# Graphons and Entropy

Andrew Hah

Spring 2025

## Contents

# 1 Introduction

This paper is, in essence, a walk through my short journey thus far in mathematics. I start with extremal graph theory, arguably the first area that really sparked my excitement, and move through several ideas that drew my attention along the way, particularly from extremal combinatorics. The path eventually leads to the theory of graphons, which I arrived at somewhat naturally. This is not necessarily a comprehensive survey, but a personal exploration of the concepts and questions that have shaped my thinking so far.

# 2 Preliminaries

## 2.1 Extremal Graph Theory

My first encounter, which quickly turned into a fascination, with extremal graph theory was in my discrete mathematics course taught by Prof. A. Razborov. One classical question that we were introduced to was of the form, given a fixed forbidden subgraph $H$, what is the maximum number of edges a graph on $n$ vertices can have without containing $H$ as a (not necessarily induced) subgraph? This number is called the Turán number, denoted $\mathrm{ex}(n, H)$. What is often considered the first result in extremal graph theory is Mantel's theorem, which answers this question for the special case when $H = K_3$ is a triangle.

**Theorem 2.1.** *(Mantel's theorem) Every $n$-vertex triangle-free graph has at most $\lfloor n^2/4 \rfloor$ edges, i.e.,* $\mathrm{ex}(n, K_3) = \lfloor n^2/4 \rfloor$.

To generalize this theorem from triangles to arbitrary cliques, we first construct the Turán graph.

**Definition 2.2.** The *Turán graph* $T_{n,r}$ is defined to be the complete $n$-vertex $r$-partite graph with part sizes differing by at most 1 (so each part has size $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$.

For example, $T_{3,1} = K_3$ and $T_{10,3} = K_{3,3,4}$.

**Theorem 2.3.** *(Turán's theorem) The Turán graph $T_{n,r}$ maximizes the number of edges among all $n$-vertex $K_{r+1}$-free graphs. It is also the unique maximizer.*

**Corollary 2.4.** $|E(T_{n,r})| \sim \binom{r}{2} \left(\frac{n}{r}\right)^2$.

## 2.2 Graph Homomorphisms

While extremal graph theory traditionally focuses on maximizing or minimizing the number of edges under subgraph constraints, many of its central questions can be reframed more generally in terms of how frequently small patterns occur within larger graphs. This motivates the study of graph homomorphisms (adjacency-preserving maps), which capture not just the presence of a subgraph, but the frequency with which it appears.

Note. Although graph homomorphisms were not the next topic I encountered chronologically, presenting them here provides a natural and useful continuation of the ideas introduced in the previous section.

**Definition 2.5.** For two finite simple graphs $F$ and $G$, a *homomorphism* from $F$ to $G$ is a map $\phi : V(F) \to V(G)$ such that if $uv \in E(F)$, then $\phi(u)\phi(v) \in E(G)$. We define

$$\mathrm{Hom}(F, G) = \{\text{homomorphisms from } F \text{ to } G\},$$

and

$$\mathrm{hom}(F, G) = |\mathrm{Hom}(F, G)|.$$

**Definition 2.6.** Define the *$F$-homomorphism density in $G$* as

$$t(F, G) = \frac{\mathrm{hom}(F, G)}{|V(G)|^{|V(F)|}},$$

i.e., the probability that a random map of $V(F)$ into $V(G)$ is a homomorphism.

Note. This definition works well with dense graphs, which are what we consider in this paper, but other normalizations are more appropriate for sparse graphs.

**Example 2.7.** Using this new notion of homomorphism densities, we can recall Corollary 2.4 to write

$$t(K_2, T_{n,r}) = \frac{\mathrm{hom}(K_2, T_{n,r})}{n^2} \sim \frac{2\binom{r}{2}\left(\frac{n}{r}\right)^2}{n^2} = 1 - \frac{1}{r}.$$

Similarly, the triangle density can be written

$$t(K_3, T_{n,r}) \sim \frac{6\binom{r}{3}\left(\frac{n}{r}\right)^3}{n^3} = \left(1 - \frac{1}{r}\right)\left(1 - \frac{2}{r}\right).$$

**Example 2.8.**

(i) A walk in $G$ is a homomorphism of a path into $G$, so $\mathrm{hom}(P_k, G)$ counts the number of walks with $k - 1$ steps in $G$.

(ii) In terms of colorings, $\mathrm{hom}(G, K_q)$ counts the number of colorings of the graph $G$ with $q$ colors, satisfying the usual condition that adjacent nodes must be assigned different colors.

