Coupling sparse models and dense extremal problems

by

James Hirst

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author . . . .
Department of Mathematics
July 30, 2020

Certified by . . . .
Henry Cohn
Adjunct Professor
Thesis Supervisor

Accepted by . . . .
Jonathan Kelner
Graduate Chair, Applied Math

Coupling sparse models and dense extremal problems

by

James Hirst

Submitted to the Department of Mathematics
on July 30, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We study the problem of coupling a stochastic block model with a planted bisection to a uniform random graph having the same average degree. Focusing on the regime where the average degree is a constant relative to the number of vertices n, we show that the distance to which the models can be coupled undergoes a phase transition from $O(\sqrt{n})$ to $\Omega(n)$ as the planted bisection in the block model varies. This settles half of a conjecture of Bollobás and Riordan and has some implications for sparse graph limit theory. In particular, for certain ranges of parameters, a block model and the corresponding uniform model produce samples which must converge to the same limit point. This implies that any notion of convergence for sequences of graphs with $\Theta(n)$ edges which allows for samples from a limit object to converge back to the limit itself must identify these models.

On the other hand, we demonstrate that the existing theory of dense graph limits is a powerful tool for dealing with extremal problems on graphs with $\Theta(n^2)$ edges. The language of graphons along with the flag algebra method allow us to obtain many results which would otherwise be out of reach or at least difficult to manage. We study graph profiles which capture correlations between different graphs in a larger network. Further, we give proofs in the flag algebra of some inducibility-like problems which have gained some particular interest recently.

Thesis Supervisor: Henry Cohn
Title: Adjunct Professor

Acknowledgments

My thanks go out to NSERC who partially funded my research from 2015 to 2018. I also would like to thank my advisor Henry Cohn and all the math department staff and faculty I've had the pleasure of working with over the years.

Contents

Chapter 1

Block models and sparse graph metrics

1.1 Introduction

The idea of putting a metric on the space of graphs has generated a significant amount of fruitful research in the last decade. In particular, this has led to an extremely rich theory of dense graph limits, a line of work stemming from [8] which remains active today. Here, the word *dense* refers to the scaling in the number of edges: a dense sequence of graphs with n vertices should have $\Theta(n^2)$ edges, and any set of $o(n^2)$ edges does not contribute to the limit of the sequence (they become measure 0 sets in the limit). The metric that best captures similarity between dense graphs is the *cut distance*, which, informally, is a measure of how well one can align two graphs to have a small number of edges across every cut. Closeness in cut distance turns out to be equivalent to a number of natural metrics including subgraph counts and free energies. For an excellent exposition of these results, we refer the reader to [19].

Modulo some technical conditions, a suitably re-normalized version of the cut distance is also a valid and interesting metric for graphs with $o(n^2)$ edges. In [4], the authors show that for graph sequences in which every $o(n)$ vertices of a graph span $o(n)$ edges (forbidding graphs like a clique on $o(n)$ vertices), many of the nice

Chapter 1 up to section 1.5 is from the author's paper [17].

properties of the cut distance continue to hold. The more recent work of [7] uses a relaxation of this condition based on boundedness in an ℓ_p sense (ordinary graphs can be thought of as bounded in ℓ_∞). However, in both of these results is an implicit requirement that graphs have unbounded degrees as n grows, i.e., $\omega(n)$ edges.

An attempt at treating the $\Theta(n)$ sparsity regime is given in [5]. One of the main difficulties here is that the random graph $G(n, c/n)$ concentrates much more weakly than usual. Indeed, the degree of a vertex has mean and variance of the same order c, and many key quantities have tails which decay only exponentially in n, rather than super-exponentially. As such, basic results which follow from a union bound for denser graphs break down. In this setting, the cut distance turns out to be much too strong a metric: even a sequence of i.i.d. graphs from $G(n, c/n)$ fails to be a Cauchy sequence. On the other hand, local metrics based on subgraph counts are in some sense too weak, as they fail to distinguish between a random graph and a random bipartite graph of the same average degree.

The most basic object in dense graph limit theory is the graphon, a symmetric and measurable function $W : [0,1]^2 \to [0,1]$. Given a graphon W, there is a natural way to sample an n-vertex graph G_n: choose $x_1, \ldots, x_n$ uniformly from $[0,1]$ and independently place each edge (i,j) with probability $W(x_i, x_j)$. This produces a graph with $\Theta(n^2)$ edges (assuming $W \neq 0$), and such models are well studied. For instance a basic result from graph limit theory says that the sequence $\{G_n\}$ converges to W (under the cut distance) almost surely.

With a slight modification, we can use this construction to sample graphs with $\Theta(n)$ edges. To avoid confusion, we will call a symmetric and measurable function $\kappa : [0,1]^2 \to \mathbb{R}$ a *kernel* to distinguish it from a graphon which typically maps to $[0,1]$. Given a bounded kernel $\kappa : [0,1]^2 \to \mathbb{R}$, again take $x_1, \ldots, x_n$ as i.i.d. $U(0,1)$ but now place edge (i,j) with probability $\kappa(x_i, x_j)/n$ (projecting this value into $[0,1]$ as required). Denote this random graph model by $G(n, \kappa)$.

Can we still say in some sense that the (sparsely) sampled sequence $\{G_n\}$ converges to κ? We will see momentarily that there is a natural barrier to this which is not present in the dense case. Given n-vertex graphs G and H on the same vertex set, we define the *edit distance* between G and H to be the minimum number of edges which need to be changed to turn H into an isomorphic copy of G, and we write

$$\mathrm{d}_1(G,H) = \min_{G'\cong G} |E(G')\triangle E(H)|\,.$$

As in the dense case, it is reasonable to assume that any notion of convergence for sparse graphs should be invariant under changes to $o(n)$ edges. Following [5], we will call the models $G(n,\kappa_1)$ and $G(n,\kappa_2)$ (or sometimes just κ_1 and κ_2) *essentially equivalent* if there is a coupling μ between $G(n,\kappa_1)$ and $G(n,\kappa_2)$ such that $\mathbf{E}_\mu \mathrm{d}_1(G,H) = o(n)$, where $(G,H)\sim\mu$. We will call them *essentially different* otherwise. Returning to our question above, if we can find $\kappa_1 \neq \kappa_2$ which are essentially equivalent, then samples from $G(n,\kappa_1)$ and $G(n,\kappa_2)$ must converge in some sense to both κ_1 and κ_2.

Consider the following example. Let κ_1 be identically c, and take κ_2 to have value $c+\delta$ on $[0,1/2]\times[1/2,1]$ (symmetrically) and $c-\delta$ elsewhere. Henceforth we will set $\mathbf{Q} = G(n,\kappa_1)$ and $\mathbf{P} = G(n,\kappa_2)$, suppressing the dependence on n, c, and δ wherever it avoids confusion. In [5], the authors give a simple example to show that when $c \geq \delta$ and $0 \leq \delta \leq 1$, $\mathbf{P}$ and $\mathbf{Q}$ are in fact essentially equivalent. The following extension of this is half of Conjecture 6.3 from [5].

Conjecture 1.1.1. *Let $0 \leq \delta \leq c$. If $\delta < \sqrt{c}$, then $\mathbf{P}$ and $\mathbf{Q}$ are essentially equivalent.*

Note that here the underlying kernels are truly different: One cannot compose κ_1 and κ_2 with (potentially different) measure preserving maps to make them equal almost everywhere. This makes the idea that $\mathbf{P}$ and $\mathbf{Q}$ can be essentially equivalent somewhat surprising at first glance.

Our main result is a proof of Conjecture 1.1.1. In fact, we prove an even stronger result: when $\delta < \sqrt{c}$, the two models can be coupled to an expected edit distance of $O(\sqrt{n})$.

Theorem 1.1.2. *Let $0 \leq \delta \leq c$ and denote by $\Pi(\mathbf{P}, \mathbf{Q})$ the space of coupling measures between $\mathbf{P}$ and $\mathbf{Q}$. If $\delta < \sqrt{c}$, then there exists a coupling $\mu \in \Pi(\mathbf{P}_n, \mathbf{Q}_n)$ such that*

$$\mathbf{E}_\mu \left[\mathrm{d}_1(G, H) \right] = O(\sqrt{n}).$$

We also give a lower bound on the expected edit distance under any coupling.

