likewise, $x_1 = \cdots = x_k$, as claimed. $\square$

Theorem 5.4.4 (Linear inequalities between clique densities)
Let $c_1, \cdots, c_\ell \in \mathbb{R}$. The inequality

$$\sum_{r=1}^{\ell} c_r t(K_r, G) \geq 0$$

is true for every graph G if and only if it is true with $G = K_n$ for every positive integer n.

More explicitly, the above inequality holds for all graphs G if and only if

$$\sum_{r=1}^{\ell} c_r \cdot \frac{n(n-1)\cdots(n-r+1)}{n^r} \geq 0 \qquad \text{for every } n \in \mathbb{N}.$$

Since this is a single variable polynomial in m, it is usually easy to check this inequality. We will see some examples right after the proof.

Proof. Suppose the displayed inequality holds for all cliques G. Let G be an arbitrary graph with vertex set $[n]$. Let

$$f(x_1, \ldots, x_n) = \sum_{r=1}^{\ell} r! c_r \sum_{\substack{\{i_1,\ldots,i_r\} \\ r\text{-clique in } G}} x_{i_1} \cdots x_{i_r}.$$

So

$$f(1/n, \ldots, 1/n) = \sum_{r=1}^{\ell} c_r t(K_r, G).$$

It suffices to prove that

$$\min_{\substack{x_1,\ldots,x_n \geq 0 \\ x_1+\cdots+x_n=1}} f(x_1, \ldots, x_n) \geq 0.$$

By compactness, we can assume that the minimum is attained at some x. Among all minimizing x, choose one with the smallest support (i.e., the number of nonzero coordinates).

As in the previous proof, if $ij \notin E(G)$ for some pair of distinct $x_i, x_j > 0$, then, replacing (x_i, x_j) by $(s, x_i + x_j - s)$, f changes linearly in s. Since f is already maximized at x, it must not change with s. So we can replace (x_i, x_j) by $(x_i + x_j, 0)$, which keeps f the same while decreasing the number of nonzero coordinates of x. Thus the support of x is a clique in G. Suppose x is supported on coordinates $[k]$ So f is a symmetric polynomial in $x_1, \ldots, x_k$. Lemma 5.4.3 implies that $x_1 = \cdots = x_k = 1/k$. Then $f(x) = \sum_{r=1}^{\ell} c_r t(K_r, K_k) \geq 0$ by hypothesis. $\square$

Remark 5.4.5. This proof technique can be adapted to show the stronger result that among all graphs G with a given number of vertices, the quantity $\sum_{r=1}^{\ell} c_r t(K_r, G)$ is minimized when G is a multipartite graph. Compare with the Zykov symmetrization proof of Turán's theorem (Theorem 1.2.4).

The theorem only considers linear inequalities between clique densities. The statement fails in general for inequalities with other graph densities (why?).

Theorem 5.4.4 can be equivalently instated in terms of the convex hull of the region of all possible clique density tuples.

Corollary 5.4.6 (Convex hull of feasible clique densities)
Let $\ell \geq 3$. In $\mathbb{R}^{\ell-1}$, the convex hull of

$$\{(t(K_2, W), t(K_3, W), \cdots, t(K_\ell, W)) : \text{graphons } W\}$$

is the same as the convex hull of

$$\{(t(K_2, K_n), t(K_3, K_n), \cdots, t(K_\ell, K_n)) : n \in \mathbb{N}\}\,.$$

For $\ell = 3$, the points

$$(t(K_2, K_n), t(K_3, K_n)) = \left(1 - \frac{1}{n}, \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\right), \qquad n \in \mathbb{N},$$

are the extremal points of the convex hull of the edge-triangle region from (5.1.1). The actual region, illustrated in Figure 5.1.1, has lower boundary consisting of concave curves connecting the points $(t(K_2, K_n), t(K_3, K_n))$.

This convex hull description easily implies Turán's theorem (exercise).