As the simple examples above illustrate, homomorphism numbers have many applications. We now present several properties that emphasize their utility and what they can express.

**Proposition 2.9.**

*(i) If $F_1$ and $F_2$ are node-disjoint, then*

$$\mathrm{hom}(F_1 \cup F_2, G) = \mathrm{hom}(F_1, G)\mathrm{hom}(F_2, G).$$

*(ii) If $F$ is connected and $G_1$ and $G_2$ are node-disjoint, then*

$$\mathrm{hom}(F, G_1 \cup G_2) = \mathrm{hom}(F, G_1) + \mathrm{hom}(F, G_2).$$

*(iii)*

$$\hom(F, G_1 \times G_2) = \hom(F, G_1)\hom(F, G_2).$$

*Proof.*

(i) Since $F_1$ and $F_2$ are node-disjoint, a homomorphism $\phi : V(F_1 \cup F_2) \to V(G)$ is equivalent to specifying a homomorphism $\phi_1 : V(F_1) \to V(G)$ and a homomorphism $\phi_2 : V(F_2) \to V(G)$ independently.

(ii) Since $G_1$ and $G_2$ are node-disjoint, a homomorphism $\phi : V(F) \to V(G_1 \cup G_2)$ must map all of $F$ into just one of $G_1$ or $G_2$.

(iii) Each homomorphism $\phi : V(F) \to V(G_1 \times G_2)$ corresponds uniquely to a pair of homomorphisms $\phi_1 : F \to G_1$ and $\phi_2 : F \to G_2$. ■

**Theorem 2.10.** *[Lov12, Theorem 5.29] Either one of the simple graph parameters* $\hom(., G)$ *and* $\hom(G, .)$ *determines a simple graph* $G$*.*

*Proof Sketch.* Consider two simple graphs $G, G'$ such that $\mathrm{inj}(F, G) = \mathrm{inj}(F, G')$ for every simple graph $F$, where $\mathrm{inj}(F, G)$ is the number of injective homomorphisms from $F$ into $G$. Then $G$ and $G'$ have injective homomorphisms into each other and thus are isomorphic. A nontrivial result (see [Lov12, Section 5.4]) shows that the injective homomorphism counts are determined by the ordinary homomorphism counts. Hence, if $\hom(F, G) = \hom(F, G')$ for all simple graphs $F$, then $G \cong G'$. □

## 2.3 Entropy in Combinatorics

We now return to the actual chronological path I took. After first encountering extremal graph theory, I soon found myself drawn to the field of extremal combinatorics as a whole, to which I give credit to Extremal Combinatorics by S. Jukna [Juk11] and The Probabilistic Method by N. Alon and J. H. Spencer [AS16]. After lots of reading, I realized that I was especially excited by the elegance and power of entropy, as it offered a way to reason about combinatorial structure through information, rather than direct counting. In this section, I introduce the basic definitions and ideas behind entropy, with a focus on results that will later connect to graph theory and extremal problems.

**Definition 2.11.** Let $X$ be a random variable taking values in some range $S$. For each $s \in S$, let $p_s = \mathbb{P}(X = s)$. The *(binary) entropy* of $X$ is defined by

$$H(X) = \sum_{s \in S} p_s \log_2 \frac{1}{p_s} = \sum_{s \in S} -p_s \log_2 p_s,$$

i.e., the expected "information gain" of $X$.

**Remark 2.12.** Going forward, we use log to denote $\log_2$, the binary logarithm.

We now prove several basic properties of entropy.

**Lemma 2.13.** *(Uniform Bound) If $X$ is a random variable with finite support $S$,*

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

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

*Proof.* Since the function $f(x) = -x \log x$ is concave, by Jensen's inequality, we have

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

For the entropy of *joint random variables* $(X, Y)$, we write,

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

In particular, $H(X, Y) = H(X) + H(Y)$ if $X$ and $Y$ are independent.

**Definition 2.14.** Given jointly distributed random variables $X$ and $Y$, we define the *conditional entropy*,

$$H(X \mid Y) = \sum_{y} \mathbb{P}(Y = y) H(X \mid Y = y),$$

i.e., $H(X \mid Y)$ is the information gain of $X$ given a particular value of $Y$, averaged over the range of values that $Y$ can take.

**Example 2.15.**

(i) If $X$ and $Y$ are independent, $H(X \mid Y) = H(X)$.
(ii) If $X$ is completely determined by $Y$, $H(X \mid Y) = 0$.