Theorem 1.1.3. *Let $1 < \delta \leq c$. For any coupling $\mu \in \Pi(\mathbf{P}, \mathbf{Q})$,*

$$\mathbf{E}_\mu \left[\mathrm{d}_1(G, H) \right] = \Omega \left(\frac{\delta^{\log(n)^{1/3}}}{\log(n)^{1/3}} \right).$$

1.2 Preliminaries

1.2.1 Optimal transport

The proof strategy for both Theorem 1.1.2 and Theorem 1.1.3 is to obtain bounds on the quantity

$$\inf_\mu \mathbf{E}_\mu \left[\mathrm{d}_1(G, H) \right], \tag{1.1}$$

where the infimum is taken over $\mu \in \Pi(\mathbf{P}, \mathbf{Q})$. This is actually a well known type of problem called an optimal transport problem (for the cost d_1) and as such we will make use of a number of tools from the area. For an excellent survey of optimal transport, see [28]. The main item we will make use of is the following duality formula.

Proposition 1.2.1 (Monge-Kantorovich Duality)**.** *We have the equality*

$$\inf_\mu \mathbf{E}_\mu \left[\mathrm{d}_1(G, H) \right] = \sup_f \left| \mathbf{E}_P f - \mathbf{E}_Q f \right|, \tag{1.2}$$

where $\mu \in \Pi(\mathbf{P}, \mathbf{Q})$, $(G, H) \sim \mu$, and f is 1-Lipschitz in the edit distance.

This is in fact a special case of Monge-Kantorovich duality for cost functions which satisfy the triangle inequality. Note that Proposition 1.2.1 immediately implies the useful inequality

$$\inf_\mu \mathbf{E}_\mu \left[\mathrm{d}_1(G, H) \right] \geq \sup_f \left| \mathbf{E}_P f - \mathbf{E}_Q f \right|,$$

which is often referred to as *weak duality.*

It is not hard to see that the infimum in (1.2) is always attained: the space $\Pi(\mathbf{P}, \mathbf{Q})$ is compact for the weak convergence in duality with the space of continuous functions. Hence, by the definition of weak convergence, the functional sending μ to $\mathbf{E}_\mu[d_1(G, H)]$ is continuous. This is enough to guarantee the existence of a minimizer μ^* (Weierstrass). As it turns out, the supremum is also attained, but the proof is somewhat more involved (see Section 1.6 in [28]).

Proposition 1.2.2. *There exists a coupling $\mu^* \in \Pi(\mathbf{P}, \mathbf{Q})$ and a 1-Lipschitz function f^* such that*

$$\mathbf{E}_{\mu^*}[\mathrm{d}_1(G^*, H^*)] = |\mathbf{E}_P f^* - \mathbf{E}_Q f^*|,$$

where $(G^*, H^*) \sim \mu^*$.

1.2.2 Stochastic block models

The study of sparse random graphs has also seen some attention from a different research field, motivated by the problem of detecting hidden communities planted in a random model. The stochastic block model for two balanced communities is usually defined as follows. Take parameters $a > b > 0$ (for a ferromagnetic block model) and assign to each vertex i of an n vertex graph a random spin $\sigma_i \in \{-1, 1\}$. Then, connect edges i and j with probability a/n if $\sigma_i\sigma_j = 1$, and with probability b/n otherwise, independently for every edge. Notice that samples from the stochastic block model are identical to samples from $\mathbf{P}$, with $a = c + \delta$ and $b = c - \delta$.

Much of the previous work on such models is algorithmic, but some recent theoretical results have been obtained which are extremely relevant to the problem at hand. Of particular interest is [21], in which the authors settle a conjecture from [10] and give an information-theoretic lower bound for the detection problem. This lower bound comes in the form of a contiguity result.

Theorem 1.2.3 (Theorem 2.4 from [21]). *When $\delta < \sqrt{c}$, the models $\mathbf{P}_n$ and $\mathbf{Q}_n$ are mutually contiguous, i.e., for every sequence of events A_n we have*

$$\mathbf{P}_n(A_n) \to 0 \iff \mathbf{Q}_n(A_n) \to 0.$$

To be precise, contiguity is a notion that should be applied to sequences of probability measures, and so when we say the models $\mathbf{P}_n$ and $\mathbf{Q}_n$ are mutually contiguous, we really mean that the sequences $\{\mathbf{P}_n\}$ and $\{\mathbf{Q}_n\}$ are.

An easy but important corollary which shall be our main point of contact with Theorem 1.2.3 is that it is impossible to consistently distinguish between samples from $\mathbf{P}$ and $\mathbf{Q}$ below the contiguity threshold.

Corollary 1.2.4. *Generate a random pair (σ, G), where $\sigma \in \{0,1\}$ and G is an n-vertex graph as follows. Take $\sigma \sim \mathrm{Ber}(1/2)$ and sample G from $\mathbf{P}$ if $\sigma = 1$, sampling G from $\mathbf{Q}$ otherwise. When $\delta < \sqrt{c}$, there is no estimator π taking an n-vertex graph to $\{0,1\}$ such that*

$$\Pr[\pi(G) = \sigma] = 1 - o(1). \tag{1.3}$$

Proof. Given such a π, take $A_n = \{G : \pi(G) = 1\}$. Then $\mathbf{P}_n(A_n) \to 1$ while $\mathbf{Q}_n(A_n) \to 0$. □

We often call an estimator π satisfying (1.3) an estimator for the block model or simply an estimator for $\mathbf{P}$. It is known that such estimators exist for all $\delta > \sqrt{c}$ and can even be efficiently computed (see [21] and [22]).

1.3 Proofs

At a high level, the proof strategy for Theorem 1.1.2 is to use Monge-Kantorovich duality along with an upper bound for

$$\sup_f |\mathbf{E}_P f - \mathbf{E}_Q f|.$$

If $|\mathbf{E}_P f - \mathbf{E}_Q f|$ is too large for some f, then f can be used as an estimator for $\mathbf{P}$ by thresholding. In order to make this work, we need to know that f is sufficiently concentrated around its mean.

One can try to apply Azuma-Hoeffding using the vertex exposure martingale, but the martingale differences can only be bounded by the maximum degree of the graph which is a bit less than $\log(n)$ under both $\mathbf{P}$ and $\mathbf{Q}$. This gives a tail bound of the form

$$\Pr\left[|f - \mathbf{E}f| > t\right] < 2\exp\left(-\frac{2t^2}{n\log(n)^2}\right)$$

at best, losing a $\log(n)^2$ factor in the variance. Rather than looking for tails with the correct variance, we will instead use the following inequality which bounds the variance directly.

Proposition 1.3.1 (Efron-Stein [30])**.** *Let $X_1, \ldots, X_n$ be independent random variables and set $X = (X_1, \ldots, X_n)$. Denote by $X^{(i)}$ the random variable obtained from X by re-sampling X_i. Then*

$$\mathrm{Var}(f(X)) \leq \frac{1}{2}\sum_{i=1}^{n} \mathbf{E}\left[\left(f(X) - f(X^{(i)})\right)^2\right].$$

With Efron-Stein, we can prove concentration for Lipschitz functions for samples from any bounded kernel.

Lemma 1.3.2. *Let κ be a bounded kernel with $d = \sup \kappa$, take $G_n \sim G(n, \kappa)$, and let $f : \mathcal{G}_n \to \mathbb{R}$ be 1-Lipschitz in the edit distance. Then*

$$\mathrm{Var}(f(G_n)) \leq dn.$$

Proof. The proof is a direct application of Efron-Stein. With $m = \binom{n}{2}$, $f(G_n)$ is a function of $n + m$ independent $U(0,1)$ random variables, namely, n type (or vertex) variables and m edge variables. Notice that for a type variable i, $f(X) - f(X^{(i)})$ is stochastically dominated by a $\mathrm{Bin}(n, d/n)$ random variable since f is Lipschitz. So we can bound the marginal variance as

$$\mathbf{E}\left[\left(f(X)-f(X^{(i)})\right)^2\right] \leq \mathrm{Var}(\mathrm{Bin}(n,d/n)) = d\left(1-\frac{d}{n}\right) \leq d.$$

Similarly for an edge variable j, $f(X)-f(X^{(j)})$ is stochastically dominated by a $\mathrm{Ber}(d/n)$ random variable which gives

$$\mathbf{E}\left[\left(f(X)-f(X^{(j)})\right)^2\right] \leq \mathrm{Var}(\mathrm{Ber}(d/n)) = \frac{d}{n}\left(1-\frac{d}{n}\right) \leq \frac{d}{n}.$$

Putting these together yields

$$\mathrm{Var}(f(G_n)) \leq \frac{1}{2}\left(dn+\frac{d}{n}m\right) \leq dn. \qquad \square$$

We are now ready to prove Theorem 1.1.2.