Exercise 5.4.7. For each graph F, let $c_F \in \mathbb{R}$ be such that $c_F \geq 0$ whenever F is not a clique (no restrictions when F is a clique). Assume that $c_F \neq 0$ for finitely many F's. Prove that the inequality

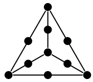
$$\sum_F c_F t_{\text{inj}}(F, G) \geq 0$$

is true for every graph G if and only if it is true with $G = K_n$ for every positive integer n.

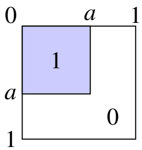
Exercise 5.4.8 (Cliquey edges)**.** Let n, r, t be nonnegative integers. Show that every n-vertex graph with at least $(1 - \frac{1}{r})\frac{n^2}{2} + t$ edges contains at least rt edges that belong to a K_{r+1}.

Hint: Rephrase the statement as an inequality between the number of edges and the number of cliquey edges in every graph.

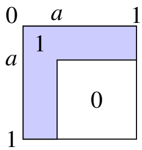
Exercise 5.4.9. Let F be the 3-graph with 10 vertices and 6 edges illustrated below (lines denotes edges). Prove that the hypergraph Turán density of F is $2/9$.



Exercise 5.4.10* (Maximizing $K_{1,2}$ density)**.** Prove that, for every $p \in [0, 1]$, among all graphons W with $t(K_2, W) = p$, the maximum possible value of $t(K_{1,2}, W)$ is attained by either a "clique" or a "hub" graphon, illustrated below.



clique graphon
$W(x, y) = 1_{\max\{x,y\} \le a}$



hub graphon
$W(x, y) = 1_{\min\{x,y\} \le a}$

5.5 Entropy

In this section, we explain how to use entropy to prove certain graph homomorphism inequalities.

Entropy basics

Definition 5.5.1 (Entropy)

Let X be a discrete random variable taking values in some set S. For each $s \in S$, let $p_s = \mathbb{P}(X = s)$. We define the **(binary) entropy** of X to be

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

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

Exercise 5.5.2. Show that $H(X) \ge 0$ always.

Intuitively, $H(X)$ measures the amount of "surprise" in the randomness of X. A more rigorous interpretation of this intuition is given by the **Shannon noiseless coding theorem**, which says that the minimum number of bits needed to encode n independent copies of X is $nH(X) + o(n)$.

Here are some basic properties of entropy.

Lemma 5.5.3 (Uniform bound)

If X is a random variable supported on a finite set S, then

$$H(X) \leq \log_2 |S| .$$

Equality holds if and only if X is uniformly distributed on S.

Proof. Let function $f(x) = -x \log_2 x$ is concave for $x \in [0, 1]$. We have, by concavity,

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

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

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

In particular,

$$H(X, Y) = H(X) + H(Y) \qquad \text{if } X \text{ and } Y \text{ are independent.}$$

We can similarly define $H(X, Y, Z)$, etc.

Definition 5.5.4 (Conditional entropy)

Given jointly distributed discrete random variables X and Y, define

$$H(X|Y) := \sum_{y} \mathbb{P}(Y = y) H(X|Y = y).$$

Here $H(X|Y = y) = \sum_x -\mathbb{P}(X = x|Y = y) \log_2 \mathbb{P}(X = x|Y = y)$ is entropy of the random variable X conditioned on the event $Y = y$.

Intuitively, $H(X|Y)$ measures the expected amount of new information or surprise in X after Y has already been revealed. For example:

- If X is completely determined by Y, i.e., $X = f(Y)$ for some function f, 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 5.5.5 (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 (below we skip y if $p(y) = 0$)

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

□

Lemma 5.5.6 (Subadditivity)
$H(X, Y) \leq H(X) + H(Y)$. More generally,

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

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

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

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

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

□

Remark 5.5.7. 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 5.5.8 (Dropping conditioning)
$H(X|Y) \le H(X)$. 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 5.5.9. Another way to state the dropping condition inequality is the **data processing inequality**: $H(X|f(Y)) \ge H(X|Y)$ for any function f.

Applications to Sidorenko's conjecture

Now let us use entropy to establish some interesting cases of Sidorenko's conjecture. Recall that a bipartite graph F is said to be **Sidorenko** if

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

for every graph G. Sidorenko's conjecture says that every bipartite graph is Sidorenko.