**Lemma 2.16.** *(Chain Rule)*

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

*Proof.* We write $p(x) = \mathbb{P}(X = x)$. Then

$$\begin{aligned}
H(Y \mid X) &= \sum_{x} p(x) H(Y \mid X = x) = \sum_{x} -p(x) \sum_{y} p(y \mid x) \log p(y \mid x) \\
&= \sum_{x} \sum_{y} -p(x, y) \log \left(\frac{p(x, y)}{p(x)}\right) = \sum_{x,y} -p(x, y) \log p(x, y) + \sum_{x} p(x) \log p(x) \\
&= H(X, Y) - H(X).
\end{aligned}$$

■

**Lemma 2.17.** *(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 $f(x) = \log(1/x)$. Then

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

We can iterate to obtain the general case. □

**Lemma 2.18.**

$$H(X \mid Y) \leq H(X).$$

*Proof.* By the chain rule and subadditivity,

$$H(X \mid Y) = H(X, Y) - H(Y) \leq H(X)$$

□

**Theorem 2.19.**

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

*Proof.* By the chain rule and Lemma 2.18,

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

Summing,

$$\begin{aligned}
H(X, Y) + H(X, Z) + H(Y, Z) &\geq 2H(X) + 2H(Y \mid X) + 2H(Z \mid X, Y) \\
&= 2H(X, Y, Z).
\end{aligned}$$

□

This theorem is a special case of a much more useful inequality, Shearer's inequality.

**Theorem 2.20.** *(Shearer's Inequality) Let $A_1, \ldots, A_s \subseteq [n]$, where each $i \in [n]$ appears in at least $k$ sets $A_i$. Let $X_1, \ldots, X_n$ be jointly distributed discrete random variables. Writing $X_A = (X_i)_{i \in A}$, we have*

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

# 3 Graph Limits

The next topic in my studies was the theory of graph limits. My interest in it was deepened by the fact that I was concurrently taking the real analysis sequence here, so the idea that large, discrete graphs could converge to a continuous object was not only surprising, but also very satisfying.

As we saw in Section 2.1, many questions in extremal graph theory ask about the asymptotic behavior of graphs as the number of vertices grows large, so it is a natural idea that large graphs should be studied not as discrete objects, but as approximations of some continuous limit. However, there are several challenges in making this idea rigorous. Graphs are inherently combinatorial, and standard notions of convergence, such as pointwise convergence of adjacency matrices, fail to capture the structural similarities we care about.

This motivates the development of a continuous theory of limits for graphs, one in which graphs are viewed through the lens of their statistical behavior. The framework that achieves this is the theory of graph limits, where the central objects are graphons, which provide not only a natural language for describing convergence and similarity, but also connect with powerful analytic tools, such as compactness and continuity. I develop this theory step by step in this section.

## 3.1 Graphons

To define a meaningful limit object for dense graphs, Lovász and Szegedy introduced graphons [LS06], analytic objects that generalize the adjacency matrix of a graph into a measurable function on the unit square.

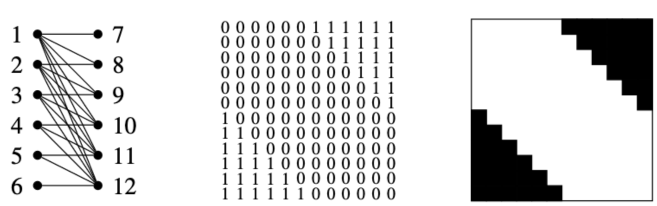
**Definition 3.1.** A *graphon* is a symmetric measurable function $W : [0,1]^2 \to [0,1]$.

Every graph $G$ has an associated graphon $W_G$.

**Definition 3.2.** Given a graph $G$ with $n$ vertices labeled $1, \ldots, n$, we define its *associated graphon* $W_G : [0,1]^2 \to [0,1]$ by first partitioning $[0,1]$ into $n$ equal-length intervals $I_1, \ldots, I_n$ and setting $W_G$ to be 1 on all $I_i \times I_j$ where $ij$ is an edge of $G$, and 0 on all other $I_i \times I_j$.

**Definition 3.3.** A *step graphon* $W$ with $k$ steps consists of first partitioning $[0,1]$ into $k$ intervals $I_1, \ldots, I_k$ and then setting $W$ to be a constant on each $I_i \times I_j$.

**Example 3.4.** Consider the bipartite graph on $2n$ vertices with one vertex part $\{v_1, \ldots, v_n\}$ and the other $\{w_1, \ldots, w_n\}$, with edges $v_i w_j$ whenever $i \leq j$. The graph, its adjacency matrix, and associated graphon are pictured below.