Proof of Theorem 1.1.2. Suppose that in (1.2) we have

$$\sup_f |\mathbf{E}_{\mathbf{P}}f - \mathbf{E}_{\mathbf{Q}}f| = \omega(\sqrt{n}).$$

By Proposition 1.2.2 there exists a 1-Lipschitz function f such that

$$|\mathbf{E}_{\mathbf{P}}f - \mathbf{E}_{\mathbf{Q}}f| = \omega(\sqrt{n}) = \alpha(n)\sqrt{n},$$

where $\alpha(n)$ is a function with $\lim_{n\to\infty}\alpha(n)=\infty$. Using Lemma 1.3.2 and Chebyshev's inequality we have

$$\Pr\left[|f-\mathbf{E}f| > \sqrt{\alpha(n)}\sqrt{(c+\delta)n}\right] \leq \frac{1}{\alpha(n)} \to 0 \tag{1.4}$$

for both $\mathbf{P}$ and $\mathbf{Q}$. Now, one of the terms $\mathbf{E}_{\mathbf{P}}f$ and $\mathbf{E}_{\mathbf{Q}}f$ must be at least the difference $\alpha(n)\sqrt{n}$. For every n, define the estimator π as follows. If $E_{\mathbf{P}}f > \mathbf{E}_{\mathbf{Q}}f$ then let

$$\pi(G) = \begin{cases} 1 & \text{if } f(G) \geq \mathbf{E}_{\mathbf{P}}f - \frac{\alpha(n)\sqrt{n}}{2} \\ 0 & \text{otherwise} \end{cases}.$$

Similarly if $\mathbf{E_Q} f > \mathbf{E_P} f$ let

$$\pi(G) = \begin{cases} 0 & \text{if } f(G) \geq \mathbf{E_Q} f - \frac{\alpha(n)\sqrt{n}}{2} \\ 1 & \text{otherwise} \end{cases}.$$

It follows from (1.4) that π is an estimator for $\mathbf{P}$ which is impossible by Corollary 1.2.4.

Returning to (1.2) we get the upper bound

$$\inf_{\mu} \mathbf{E}_{\mu} \left[\mathrm{d}_1(G, H) \right] = O(\sqrt{n}).$$

Again referring to Proposition 1.2.2, there exists a coupling $\mu \in \Pi(\mathbf{P}, \mathbf{Q})$ such that

$$\mathbf{E}_{\mu} \left[\mathrm{d}_1(G, H) \right] = O(\sqrt{n}).$$ □

We should note that it is important to make the distinction between $\mathbf{E_P} f > \mathbf{E_Q} f$ and $\mathbf{E_Q} f > \mathbf{E_P} f$ for every n, as even though f concentrates around its mean for a fixed n, the values of $\mathbf{E_P} f$ and $\mathbf{E_Q} f$ might not be converging as n grows.

Let us now turn to the proof of Theorem 1.1.3.

Proof of Theorem 1.1.3. From weak duality, for any 1-Lipschitz function f we have the inequality

$$\inf_{\mu} \mathbf{E}_{\mu} \left[\mathrm{d}_1(G, H) \right] \geq \left| \mathbf{E_P} f - \mathbf{E_Q} f \right|.$$

Given a graph G and a parameter $k \in \mathbb{N}$, define Y_k to be the maximum number of k-cycles in a disjoint cycle packing of G. Set $f(G) = Y_k$, noting that f is certainly 1-Lipschitz, as an edge can contribute to a single cycle in any disjoint packing.

It remains to control $\mathbf{E_P} f$ and $\mathbf{E_Q} f$. To this end, define X_k to be simply the number of (potentially overlapping) k-cycles in G, counted as ordered sets of k vertices, up to automorphism. Here, we have

$$\mathbf{E_Q} X_k = \frac{n(n-1)\cdots(n-k+1)}{2k} \left(\frac{c}{n} \right)^k = \frac{c^k}{2k} + o(1),$$

where the second equality holds for $k = o(\sqrt{n})$. Indeed, there are $\sim n^k/2k$ ways to choose the vertices along with an ordering for the cycle modulo automorphism, and a probability $(c/n)^k$ of the edges appearing in sequence. Extending this result to $\mathbf{P}$, it is shown in [21] (Section 3) that

$$\mathbf{E}_{\mathbf{P}} X_k = \frac{c^k + \delta^k}{2k} + o(1),$$

again for $k = o(\sqrt{n})$.

Consider the set $\mathcal{H}$ of all graphs H which can be obtained as the union of two distinct, non-disjoint k-cycles. Note that every graph $H \in \mathcal{H}$ has $m \leq 2k$ vertices and at least $m+1$ edges, since it cannot be a cycle itself. For every such H, define X_H to be the number of injective homomorphisms from H into G . We can easily compute

$$\mathbf{E}_{\mathbf{Q}} X_H \leq n^m \left(\frac{c}{n}\right)^{m+1} \leq \frac{c^{2k+1}}{n}.$$

Define Z_k by

$$Z_k = \sum_{H \in \mathcal{H}} X_H.$$

Relating X_k and Y_k through Z_k, we have the inequalities

$$X_k \geq Y_k \geq X_k - (2k)^k Z_k,$$

since every embedding of a graph $H \in \mathcal{H}$ into G contributes at most $(2k)^k$ k-cycles to X_k, and removing all such edges yields a graph with $X_k = Y_k$. A graph with $2k$ vertices has at most $(2k)^2/2 = 2k^2$ edges so, counting over all possible labelled graphs, we get $|\mathcal{H}| \leq 2^{2k^2}$, where graphs on $< 2k$ vertices are identified with graphs on $2k$ vertices with isolated vertices. This gives

$$\mathbf{E}_{\mathbf{Q}} Z_k \leq \frac{2^{2k^2} c^{2k+1}}{n}.$$

Taking $k = \log(n)^{1/3}$ (in fact, any $k = \log(n)^{\alpha}$, $\alpha < 1/2$ will work), the numerator is $o(n)$ and so $\mathbf{E}_{\mathbf{Q}} Z_k = o(1)$. Since the number of copies of any graph in $\mathbf{P}$ is stochastically dominated by the number of copies in a $G(n,(c+\delta)/n)$, $\mathbf{E_P}Z_k = o(1)$ as well. Putting everything together, we end up with

$$|\mathbf{E_P}f - \mathbf{E_Q}f| \geq \frac{\delta^k}{2k} - o(1) = \Omega\left(\frac{\delta^{\log(n)^{1/3}}}{\log(n)^{1/3}}\right). \qquad \square$$

As noted, taking $k = \log(n)^\alpha$ for any $\alpha < 1/2$ improves the bound in Theorem 1.1.3 to

$$\frac{\delta^{\log(n)^\alpha}}{\log(n)^\alpha},$$

but the value of this function remains somewhere between polylog(n) and poly(n) for all $\alpha < 1/2$ and $\delta > 1$ and so this is not a particularly meaningful improvement.

1.4 Coupling remarks

As mentioned, Conjecture 1.1.1 is only half of the conjecture made in [5]. In addition, the authors conjecture that for $\delta > \sqrt{c}$ and $c > 1$ (to ensure a giant component), the models $\mathbf{P}$ and $\mathbf{Q}$ are essentially different. This means that it should be impossible to couple $\mathbf{P}$ and $\mathbf{Q}$ to an expected edit distance of $o(n)$. Now if we allow $\delta > A\sqrt{c}$ for some sufficiently large constant A, then this is certainly the case. This is because the minimum bisection in a graph $G \sim \mathbf{Q}$ is of the order $cn/4 - \Theta(\sqrt{c}n)$ with high probability, whereas $G \sim \mathbf{P}$ almost always has a bisection with $cn/4 - \Theta(\delta n/4)$ edges. Plugging the minimum bisection into (1.2) gives the desired $\Theta(n)$ lower bound.