The entropy approach to Sidorenko's conjecture was first introduced by Li and Szegedy (2011) and further developed in subsequent works (Szegedy (2015); Conlon, Kim, Lee, and Lee (2018)). Here we illustrate the entropy approach to Sidorenko's conjecture with several examples.

To show that F is Sidorenko, we need to show that for every graph G,

$$\frac{\hom(F,G)}{v(G)^{v(F)}} \ge \left(\frac{2e(G)}{v(G)^2}\right)^{e(F)}. \tag{5.5.1}$$

We write Hom(F, G) for the the set of all maps $V(F) \to V(G)$ that give a graph homomorphism $F \to G$. This set has cardinality hom(F, G). Our strategy is to construct a random element $\Phi \in \mathrm{Hom}(F, G)$ whose entropy satisfies

$$H(\Phi) \geq e(F) \log_2(2e(G)) - (2e(F) - v(F)) \log_2 v(G). \tag{5.5.2}$$

The uniform bound $H(\Phi) \leq \log_2 \mathrm{hom}(F, G)$ then implies (5.5.1).

Let us illustrate this technique for a three-edge path. We had already seen two proofs of the following inequality in Section 5.3. Now we present a different proof using the entropy method along with generalizations.

Theorem 5.5.10

The 3-edge path is Sidorenko.

Proof. Let P_4 denote the 3-edge path and G a graph. An element of Hom(P_4, G) is a walk of length three. 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.

A key observation is that YZ is also distributed as a uniform random edge of G (pause and think about why). 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 identically distributed, and hence YZ is a uniform random edge of G.

Similarly, ZW is distributed as uniform random edge.

Also, since X and Z are conditionally independent given Y

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

Furthermore,

$$H(Y|X) = H(Z|Y) = H(W|Z)$$

since XY, YZ, ZW are identically distributed as a uniform random edge.

Thus

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

This proves (5.5.2), and thus shows that P_4 is Sidorenko. Indeed, by the uniform bound,

$$\log_2 \hom(P_4, F) \geq H(X,Y,Z,W) \geq 3\log_2(2e(G)) - 2\log_2 v(G),$$

and hence

$$t(P_4, G) = \frac{\hom(P_4, G)}{v(G)^4} \geq \left(\frac{2e(G)}{v(G)^2}\right)^3 = t(K_2, G)^3. \qquad \square$$

Let us outline how to extend the above proof strategy from the 3-edge path to any tree T. Define a **T-branching random walk** in a graph G to a random $\Phi \in \mathrm{Hom}(T, G)$ defined by fixing an arbitrary root v of T (the choice of v will not matter in the end). Then set $\Phi(v)$ to be a random vertex of G with each vertex of G chosen proportional to its degree. Then extend Φ to a random homomorphism $T \to G$ one vertex at a time: if $u \in V(T)$ is already mapped to $\Phi(u)$ and $w \in V(T)$ has not yet been mapped, then set $\Phi(w)$ to be a uniform random neighbor of $\Phi(u)$, independent of all other choices. The resulting random $\Phi \in \mathrm{Hom}(T, G)$ has the following properties:

- for each edge of T, its image under Φ is a uniform random edge of G (in the sense of the proof of Theorem 5.5.10); and
- for each vertex v of T, conditioned on $\Phi(v)$, the neighbors of v in T are mapped by Φ to conditionally independent and uniform neighbors of $\Phi(v)$ in G.

Furthermore, as in the proof of Theorem 5.5.10,

$$\begin{aligned} H(\Phi) &= e(T)\log_2(2e(G)) - (e(T)-1)H(\Phi(v)) \\ &\geq e(T)\log_2(2e(G)) - (e(T)-1)\log_2 v(G). \end{aligned} \tag{5.5.3}$$

(Exercise: fill in the details.) Together with the uniform bound $H(\Phi) \leq \log_2 \hom(T, G)$, we proved the following.

Theorem 5.5.11

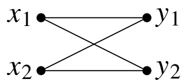
Every tree is Sidorenko.