Intuitively, we can think of $W(x, y)$ as describing the "edge density" or probability of an edge between the points $x$ and $y$ in the unit interval. In this setting, the interval $[0, 1]$ replaces the finite vertex set of a graph, and the function $W$ plays the role of a generalized adjacency matrix.

To develop the theory of graph limits further, we need a way to measure how "close" two graphons are. There are two natural and ultimately equivalent approaches to defining this similarity, one based on the cut distance, and the other based on comparing homomorphism densities.

### 3.2 Cut Distance

The cut distance provides a natural notion of similarity between graphons, measuring how close two graphons are up to relabeling of the underlying vertex space. It is based on the cut norm, which generalizes the notion of edge density across subsets of a graph.

**Definition 3.5.** The *cut norm* of a measurable $W : [0, 1]^2 \to \mathbb{R}$ is defined as

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W \right|,$$

where the supremum is taken over all measurable subsets $S, T \subseteq [0, 1]$.

To define the cut distance in a meaningful, label-invariant way, we need to formalize what it means to relabel a graphon. This is done using measure-preserving maps, which intuitively represent relabelings that preserve the uniform distribution on the unit interval.

**Definition 3.6.** We say that $\phi : [0, 1] \to [0, 1]$ is a *measure preserving map* if

$$\lambda(A) = \lambda(\phi^{-1}(A)) \qquad \forall \text{ measurable A } \subseteq [0, 1].$$

We say that $\phi$ is an *invertible* measure preserving map if there is another measure preserving map $\psi : [0, 1] \to [0, 1]$ such that $\phi \circ \psi$ and $\psi \circ \phi$ are both identity maps outside sets of measure zero.

Given $W : [0, 1]^2 \to \mathbb{R}$ and an invertible measure preserving map $\phi : [0, 1] \to [0, 1]$, we write

$$W^\phi(x, y) = W(\phi(x), \phi(y)).$$

This can be seen as a relabeling of the underlying vertex space, analogous to permuting the vertices of a finite graph. With this in place, we define the cut distance between two graphons as the smallest possible cut norm distance over all such relabelings.

**Definition 3.7.** Given two symmetric measurable functions $U, W : [0, 1]^2 \to \mathbb{R}$, we define the *cut distance* to be

$$\begin{aligned}
\delta_\square(U, W) &= \inf_\phi \|U - W\|_\square \\
&= \inf_\phi \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} (U(x, y) - W(\phi(x), \phi(y)))dxdy \right|,
\end{aligned}$$

where the infimum is taken over all invertible measure preserving maps $\phi : [0,1] \to [0,1]$. We define the cut distance between two graphs $G$ and $G'$ by the cut distance of their associated graphons, i.e., $\delta_\square(G, G') = \delta_\square(W_G, W_{G'})$.

**Definition 3.8.** We say that a sequence of graphs or graphons *converges in cut metric* if they form a Cauchy sequence with respect to $\delta_\square$. Furthermore, we say $W_n$ *converges to* $W$ *in cut metric* if $\delta_\square(W_n, W) \to 0$ as $n \to \infty$.

It is important to note that $\delta_\square$ is not a true metric on the space of all graphons, because two graphons that differ only by relabeling have zero distance. To fix this, we identify graphons that are equivalent in this sense.

**Definition 3.9.** Let $\widetilde{\mathcal{W}_0}$ be the set of graphons where any pair of graphons with cut distance zero are considered the same point in the space. This is a metric space under cut distance, $\delta_\square$.

The following foundational results demonstrate the power and completeness of the graphon formalism. The proofs are beyond the scope of this paper.

**Theorem 3.10.** *The set of graphs is dense in* $(\widetilde{\mathcal{W}_0}, \delta_\square)$.

**Theorem 3.11.** *The graphon space* $(\widetilde{\mathcal{W}_0}, \delta_\square)$ *is compact.*

**Theorem 3.12.** *The graphon space* $(\widetilde{\mathcal{W}_0}, \delta_\square)$ *is the completion of the space of graphs with respect to the cut metric.*

Together, these results tell us that every sequence of large graphs, provided it doesn't escape to infinity, has a limiting object in the form of a graphon. Moreover, this limit is well-defined in the cut metric, and small-scale statistics (like homomorphism densities) behave continuously with respect to this metric. This is precisely what makes graphons such a powerful and elegant tool in modern combinatorics.