It is not clear at all that we should be able to take $A = 1$ in the above. One might hope that for $\delta > \sqrt{c}$, the minimum bisections of $\mathbf{P}$ and $\mathbf{Q}$ would differ by $\Theta(n)$. One of the main factors mitigating progress in this area is the complexity of even computing the minimum bisection of a graph $G \sim \mathbf{Q}$. Writing $\mathrm{m}(G)$ for the value of the minimum bisection of G, the authors of [11] recently determined the second order growth of $m(G)$ through the formula

$$\mathbf{E_Q}\left[\mathrm{m}(G)\right] = n\left(\frac{c}{4} - \frac{\mathrm{P}^*}{2}\sqrt{c} + o(\sqrt{c})\right) + o(n). \tag{1.5}$$

The constant $\mathrm{P}^* \approx 0.7632$ here is the ground state energy of the SK spin glass model, and has no known simple closed form. A similar formula for $\mathbf{E_P}[\mathrm{m}(G)]$ is given in [29] with an even more mysterious constant C^*. It would be a significant breakthrough to pinpoint the value of δ for which the minimum bisections in $\mathbf{P}$ and $\mathbf{Q}$ transition from having a difference $O(\sqrt{n})$ to $\Theta(n)$. It could even be that two separate phase transitions occur.

The work of [18] uses some strong heuristic methods from statistical physics to make predictions in this direction. Although non-rigorous, these heuristics (including the so-called cavity method) have predicted a number of results which have since been proven rigorously (like Theorem 1.2.3 above). They study an SDP relaxation of minimum bisection and predict that it detects communities down to $\delta = \sqrt{c}(1 + (8c)^{-1} + O(c^{-2}))$. This lower bound is at least partially vindicated by a matching upper bound from [20], who show that the SDP based estimator in fact does detect communities at $\delta = \sqrt{c} + \varepsilon$ for any ε, provided $c = c(\varepsilon)$ can be taken sufficiently large. It seems as though minimum bisection should not do any worse than the SDP relaxation, which would allow us to take $A = 1 + o_c(1)$ above, but this is not a proof by any means.

Ignoring the question of bisections, it may well be that the models $\mathbf{P}$ and $\mathbf{Q}$ do in fact become essentially different at the contiguity threshold $\delta = \sqrt{c}$. Our choice of f in the proof of Theorem 1.1.3 was based on the simple fact that $\mathbf{P}$ is biased towards having more k-cycles than $\mathbf{Q}$ on average. Taking k larger gives a better lower bound, but one cannot take k past $o(\sqrt{\log(n)})$ without overlapping cycles contributing to the expectation. To normalize f into a Lipschitz function would then require an analysis of this overlap and seems difficult. A radically different choice of f may be able to give improved lower bounds. For $\delta < \sqrt{c}$, it seems as though $\Theta(\sqrt{n})$ should be the correct answer: If not, then the minimum bisections in $\mathbf{P}$ and $\mathbf{Q}$ agree to within $o(\sqrt{n})$, which would be somewhat surprising.

As a final remark, we should note that the notion of being essentially different is robust under $o(n)$ adversarial edits to our samples. If $\mathbf{P}$ and $\mathbf{Q}$ become essentially different when $\delta > \sqrt{c}$, then Monge-Kantorovich duality along with Lemma 1.3.2

implies the existence of a *robust* estimator for $\mathbf{P}$. As far as we are aware, no estimator which achieves the $\delta > \sqrt{c}$ threshold is known to be robust to $o(n)$ edits. For instance, one can remove all cycles of length $k = o(\log(n))$ with $o(n)$ edits, breaking any current cycle-based estimators which only count short cycles.

1.5 Sparse Graph Metrics

While it should be clear at this point that metrics for sparse graphs should not be able to enjoy all the properties of their counterparts for dense graphs, it is worth examining some candidate metrics for the sparse regime and trying to understand their potential. Consider the following rule for sampling from a graph. Given a graph G, sample a vertex $v \in V(G)$ uniformly at random and look at $B_r(v)$, the ball of radius r around v (considered as a BFS tree). By this sampling rule, G induces a distribution on rooted trees of depth at most r, denoted $\Gamma(G, r)$. We can now define the local metric, d_{loc} by

$$d_{\text{loc}}(G, G') = \sum_{r=1}^{\infty} 2^{-r} d(\Gamma(G, r), \Gamma(G', r)),$$

where $d(\cdot, \cdot)$ is any metric on probability measures, e.g., total variation distance. The normalization by 2^{-r} here in each term is somewhat arbitrary, and is only to ensure that d_{loc} induces the product topology, i.e., so that a sequence of graph is convergent with respect to d_{loc} if and only if the distributions $\Gamma(G, r)$ converge with respect to d for every r.

On the positive side, it is known and easy to show that a sequence of draws from $G(n, c/n)$ converges with probability 1 with respect to d_{loc}. However, d_{loc} is clearly flawed, as seen by the following example which we alluded to previously. Consider on one hand $G \sim G(n, c/n)$, and on the other a uniformly random bipartite graph H with the same average degree. For any constant (with respect to n) radius r, the distributions $\Gamma(G, r)$ and $\Gamma(H, r)$ are identical with high probability, and so we will have $d_{\text{loc}}(G, H) \to_n 0$ despite a glaring global dissimilarity between G and H.

In response to this, the following metric, introduced in [4], but studied more thoroughly in [5], has become a promising candidate in the linear sparsity regime. For two graphs G and H, we define

$$d_{\text{part}}(G,H) = \sum_{k=1}^{\infty} 2^{-k} d_H(\mathcal{M}_k(G), \mathcal{M}_k(H)),$$

where $\mathcal{M}_K(G)$ is the set of all $k \times k$ matrices which are obtained from G by averaging its edges over nearly balanced partitions $\Pi = (P_1, \ldots, P_k)$ of the vertices, and d_H is the Hausdorff metric on the space of such matrices, defined with respect to the ℓ_∞ norm. For graphs with $\leq Cn$ edges, C some constant, the entries of matrices in $\mathcal{M}_k(G)$ are bounded and so the resulting metric space is compact.

As an example, at the $k = 2$ scale, d_{part} sees the value of every bisection in G and H, and so is certainly able to distinguish between a random graph and a random bipartite graph of the same average degree. Unfortunately, it is not even known whether a sequence of i.i.d. draws from a $G(n, c/n)$ are Cauchy with respect to d_{part}. It is known that the draws concentrate as n grows, but this does not exclude the possibility of fluctuations as n varies. This is a special case of a conjecture from [5].

Conjecture 1.5.1. *For every c, a sequence $\{G_n\}$ with $G_n \sim G(n, c/n)$ is Cauchy with respect to d_{part} with probability 1.*

This issue of proving convergence for metrics on sparse sequences turns out to be rather rampant in the field. The only non-trivial metric known to converge on samples from $G(n, c/n)$ is d_{loc} (although one might argue that d_{loc} *is* a trivial metric in itself). The underlying issue is really that many basic parameters of $G(n, c/n)$ are still not even known to converge in the limit as $n \to \infty$. Some recent progress has been made along these lines: the authors of [3] show that many quantities which would be implied to converge by the convergence of d_{part}, including the (normalized) independence number of $G(n, c/n)$ and its max-cut, do actually converge. Their method is using an interpolation technique from [13] that has seen numerous applications and refinements over the years. It seems as though these techniques may give a method for proving convergence of d_{part} and other candidate metrics.

On the other hand, we can ask what types of graphs are distinguished by d_{part}. In particular, can the partition metric distinguish between $\mathbf{P}_n$ and $\mathbf{Q}_n$ all the way down to the detection threshold $\delta = \sqrt{c}$? Certainly when δ is large enough compared to c, the answer is yes, for the same reason why it is impossible to couple the models to $o(n)$ edit distance in this regime. However, the answer at the threshold is not at all clear, mostly due to our lack of understanding of bisections.

We would like to say that, given the lower bound for detection in the planted bisection model $\mathbf{P}_n$, no metric should be able to separate i.i.d. draws from $\mathbf{P}_n$ and $\mathbf{Q}_n$ below the detection threshold $\delta = \sqrt{c}$, but this is certainly not the case. Indeed, the cut distance is an easy counterexample to this statement. However, the cut distance does not have *any* non-trivial convergent sequences, and so should be omitted from our desired result. The following theorem makes this precise and is seemingly the best possible statement that can be made here.