We saw earlier that $K_{s,t}$ is Sidorenko, which can be proved by two applications of Hölder's inequality (see Section 5.3). Here let us give another proof using entropy. This entropy proof is subtler than the earlier Hölder's inequality proof, but it will soon lead us more naturally to the next generalization.

Theorem 5.5.12

Every complete bipartite graph is Sidorenko.

Let us demonstrate the proof for $K_{2,2}$ for concreteness. The same proof extends to all $K_{s,t}$.



Proof that $K_{2,2}$ is Sidorenko. As earlier, we construct a random element of $\mathrm{Hom}(K_{2,2}, G)$. Pick a random $(X_1, X_2, Y_1, Y_2) \in V(G)^4$ 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 some attention. It does *not* say that we choose a uniform random common neighbor of Y_1 and Y_2, as one might naively attempt. Instead, one can think of the first two steps as defining the $K_{1,2}$-branching random walk for (X_1, Y_1, Y_2). Under this distribution, we can first sample (Y_1, Y_2) according to its marginal, and then produce two conditionally independent copies of X_1 (with the second copy now called X_2).

We have

$$
\begin{aligned}
&H(X_1, X_2, Y_1, Y_2) \\
&= H(Y_1, Y_2) + H(X_1, X_2 | Y_1, Y_2) && \text{[chain rule]} \\
&= H(Y_1, Y_2) + 2H(X_1 | Y_1, Y_2) && \text{[cond. indep.]} \\
&= 2H(X_1, Y_1, Y_2) - H(Y_1, Y_2) && \text{[chain rule]} \\
&\geq 2(2\log_2(2e(G)) - \log_2 v(G)) - H(Y_1, Y_2). && \text{[(5.5.3)]} \\
&\geq 2(2\log_2(2e(G)) - \log_2 v(G)) - 2\log_2 v(G). && \text{[uniform bound]} \\
&= 4\log(2e(G)) - 4\log_2 v(G).
\end{aligned}
$$

Together with the uniform bound $H(X_1, X_2, Y_1, Y_2) \leq \log_2 \hom(K_{2,2}, G)$, we deduce that $K_{2,2}$ is Sidorenko. □

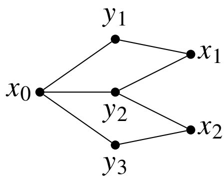
Exercise 5.5.13. Complete the proof of Theorem 5.5.12 for general $K_{s,t}$.

The following result was first proved by Conlon, Fox, and Sudakov (2010) using the dependent random choice technique. The entropy proof was found later by Li and Szegedy (2011).

Theorem 5.5.14

Let F be a bipartite graph that has a vertex adjacent to all vertices in the other part. Then F is Sidorenko.

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



Proof that the above graph is Sidorenko. Pick $(X_0, X_1, X_2, Y_1, Y_2, Y_3) \in V(G)^6$ randomly 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).

We have the following properties:

- 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).

So (the 1st and 4th steps by chain rule, and the 2nd and 3rd steps by conditional independence)

$$\begin{aligned}
&H(X_0, X_1, X_2, Y_1, Y_2, Y_3) \\
&= H(X_0, X_1, X_2|Y_1, Y_2, Y_3) + H(Y_1, Y_2, Y_3) \\
&= H(X_0|Y_1, Y_2, Y_3) + H(X_1|Y_1, Y_2, Y_3) + H(X_2|Y_1, Y_2, Y_3) + H(Y_1, Y_2, Y_3) \\
&= H(X_0|Y_1, Y_2, Y_3) + H(X_1|Y_1, Y_2) + H(X_2|Y_2, Y_3) + H(Y_1, Y_2, Y_3) \\
&= H(X_0, Y_1, Y_2, Y_3) + H(X_1, Y_1, Y_2) + H(X_2, Y_2, Y_3) - H(Y_1, Y_2) - H(Y_2, Y_3).
\end{aligned}$$

By (5.5.3),

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

And by the uniform bound,

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

Putting everything together, we have

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

Thereby verifying (5.5.2), showing that F is Sidorenko. □

(Where did we use the assumption that F has vertex complete to the other part?)

Exercise 5.5.15. Complete the proof of Theorem 5.5.14.