### 3.3 Homomorphism Densities in Graphons

We now revisit the notion of homomorphism densities in the broader context of graphons. Earlier, we defined the homomorphism density $t(F, G)$ to be the probability that a random map from $V(F)$ to $V(G)$ preserves all adjacencies. In the graphon setting, the vertex set of the host graph becomes the unit interval $[0,1]$, and instead of a finite summation over vertex maps, we take an integral over all possible assignments of the vertices of $F$ into $[0,1]$.

**Definition 3.13.** Let $F$ be a graph and $W$ a graphon. The $F$*-density in* $W$ is defined to be

$$t(F, W) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i \in V(F)} dx_i.$$

This definition is a natural extension of the discrete case. In fact, when $W = W_G$ is the graphon associated to a finite graph $G$, this integral computes the exact same quantity as $t(F, G)$, since the graphon $W_G$ takes only finitely many values and is constant on blocks of a partition of $[0,1]$. Importantly, $t(F, W)$ can be interpreted as the limiting relative frequency of observing a copy of $F$ inside some large, random graph modeled by $W$. This perspective leads us to an important type of convergence for sequences of graphons.

**Definition 3.14.** We say that a sequence of graphons $W_n$ is *left-convergent* if for every graph $F$, $t(F, W_n)$ converges as $n \to \infty$. We say that this sequence left-converges to a graphon $W$ if $\lim_{n\to\infty} t(F, W_n) = t(F, W)$.

The terminology "left-convergent" comes from the perspective that the test graphs $F$ are placed on the left of the homomorphism $\hom(\mathrm{F}, \mathrm{G})$. That is, we're looking at how well various small graphs $F$ embed into the sequence $W_n$, and whether the pattern statistics they produce are converging. This raises the question, does this notion of convergence match our earlier, more analytic notion of convergence in the cut distance?

Amazingly, the answer is yes.

**Theorem 3.15.** *A sequence of graphons is left-convergent if and only if it is a Cauchy sequence with respect to the cut metric.*

### 3.4 Applications

Now that we have developed the formalism of graphons and convergence in the cut metric, here are a number of examples that illustrate how graphons can be used to analyze large graphs and reason about structural properties such as randomness and symmetry.

**Example 3.16.** Let $G(n, p)$ denote the Erdős–Rényi random graph on $n$ vertices where each edge is included independently with probability $p \in (0, 1)$. As $n \to \infty$, the sequence $(G(n, p))_{n\geq 1}$ converges in the cut metric to the constant graphon $W(x, y) = p$. This example shows that random graphs with a fixed edge probability can be modeled in the limit by a deterministic, constant graphon. The randomness is encoded in the uniform distribution over the unit interval, while the edge structure is encoded in the constant value of $W$.

**Example 3.17.** Let $G_n = K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ be the complete bipartite graph on $n$ vertices. The graphon limit of this sequence is the step-function graphon:

$$W(x, y) = \begin{cases} 1 & \text{if } x \leq 1/2 < y \text{ or } y \leq 1/2 < x, \\ 0 & \text{otherwise.} \end{cases}$$

This limit graphon captures the bipartite structure of $G_n$ in the form of a block matrix. In general, structured deterministic graphs correspond to simple, piecewise-constant graphons.

**Example 3.18.** A sequence of graphs $(G_n)$ is said to be *quasirandom* if it behaves like $G(n, p)$ in terms of edge distribution and subgraph densities. One characterization of quasirandomness is that the sequence converges in the cut metric to the constant graphon $W(x, y) = p$ and satisfies

$$t(C_4, G_n) \to p^4.$$

# 4 Extremal Graph Theory and Entropy Revisited

Finally, we arrive at my favorite section, which is what I am currently still exploring and having lots of fun with. Having developed the machinery of graphons and their applications, we can now return to extremal graph theory with a more powerful set of tools. One of the major advantages of the graphon framework is that it allows us to recast extremal problems as variational problems over a compact space.

## 4.1 Sidorenko's Conjecture

Sidorenko's conjecture, one of the central open problems in extremal graph theory, becomes especially elegant when expressed in terms of graphons.

**Proposition 4.1.** *Let $F$ be a bipartite graph. Then for every graphon $W$, we have:*

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

This inequality trivially holds for constant graphons (i.e., the Erdős–Rényi limit).
We say that a graph $F$ is Sidorenko if the condition in the conjecture holds, i.e., $t(F, W) \geq t(K_2, W)^{|E(F)|}$. I end this paper by proving that several basic graphs are Sidorenko using the techniques discussed throughout the paper.