Theorem 1.5.2. *Let $d(\cdot,\cdot)$ be any metric on $\mathcal{G}$, and δ, c with $\delta < \sqrt{c}$ be given. Take $\{G_n\}_{n\in\mathbb{N}}$ and $\{H_n\}_{n\in\mathbb{N}}$ to be sequences of i.i.d. draws from $\mathbf{Q}_n$ and $\mathbf{P}_n$, respectively. If both $\{G_n\}$ and $\{H_n\}$ are convergent, in the sense that for every $\varepsilon > 0$ there exist constants $N_1(\varepsilon)$ and $N_2(\varepsilon)$ such that for every $n, m \geq N_1(\varepsilon)$ we have*

$$d(G_n, G_m) < \varepsilon \text{ with probability } 1 - o(1),$$

and for every $n, m \geq N_2(\varepsilon)$ we have

$$d(H_n, H_m) < \varepsilon \text{ with probability } 1 - o(1),$$

then any sequence obtained by interleaving the two sequences $\{G_n\}$ and $\{H_n\}$ is also convergent with high probability.

Proof. Suppose $G \sim \mu_n$, and consider the following randomized test for classifying G. Independently generate $H_n \sim \mathbf{P}_n$ and $G_n \sim \mathbf{Q}_n$, and output that $G \sim \mathbf{P}_n$ if $d(G, H_n) \leq d(G, G_n)$, outputting $G \sim \mathbf{Q}_n$ otherwise. If d were to separate G_n and

H_n, in the sense that there exists an $\eta > 0$ and $N \in \mathbb{N}$ such that

$$\forall n \geq N, \; d(G_n, H_n) \geq \eta \;\text{ with probability } 1 - o(1),$$

then for every $n \geq N = \max(N, N_1(\eta), N_2(\eta))$, if $G \sim \mathbf{P}_n$ we have

$$d(G, H_n) < \eta \leq d(G, G_n) \;\text{ with probability } 1 - o(1),$$

and similarly if $G \sim \mathbf{Q}_n$, so this test succeeds at classifying G with high probability. Negating the assumption, it must be the case that for every $\eta > 0$ and $N \in \mathbb{N}$ there exists an $n \geq N$ such that

$$d(G_n, H_n) < \eta \text{ with some constant probability.}$$

Now let $\varepsilon > 0$, and choose $N = \max(N_1(\varepsilon/3), N_2(\varepsilon/3), N_3(\varepsilon/3))$. Then, for every $n, m \geq N$, let the index $k \geq N$ be such that $d(G_k, H_k) < \varepsilon/3$. Then by the triangle inequality we have

$$d(G_n, H_m) \leq d(G_n, G_k) + d(G_k, H_k) + d(H_k, H_m) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \qquad \square$$

An ideal metric should distinguish between block models right down to the detection threshold, and so it would be interesting to know whether any existing metrics can do this. There are at least three more interesting candidate metrics that we have not treated here. One of these is a metric based on neighborhoods of properly colored graphs, which can be shown to refine both d_{loc} and d_{part}. This metric has been studied on bounded degree graphs and has some nice properties in that setting (see [14]), but little is known about how it behaves on block models. Another interesting metric is based on a notion of 'right convergence' which defines a metric based on homomorphisms *out* of a graph (as opposed to d_{loc} which is defined based on homomorphisms in). This seems like a particularly interesting metric, as it encodes information like free energies which in turn appear in much of the block-model literature, and may be the best chance at passing the block model benchmark. Finally, a metric based on large deviation rates from [6] is known to imply both partition and right convergence.

Chapter 2

Dense extremal graph problems

2.1 Introduction

Extremal graph theory is the study of asymptotic relations between graphs and their subgraphs. In an ambitious attempt to characterize *all* such relations, Erdös, Lovász and Spencer, in [12], initiate the study of a rather complex object. The *profile* of a graph G contains all the information about its subgraphs. Precisely, it is a (discrete) distribution on subgraphs, where the probability of any $F \subseteq G$ is just the probability that $v(F)$ vertices taken uniformly from $V(G)$ induce a subgraph isomorphic to F. It should be clear that not all distributions can be obtained as profiles of graphs: Mantel's theorem, stating that a graph without triangles can have at most $1/2$ of all possible edges, is a nice example of this. It suffices to note that a complete characterization of graph profiles would solve all extremal problems of a very general type to understand the complexity of this question.

Since [12], very little progress has been made towards a better understanding of graph profiles. However, in the last decade, extremal graph theory has seen the development of a number of new powerful tools. Of particular note is the theory of dense graph limits which we have briefly touched on already and also the flag algebra method from [25]. Using what can be thought of as a combination of these techniques, Razborov [26] was able to answer possibly the simplest (smallest at least) non-trivial question about graph profiles. This theorem, which we will see in Section 2.3, is in fact highly non-trivial despite its seeming simplicity.

2.2 Basic background

Rather than counting subgraphs directly, it is often much more convenient to work with *homomorphisms*. A homomorphism from F to G is a map $\varphi : V(F) \to V(G)$ such that whenever $(u, v) \in E(F)$, we have $(\varphi(u), \varphi(v)) \in E(G)$. Note that when a homomorphism is injective, its image in G contains subgraph isomorphic to F, but a homomorphism may also map two non-adjacent vertices to a single vertex in G. As G grows, the number of homomorphisms from F to G is unbounded, so we should normalize this by $v(G)^{v(F)}$ to obtain the *homomorphism density* of F in G, denoted $t(F, G)$, which is simply the probability that a uniformly random map from $V(F)$ to $V(G)$ is a homomorphism.

Given a fixed collection of graphs $\mathcal{F} = \{F_1, \ldots, F_m\}$, we can associate to every graph G a vector in $\mathbb{R}^m$, namely $t_G = (t(F_1, G), \ldots, t(F_m, G))$. Here, t_G plays the role of the profile of G (with respect to $\mathcal{F}$), and so the set of all possible profiles can be written as

$$T_{\mathcal{F}} = \{t_G \mid G \in \mathcal{G}\}.$$

In [12], the collection $\mathcal{F}$ under consideration is the set of all connected graphs on $\leq n$ vertices, and they define the sets

$$S(n) = \overline{T_{\mathcal{F}}}, \tag{2.1}$$

taking the closure of the set of possible profiles.

The closure operation here may seem awkward to work, but there are many extremal problems for which finite graphs cannot obtain the optimum value, and a growing sequence of graphs must be specified. This is where working with graph limits will prove to be particularly useful. For a graph F, we can extend the notion of homomorphism density in a graph to graphons via the formula

$$t(F,W) = \int_{[0,1]^{v(F)}} \prod_{(u,v)\in E(F)} W(x_u,x_v) \prod_{u\in V(F)} \mathrm{d}x_u.$$

To see how this relates to homomorphism density in a graph, consider the following construction of a graphon. Given a weighted graph H, i.e., a collection of node weights $\alpha_1,\dots,\alpha_k$ and edge weights b_{ij}, we define a graphon W_H as follows. Partition $[0,1]$ as $I_1 \cup \cdots \cup I_k$ according to the nodeweights and then set

$$W_H(x,y) = \sum_{i,j} \beta_{ij} \cdot \mathbf{1}\left((x,y) \in I_i \times I_j\right). \tag{2.2}$$

The quantity $t(F, W_H)$ has a nice interpretation in terms of H: it is just the expectation

$$\mathbf{E}_{\varphi\sim\alpha} \prod_{(u,v)\in E(F)} \beta_{\varphi(u)\varphi(v)},$$

where the expectation is over a map $\varphi : V(F) \to V(H)$ taken according to the nodeweights of H. In the special case when all the nodeweights are $1/n$ and the edgeweights lie in $\{0,1\}$, we recover the definition of homomorphism density in a simple graph, and here the graphon W_H is called the *blowup* of H. Graphons of the form W_H for some weighted graph H are always step functions, and we will make heavy use of the construction (2.2) in Section 2.3.