Shearer's inequality

Another important tool in the entropy method is Shearer's inequality, which is a powerful generalization of subadditivity. Before stating it in full generality, let us first see a simple instance of Shearer's lemma.

Theorem 5.5.16 (Shearer's entropy inequality, 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

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X), \\ H(X, Z) &= H(X) \qquad\qquad\quad + H(Z|X), \\ \text{and} \quad H(Y, Z) &= \qquad\qquad\quad H(Y) + H(Z|Y). \end{aligned}$$

Adding up, and applying conditioning dropping $H(Y) \ge H(Y|X)$, we see that their sum is at at least

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

with the final equality due to the chain rule. □

Here is the general form of Shearer's inequality (Chung, Graham, Frankl, and Shearer 1986).

Theorem 5.5.17 (Shearer's entropy inequality)

Let $A_1, \dots, A_s \subset [n]$ where each $i \in [n]$ appears in at least k sets A_j's. Let $X_1, \dots, X_n$ be a jointly distributed discrete random variables. Writing $X_A := (X_i)_{i \in A}$, we have

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

Exercise 5.5.18. Prove Theorem 5.5.17 by generalizing the proof of Theorem 5.5.16.

Shearer's entropy inequality is related to the generalized Hölder inequality from Section 5.3. It is a significant generalization of the projection inequality discussed in Remark 5.3.6. See Friedgut (2004) for more discussion about these connections.

Let us use the entropy method to give another proof of Theorem 5.3.12, restated below.

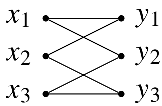
Theorem 5.5.19

For every n-vertex d-regular bipartite graph F, and any graph G (allowing looped vertices on G)

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

This proof follows the original entropy proof of Galvin and Tetali (2004), which was in turn based on the proof by Kahn (2001) for independent sets.

Proof. Let us first illustrate the proof for F being the following graph



Choose $\Phi \in \operatorname{Hom}(F, G)$ uniformly at random among all homomorphisms from F to G. Let $X_1, X_2, X_3, Y_1, Y_2, Y_3 \in V(G)$ be the respective images of the vertices of G. We have

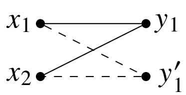
$$\begin{aligned}
&2 \log_2 \hom(F, G) \\
&\quad = 2H(X_1, X_2, X_3, Y_1, Y_2, Y_3) \\
&\quad = 2H(X_1, X_2, X_3) + 2H(Y_1, Y_2, Y_3 | X_1, X_2, X_3) && \text{[chain rule]} \\
&\quad \le H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\
&\qquad + 2H(Y_1 | X_1, X_2, X_3) + 2H(Y_2 | X_1, X_2, X_3) + 2H(Y_3 |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

In the final step, we use that X_3 and Y_1 are conditionally independent given X_1 and X_2 (why?), along with two other analogous statements. A more general statement is that if $S \subset V(F)$, then the restrictions to the different connected components of $F - S$ are conditionally independent given $(X_s)_{s \in S}$.

To complete the proof, it remains to show

$$\begin{aligned}
&H(X_1, X_2) + 2H(Y_1 | X_1, X_2) \le \log_2 \hom(K_{2,2}, G), \\
&H(X_1, X_3) + 2H(Y_2 | X_1, X_3) \le \log_2 \hom(K_{2,2}, G), \\
\text{and} \quad &H(X_2, X_3) + 2H(Y_3 | X_2, X_3) \le \log_2 \hom(K_{2,2}, G).
\end{aligned}$$

They are analogous so let us just show the first inequality. Let Y_1' be a conditionally independent copy of Y_1 given (X_1, X_2). Then (X_1, X_2, Y_1, Y_1') is the image of a homomorphism from $K_{2,2}$ to G (though not necessarily chosen uniformly).



Thus we have

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

This concludes the proof for $F = K_{2,2}$.