**Theorem 4.2.** *$K_{2,2}$ is Sidorenko, i.e., $t(K_{2,2}, W) \geq t(K_2, W)^4$.*

*Proof.* Claim. $t(K_{1,2}, W) \geq t(K_2, W)^2$.

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

Claim. $t(K_{2,2}, W) \geq t(K_{1,2}, W)^2$.

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

Putting the two claims together gives us our result. □

**Theorem 4.3.** *The 3-edge path is Sidorenko.*

*Proof 1.* Let $P_4$ be the 3-edge path, let $W$ be a graphon, and let $g(x) = \int_y W(x,y)$ represent 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 relabling, we can also write

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

Then,

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

□

Excitingly, this theorem can also be proven using entropy methods.

*Proof 2.* Let $P_4$ denote the 3-edge path, and let $G$ be a graph. An element of $\mathrm{Hom}(P_4, G)$ is a walk of length three. We randomly choose walk $XYZW$ in $G$ as follows.

(i) $XY$ is a uniformly random edge of $G$ (i.e., choose an edge of $G$ uniformly at random, then choose one of its endpoints uniformly to be $X$, and the other to be $Y$);
(ii) $Z$ is a uniform random neighbor of $Y$;
(iii) $W$ is a uniform random neighbor of $Z$.

A key observation is that $YZ$ is also distributed as a uniform random edge of $G$. Indeed, conditioned on the choice of $Y$, the vertices $X$ and $Z$ are independent and uniformly distributed neighbors of $Y$. Hence $XY$ and $YZ$ are identically distributed, so $YZ$ is a uniform random edge. Similarly, $ZW$ is also a uniformly random edge.

Since $X$ and $Z$ are conditionally independent given $Y$, we have

$$H(Z \mid X, Y) = H(Z \mid Y) \quad \text{and similarly} \quad H(W \mid X, Y, Z) = H(W \mid Z).$$

Furthermore,

$$H(Y \mid X) = H(Z \mid Y) = H(W \mid Z),$$

since $XY, YZ, ZW$ are all identically distributed as uniform random edges. Now, we compute the joint entropy,

$$\begin{aligned} H(X,Y,Z,W) &= H(X) + H(Y \mid X) + H(Z \mid X,Y) + H(W \mid X,Y,Z) \\ &= H(X) + H(Y \mid X) + H(Z \mid Y) + H(W \mid Z) \\ &= H(X) + 3H(Y \mid X) \\ &= 3H(X,Y) - 2H(X) \\ &= 3\log(2|E(G)|) - 2H(X) \\ &\geq 3\log(2|E(G)|) - 2\log|V(G)|. \end{aligned}$$

This proves the entropy lower bound:

$$\log \hom(P_4, G) \geq H(X,Y,Z,W) \geq 3\log(2|E(G)|) - 2\log|V(G)|.$$

Exponentiating both sides and normalizing,

$$\begin{aligned} t(P_4, G) &= \frac{\hom(P_4, G)}{|V(G)|^4} \\ &\geq \left(\frac{2|E(G)|}{|V(G)|^2}\right)^3 = t(K_2, G)^3. \end{aligned}$$ □

# 5 Conclusion

I hope that my reflection of my journey through the many topics I explored this quarter has been somewhat interesting to read. I think that what has stayed with me most throughout this exploration is not any particular result, but the almost cathartic feeling that emerges when seemingly disparate ideas begin to align. This process of discovering unexpected connections has been the most rewarding part of my experience so far, and though I do not yet know where this path will lead, I remain eager to continue following it.

# References


[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 2016. ISBN: 978-1-119-06195-3.

[Juk11] Stasys Jukna. *Extremal Combinatorics: With Applications in Computer Science*. Springer, 2011. ISBN: 978-3-642-17363-9. DOI: `https://doi.org/10.1007/978-3-642-17364-6`.

[Lov12] László Lovász. *Large Networks and Graph Limits*. Vol. 60. Colloquium Publications. American Mathematical Society, 2012. ISBN: 978-0-8218-9085-1.

[LS06] László Lovász and Balázs Szegedy. "Limits of dense graph sequences". In: *Journal of Combinatorial Theory, Series B* 96.6 (2006), pp. 933–957. ISSN: 0095-8956. DOI: `https://doi.org/10.1016/j.jctb.2006.05.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0095895606000517`.

[Zha23] Yufei Zhao. *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness*. Cambridge University Press, 2023. ISBN: 978-1009310949.