The method of flag algebras is a powerful tool for proving inequalities between homomorphism densities. Given a collection of graphs $F_1,\dots,F_m$ and corresponding weights $\alpha_1,\dots,\alpha_m$, we can ask whether the inequality

$$\sum_{i=1}^{m} \alpha_i t(F_i, W) \geq 0 \tag{2.3}$$

is true for every $W \in \mathcal{W}$. A common shorthand used here is to define

$$f := \sum_{i=1}^{n} \alpha_i F_i$$

as a formal linear combination and then write

$$t(f, W) := \sum_{i=1}^{m} \alpha_i t(F_i, W).$$

Such formal linear combinations of graphs are called *quantum graphs.* The template (2.3) is very general and most extremal problems can be cast in this form. If the F_i and $\alpha_i \in \mathbb{Q}$ are given as input, it is known through the results of [15] to be an undecidable problem to determine whether (2.3) holds.

Despite this, the framework of flag algebras gives a method for proving such inequalities by representing the left side of (2.3) as a sum of squares in a certain algebra, and this sum of squares representation can be found using semidefinite programming. We will defer the development of this theory to Section 2.5, so for now let us mention only one easy way of obtaining inequalities of the form (2.3).

From a graph F and a graphon W, we can define an induced analog of homomorphism density as

$$t_{\text{ind}}(F, W) = \int \prod_{(u,v)\in E(F)} W(x_u, x_v) \prod_{(u,v)\notin E(F)} (1 - W(x_u, x_v)) \prod_{u\in V(F)} \mathrm{d}u.$$

When W represents a simple graph G through (2.2), $t_{\text{ind}}(F, W)$ can be interpreted simply as the probability that a random map $\varphi : V(F) \to V(G)$ preserves both adjacency and non-adjacency (a *strong* homomorphism). The following simple relation between induced and non-induced densities will be useful to us. For every graph F and graphon W we have

$$t_{\text{ind}}(F, W) = \sum_{F'\supseteq F} (-1)^{e(F')-e(F)} t(F', W),$$

where in the sum we treat the F' as labelled supergraphs of F (so their multiplicity is counted). This is just the inclusion-exclusion principle and is not a particularly deep result. Since $t_{\text{ind}}(F, W) \geq 0$ by definition, we immediately obtain an inequality of the form (2.3). It is often convenient to refer to this formula using a quantum graph, and so for a graph F we define the quantum graph $\mathrm{Ind}(F)$ by

$$\mathrm{Ind}(F) := \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} F'$$

so that

$$t_{\mathrm{ind}}(F, W) = t(\mathrm{Ind}(F), W)$$

for every graphon W.

2.3 Graph Profiles

Very little is known about the sets $S(n)$ (recall the definition (2.1)) in general. In [12], it is proved that they are connected and have a non-trivial interior, so that $S(n)$ has dimension precisely given by the number of connected graphs on $\leq n$ vertices. Let us now consider $S(3)$, for which we still have a very basic understanding. The three connected (non-trivial) graphs on at most 3 vertices are K_2 (the edge), $K_{1,2}$ (the vee), and K_3 (the triangle). Thus, S_3 is a subset of $[0,1]^3$. Unfortunately, almost nothing is known about $S(3)$ as a 3-dimensional object, and so we should first consider its 2-dimensional projections.

2.3.1 Projections

Of the three coordinate-wise projections of $S(3)$, only one has been considered in any detail as far as we know. This is the $K_{1,2}$ projection or, more typically, the set of possible edge and triangle pairs, denoted

$$D_{2,3} = \{(x, z) \mid (x, y, z) \in S(3) \text{ for some } y\}.$$

As we will see shortly, even $D_{2,3}$ is a highly non-trivial set, although at this point it is nearly completely understood. Let us also give a name to the other two projections. Let

$$D_{ev} = \{(x, y) \mid (x, y, z) \in S(3) \text{ for some } z\}$$

denote the edge-vee pairs, and

$$D_{vt} = \{(y, z) \mid (x, y, z) \in S(3) \text{ for some } x\}$$

the vee-triangle pairs.

Now let us turn to a characterization of (the boundary of) these sets. Since they are subsets of the unit square, it will suffice to give an upper and lower bounding curve. To see that these bounds are best possible, we should also give extremal graphons that attain every point on the curve. We will use the shorthands x, y and z from the definitions of $D_{2,3}$, D_{ev}, and D_{vt} to simplify our notation, and a statement of the form, e.g., $z \leq f(x)$, will mean that $t(K_3, W) \leq f(t(K_2, W))$ for every $W \in \mathcal{W}$.

It will be useful in what follows to have notation for a few special weighted graphs. Let $H_1(\alpha)$ denote the weighted graph on two vertices with nodeweights α and $(1-\alpha)$. The only non-zero edgeweight is $b_{11} = 1$. We also let W_p denote the constant function p.

Theorem 2.3.1 (Triangles vs edges)**.** *The upper bounding curve of $D_{2,3}$ is given by $z \leq x^{3/2}$.*

The lower bound is

$$z \geq \frac{(k-1)(k-2)}{k^2} \left(1 + \sqrt{1 + \frac{kx}{k-1}}\right)^2 \left(1 - 2\sqrt{1 - \frac{kx}{k-1}}\right),$$

where $k = \lceil 1/(1-x) \rceil$.

The upper bound here is a well known result which is hard to attribute to anyone in particular. It follows from a simple version of the Kruskal-Katona theorem and can also be proved using a slightly clever application of Cauchy-Schwarz in the flag algebra. It is also tight, as seen by taking the graphon $W_{H_1(\sqrt{x})}$.

The lower curve is from [26] and is somewhat easier to understand from the extremal graphs. They are W_H, where H is a weighted complete graph on k nodes: all the edgeweights are 1 and $k-1$ of the nodeweights are equal with the last being at most as large as the rest. It is a fairly routine calculation to determine what the nodeweights should be given x, and then to compute the triangle density given the nodeweights. Solving for z as a function of x gives precisely what we see in the statement of Theorem 2.3.1.

Since 2008, Razborov's result has been extended to the setting in which triangles are replaced with larger complete graphs. The first of these, due to Nikiforov [23], is able to obtain the result for K_4 by carefully optimizing certain multilinear forms. This method is also able to re-prove Razborov's triangle result, but does not seem to extend to K_5 or higher. Finally, Reiher in [27] obtains the result for every K_r using what can be seen as a generalization of Razborov's original proof.

Moving along, let us consider the case of D_{ve}. In comparison to $D_{2,3}$, the boundary of D_{ve} will look somewhat simple. Indeed, no complex machinery is required in this case.

Theorem 2.3.2 (Vees vs edges)**.** *The upper bounding curve of D_{ve} is given by*

$$y \leq \max(x^{3/2}, (1-x)^{3/2} + 2x - 1),$$

and the lower bound is given by

$$y \geq x^2.$$

After some digging, the upper curve here is probably best attributed to [1], although this was known in some special cases even earlier. For $x \geq 1/2$, the upper curve is attained by $W_{H_1(\sqrt{x})}$, and for $x \leq 1/2$, by the complements $1 - W_{H_1(\sqrt{1-x})}$. The lower curve is probably too simple to attribute to anyone in particular: it is just the Cauchy-Schwarz on the degree of a vertex, and is attained by many different graphs, for example W_x.

The boundary of D_{tv} is not yet completely understood. The upper bounding curve is simple: The inclusion bound $z \leq y$ is tight using $W_{H_1(\alpha)}$ for a suitable weight α. On the other hand, the lower bound appears to be open. However, we are fairly certain that the extremal graphs are precisely those appearing on the lower bounding curve of $D_{2,3}$ (the ones we constructed above). Our evidence is based on a number of flag algebra computations, optimizing z subject to $y = d$, where d is some fixed rational number. In all the values d we checked, the SDP was able to prove a bound which is numerically (with 10^{-16} error) matched by the conjectured extremal graphs.

2.3.2 Slices

As evidenced by the complexity of the projections $D_{2,3}$ and D_{tv}, the full structure of $S(3)$ may be extremely difficult to understand. The most natural step beyond taking projections is to consider a slice by fixing some linear combination of the coordinates. In particular, by setting the triangle density z to 0, we obtain a reasonably nice looking shape.