Now let F be an arbitrary bipartite graph with vertex bipartition $V = A \cup B$. Let $\Phi \in \operatorname{Hom}(F, G)$ be chosen uniformly at random. For each $v \in V$, let $X_v = \Phi(v)$. For each $S \subset V$, write $X_S := (X_v)_{v \in S}$. We have

$$\begin{aligned} d \log_2 \hom(F, G) = dH(\Phi) &= dH(X_A) + dH(X_B | X_A) && \text{[chain rule]} \\ &\le \sum_{b \in B} H(X_{N(b)}) + d \sum_{b \in B} H(X_b | X_A) && \text{[Shearer]} \\ &= \sum_{b \in B} H(X_{N(b)}) + d \sum_{b \in B} H(X_b | X_{N(b)}). && \text{[cond. indep.]} \end{aligned}$$

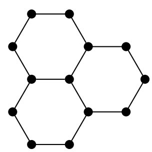
For each $b \in B$, let $X_b^{(1)}, \ldots, X_b^{(d)}$ be conditionally independent copies of X_b given $X_{N(b)}$. We have

$$\begin{aligned} H(X_{N(b)}) + dH(X_b | X_{N(b)}) &= H(X_{N(b)}) + H(X_b^{(1)}, \ldots, X_b^{(d)} | X_{N(b)}) \\ &= H(X_b^{(1)}, \ldots, X_b^{(d)}, X_{N(b)}) && \text{[chain rule]} \\ &\le \log_2 \hom(K_{d,d}, G). && \text{[uniform bound]} \end{aligned}$$

Summing over all $b \in B$, and using the previous equality, we obtain

$$d \log_2 \hom(F, G) \le d \log_2 \hom(K_{d,d}, G). \qquad \square$$

Exercise 5.5.20. Prove that the following graph is Sidorenko.



Exercise 5.5.21 (△ vs. ∧ in a directed graph)**.** Let V be a finite set, $E \subset V \times V$, 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 $\triangle \le \wedge$.

Chapter Summary

- Many problems in extremal graph theory can be phrased in terms of graph homomorphism inequalities.
  - Homomorphism density inequalities are undecidable in general, though a suite of techniques are available.
  - Many open problems remain, such as **Sidorenko's conjecture**, which says that if F is bipartite, then $t(F,G) \geq t(K_2,G)^{e(F)}$ for all graphs G.
- The set of all possible **(edge, triangle) density pairs** is known.
  - For a given edge density, the maximum triangle density is maximized by a clique.
  - For a given edge density, the minimum triangle density is given by a certain multipartite graph. (We did not prove this result in full and only established the convex hull in Section 5.4.)
- **Cauchy–Schwarz** and **Hölder** inequalities are versatile tools.
  - Simple applications of Cauchy–Schwarz inequalities can often be recognized by "reflection symmetries" in a graph, i.e., being able to "fold" a graph in half.
  - **Flag algebra** leads to computerized searches of Cauchy–Schwarz proofs of subgraph density inequalities.
  - **Generalized Hölder inequality** tells us that, as an example,

$$\int_{x,y,z} f(x,y)g(x,z)h(y,z) \leq \|f\|_2 \|g\|_2 \|h\|_2 .$$

    It can be proved by repeated applications of Hölder's inequality, once for each variable. The inequality is related to **Shearer's entropy inequality**, an example of which says that for joint random variables X, Y, Z,

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

- The **Lagranian method** relaxes an optimization problem on graphs to one about vertex-weighted graphs, and then argue by shifting weights between vertices. We used the method to prove
  - Turán's theorem (again);
  - A linear inequality between clique densities in G is true and only if it holds whenever G is a clique.
- The **entropy method** can be used to establish various cases of Sidorenko's conjecture, including for trees, as well as for a bipartite graph with one vertex complete to the other side.

Further Reading

The book *Large Networks and Graph Limits* by Lovász (2012) contains an excellent treatment of graph homomorphism inequalities in Section 2.1 and Chapter 16.

The survey *Flag Algebras: An Interim Report* by Razborov (2013) contains a survey of results obtained using the flag algebra method.

For combinatorial applications of the entropy method, see the surveys

- *Entropy and Counting* by Radhakrishnan (2003), and
- *Three Tutorial Lectures on Entropy and Counting* by Galvin (2014).