Theorem 2.3.3 (Triangle-free slice). *The slice $\{(x, y, z) \in S_3 \mid z = 0\}$ is bounded from above by the line $y = x/2$ and from below by the parabola $y = x^2$, where $x \in [0, 1/2]$.*

Proof. The triangle-free assumption here immediately implies that $x \in [0, 1/2]$. The lower curve is implied by the lower curve for D_{ve}, and is attained by W_H, where H is has two vertices, $\alpha_1 = \alpha_2 = 1/2$ and $b_{12} = 2x$.

The upper curve here comes from considering $t_{\text{ind}}(K_2 \cup K_1)$. By (2.2), this gives

$$t(K_2, W) - 2t(K_{1,2}, W) + t(K_3, W) \geq 0$$

for every $W \in \mathcal{W}$. In particular, when $t(K_3, W) = z = 0$, we get $y \leq x/2$. The extremal graph is W_H, where H has two vertices of weights α and $(1 - \alpha)$ and $b_{12} = 1$. On W_H, we have

$$x = 2\alpha(1 - \alpha), \ y = \alpha(1 - \alpha)^2 + (1 - \alpha)\alpha^2,$$

and solving for y as a function of x gives exactly $y = x/2$.

□

Although we had hoped that the slices $z = d$ might have a reasonably nice shape, especially for small d, it is not clear that this it the case at all. Even using flag algebra computations, there are gaps in both the upper and lower curve on which we have no idea what the correct value is.

2.4 Remarks

We have seen that even the simplest of the sets $S(n)$, $S(3)$ is already extremely complex. Slices seem to be the best in-road here, and although the triangle-free slice admits a particularly nice solution, there is no reason to think triangle slices are the correct answer. It would be interesting to see if there is a slice $ax + by + cz = d$ for which anything can be said for d in some interval. It may be possible to find such a slice by iterating over a number of different choices for a, b, c, d and doing flag algebra computations.

One way to obtain insight into the structure of $S(3)$ is to examine more closely the boundary of its projections. Whenever the extremal graph attaining the boundary is unique, this must correspond to a point or ridge in 3 dimensions. This is the case, e.g., for the graphs on the upper curve of $D_{2,3}$. On the lower curve here, the points on the convex hull are also unique, but the points in between have many extremal graphs. To see this, recall that the the extremal graphs here are partially unbalanced complete graphs. If we restrict to the subgraph induced by two vertices, one of which is the smaller weight vertex, the resulting graph is bipartite, and hence triangle-free. Replacing this part of W_H with any triangle-free graph with the same edge density, we obtain a new extremal graph. As Theorem 2.3.3 shows, even if the edge density $x < 1/2$ is fixed, a triangle-free graph can have between $x/2$ and x^2 vees, so there is in fact a continuum of extremal graphons. It is not even known whether this is the only way to produce new extremal graphs here. See [24] for some progress on this front.

While we have not really made this clear, so far we were only considering the boundary of these sets. One reason for this is that an extremal problem involving graphs on $\leq n$ vertices has its optimum somewhere on the boundary of $S(n)$. In fact, it is possible to linearize most extremal problems, subject to a blowup in the size of the graphs, so it would suffice to check the convex hull of $S(N)$, for some N depending on n and the degree of the extremal problem. It is not at all clear what is happening in the interior of these sets, even in the case of $D_{2,3}$. In the triangle-free slice, it is actually not hard to see that the slice in fact contains the entire region within the boundary by interpolating continuously between the extremal graphs on the upper and lower curve, so this type of technique could be useful for proving similar results about other sets.

Finally, while we have made an attempt to say something non-trivial about $S(3)$, a 3-dimensional problem, there are any number of 2-dimensional problems that remain unsolved. Moving to graphs with more than 3 vertices, the only real results are from optimizing the density of complete graphs against a fixed edge density. It seems likely that there are some easier things to be said about the possible configurations of two 4-vertex graphs, for instance. An exhaustive process would become tiresome rather quickly though, and we feel that a complete characterization of $S(3)$ is possibly within reach, given the powerful tools available in modern graph theory.

2.5 Labeled graphs and flag algebras

Now that we have seen examples of fairly simple proofs of a few inequalities, let us develop some more powerful machinery. A *k-labeled graph F* is a graph in which k of its vertices have been labeled with the integers $\{1, \ldots, k\}$. Given two k-labeled graphs F_1 and F_2, we can define their *gluing product* $F_1 \cdot F_2$ by taking the disjoint union of F_1 and F_2, identifying the labeled vertices and collapsing any resulting multiple edges. We also define the *unlabeling operator* $[\![\cdot]\!]$ on k-labeled graphs which removes all the labels and outputs the unlabeled graph. We similarly define a k-quantum graph as a formal linear combination of k-labeled graphs, and extend the unlabeling operator to k-quantum graphs by linearity.

The crux of the flag algebra method (also known as the semi-definite method) comes from the following fact. Given a vector $z = (f_1, \ldots, f_m)$ of k-quantum graphs and an $m \times m$ positive semi-definite matrix A,

$$[\![z^T A z]\!] \geq 0,$$

in the sense that

$$\sum_{i,j} A(i,j) t([\![f_i \cdot f_j]\!], W) \geq 0$$

for every graphon $W \in \mathcal{W}$. Given a quantum graph f and a parameter α, if we can find z and A such that

$$f - \alpha = [\![z^T A z]\!],$$

we obtain the inequality $f \geq \alpha$. With f and z as inputs, we can build a semi-definite program whose dual produces the matrix A and α being the value of the program itself. Hence, the program produces a proof of the best possible inequality it can find given f and also the collection of k-quantum graphs making up z. In general, taking z to be large will give a better value of α, but it is often the case that a fairly small number of graphs produces the optimal value. For the full details, we refer the reader to [16].

2.6 Edge inducibility

An innocent sounding problem is, given a graph F, to maximize $t_{\text{ind}}(F, W)$ over graphons W. This value (although usually normalized differently) is known as the *inducibility* of F, and is denoted $i(F)$. Such problems have been studied from quite long ago (see [9] and also [16]). These problems turn out to be extremely difficult and the inducibility of most graphs is unknown, barring some special families of graphs or very small graphs. Even the inducibility of the 4-vertex path is famously unknown despite our modern tools and techniques.

A related notion has recently been gaining some traction as a topic of interest. Following [2], given parameters k and ℓ we define $I(G, k, \ell)$ to be number of k-subsets of $V(G)$ inducing exactly ℓ edges. We then define

$$I(n,k,\ell) = \max_{G\in\mathcal{G}_n} I(G,k,\ell),$$

where $\mathcal{G}_n$ denotes the set of n-vertex graphs. Finally we can define the *edge-inducibility* of k and ℓ by

$$i(k,\ell) = \lim_{n\to\infty} \frac{I(n,k,\ell)}{\binom{n}{k}}.$$

Notice that if F is the unique graph with k vertices and exactly ℓ edges, then

$$i(k,\ell) = \frac{k!}{|\Gamma(F)|} i(F),$$

where $\Gamma(F)$ denotes the automorphism group of F. The factor $k!/|\Gamma(F)|$ here comes from the differently normalizing factors, along with the fact that if F occurs as an induced subgraph of a graph G, then there are precisely $|\Gamma(F)|$ distinct strong homomorphisms mapping F to G.

It turns out that edge-inducibility can always be written as a standard inducibility problem, but of a quantum graph, in the following sense.

Lemma 2.6.1. *Let $k > 0$ and $0 < \ell \leq \binom{k}{2}$ be given, and denote by $\mathcal{G}(k,\ell)$ the set of k-vertex graphs with exactly ℓ edges. Then*

$$i(k,\ell) = \max_{W\in\mathcal{W}} \left\{ \sum_{F\in\mathcal{G}(k,\ell)} \frac{k!}{|\Gamma(F)|} t_{\text{ind}}(F,W) \right\}.$$

Another way of writing this is to take the quantum graph

$$f_{k,\ell} := \sum_{F\in\mathcal{G}(k,\ell)} \frac{k!}{|\Gamma(F)|} F \tag{2.4}$$

in which case

$$i(k,\ell) = \max_{W\in\mathcal{W}} t_{\text{ind}}(f_{k,\ell}, W),$$

a fact we will make use of in our proofs below.

Looking at small values of k and ℓ, there are a number of cases which are known. When $k = 3$, there is a unique graph with ℓ edges for every possible ℓ and the inducibility of all such graphs is known. Moving to $k = 4$ however, the picture is much less complete. The value $i(4, 1) = i(4, 5)$ is equal to $i(K_{1,1,2})$ which was originally resolved in [16] using flag algebras. At $k = 5$ it seems that no values are currently known. In what follows, we will resolve the values of $i(4, 2)$ and $i(5, 6)$ exactly.

To simplify the notation in our proofs, we need to explain how we represent graphs through the flag algebra computations. Graphs are sorted according to their lexicographically smallest adjacency matrix. Labeled graphs are represented using 3 parameters, n, k, and t: n represents the number of vertices, k the number of labels, and t the graph induced by the labeled vertices (according to the aforementioned ordering). Given n, k, and t, the labeled graphs are again sorted by their lexicographically smallest adjacency matrix, subject to the labeled vertices taking their own values. We denote this list of labeled graphs by $F(n, k, t)$. Further, whenever we indicate a graph i from a set $F(n, k, t)$, we in fact mean $\mathrm{Ind}(i)$. This is due to the way our program represents the product of induced graphs and makes the calculations consistent with the gluing product we defined above. As an example, $F(3, 2, 0)$ contains four 3-vertex, 2-labeled graphs in which the labeled vertices are not connected by an edge ($t = 0$). Graph 1 here represents the empty graph, graph 4 the path with labeled endpoints, and graphs 2 and 3 represent the remaining two graphs depending on which labeled vertex is isolated (these are different as labeled graphs).

Theorem 2.6.2. *We have*

$$i(4, 2) = \frac{1}{2}.$$

Proof. Consider the graphon W which is the blowup of two disjoint triangles. It is not too hard to see that

$$t_{\mathrm{ind}}(f_{4,2}, W) = \frac{1}{2},$$

which proves that $i(4, 2) \geq 1/2$.

For the upper bound, we will need the following positive semi-definite matrices.

$$A_1 = \frac{1}{8}\begin{pmatrix} 4 & -2 & -8 \\ -2 & 1 & 4 \\ -8 & 4 & 16 \end{pmatrix}. \qquad A_2 = \frac{71}{400}.$$

$$A_3 = \frac{1}{80}\begin{pmatrix} 17 & -19 & -13 \\ -19 & 25 & 7 \\ -13 & 7 & 25 \end{pmatrix}. \qquad A_4 = \frac{697}{800}.$$

$$A_5 = \frac{1}{1600}\begin{pmatrix} 5260 & 1223 & 4006 & -5856 \\ 1223 & 291 & 929 & -1367 \\ 4006 & 929 & 3052 & -4458 \\ -5856 & -1367 & -4458 & 6524 \end{pmatrix}.$$

$$A_6 = \frac{1}{1600}\begin{pmatrix} 1380 & -1109 & -1921 & -1110 \\ -1109 & 2612 & -861 & 1576 \\ -1921 & -861 & 6310 & 314 \\ -1110 & 1576 & 314 & 1440 \end{pmatrix}.$$

The corresponding basis vectors (using the notation described above) are

$$z_1 = (1, 2+3, 4) \text{ from } F(3,2,0).$$

$$z_2 = (2-3) \text{ from } F(3,2,0).$$

$$z_3 = (1, 2+3, 4) \text{ from } F(3,2,1),$$

$$z_4 = (2-3) \text{ from } F(3,2,1).$$

$$z_5 = (1, 2+3, 4, 5) \text{ from } F(4,3,1).$$

$$z_6 = (1, 2, 6, 7) \text{ from } F(4,3,1).$$

From here it suffices to verify the calculation

$$-\mathrm{Ind}(f_{4,2}) + \frac{1}{2} = \sum_{i=1}^{6} [\![z_i^T A_i z_i]\!] \geq 0,$$

from which it follows that

$$t_{\mathrm{ind}}(f_{4,2}, W) \leq \frac{1}{2}$$

for every graphon $W \in \mathcal{W}$. In particular,

$$i(4,2) = \max_{W \in \mathcal{W}} t_{\mathrm{ind}}(f_{4,2}, W) \leq \frac{1}{2},$$

completing the proof. □

Theorem 2.6.3. *We have*

$$i(5,6) = \frac{5}{8}.$$

Proof. The lower bound here comes from taking W to be the blowup of an edge (or any balanced complete bipartite graph). For the upper bound, our matrices (all positive semi-definite) are as follows.

$$A_1 = \frac{1}{2400}\begin{pmatrix} 1434 & -2967 \\ -2967 & 6500 \end{pmatrix}. \qquad A_2 = \frac{477}{800}. \qquad A_3 = \frac{1}{8}\begin{pmatrix} 5 & -5 \\ -5 & 5 \end{pmatrix}.$$

$$A_4 = \frac{1}{4800}\begin{pmatrix} 12660 & -10848 & 10848 & -16217 \\ -10848 & 12276 & -12276 & 18760 \\ 10848. & -12276 & 12276 & -18760 \\ -16217 & 18760 & -18760 & 33240 \end{pmatrix}.$$

$$A_5 = \frac{1}{2400}\begin{pmatrix} 8464 & -209 \\ -209 & 6189 \end{pmatrix}.$$

$$A_6 = \frac{1}{4800}\begin{pmatrix} 67200 & -28195 \\ -28195 & 26862 \end{pmatrix}.$$

The corresponding basis vectors are

$$z_1 = (1, 4) \text{ from } F(3, 2, 0).$$

$$z_2 = (2 - 3) \text{ from } F(3, 2, 0).$$

$$z_3 = (1, 4) \text{ from } F(3, 2, 1).$$

$$z_4 = (1, 4, 5, 6 + 7) \text{ from } F(4, 3, 1).$$

$$z_5 = (2 - 3, 6 - 7) \text{ from } F(4, 3, 1).$$

$$z_6 = (1, 4 - 6) \text{ from } F(4, 3, 2).$$

As in the proof of the previous theorem, it suffices to verify the calculation

$$-\mathrm{Ind}(f_{5,6}) + \frac{5}{8} = \sum_{i=1}^{6} [\![z_i^T A_i z_i]\!] \geq 0,$$

from which the result follows. □

While we do not make any claims to the uniqueness of the extremal graphs, it seems likely that the construction for $i(4, 2)$ is not unique while that of $i(5, 6)$ is. This is based on looking at the solution to the primal SDP: when the values take seemingly arbitrary floating-point numbers, as in the case of $i(4, 2)$, it usually indicates multiple primal solutions. On the other hand, when the primal solution takes on only nice rational values, it generally turns out to be a unique solution.

At $k = 4$, all possible values of ℓ are now (essentially) known. The only remaining value is the self-complementary $i(4, 3)$, but floating-point calculations give an upper bound of around $0.5 + 10^{-16}$ and the lower bound of $1/2$ is attained by the blowup of an edge again. It would certainly be possible to convert this floating-point proof to a rational one as we have done above, but the proof requires passing to graphs on 6 vertices which blows up the size of the matrices considerably.

Moving to $k = 5$, the value of $i(5, 4) = i(5, 6)$ given in Theorem 2.6.3 is seemingly the only one in reach currently. As above, $i(5, 1) = i(5, 9)$ is now $i(K_{1,1,1,2})$, which is open, and none of $i(5,2)$, $i(5,3)$, or $i(5,5)$ seem amenable to a flag algebra solution. However, it is certainly possible that other techniques could lead to some easy proofs of those cases.

Bibliography

[1] Rudolf Ahlswede and Gyula O. H. Katona. Graphs with maximal number of adjacent pairs of edges. *Acta Math. Acad. Sci. Hungar.*, 32(1-2):97–120, 1978.

[2] Noga Alon, Dan Hefetz, Michael Krivelevich, and Mykhaylo Tyomkyn. Edge-statistics on large graphs. *Combinatorics, Probability and Computing*, 29(2):163–189, 2020.

[3] Mohsen Bayati, David Gamarnik, and Prasad Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.*, 41(6):4080–4115, 2013.

[4] Béla Bollobás and Oliver Riordan. Metrics for sparse graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 211–287. Cambridge Univ. Press, Cambridge, 2009.

[5] Béla Bollobás and Oliver Riordan. Sparse graphs: metrics and random models. *Random Structures Algorithms*, 39(1):1–38, 2011.

[6] Christian Borgs, Jennifer Chayes, and David Gamarnik. Convergent sequences of sparse graphs: A large deviations approach. *Random Structures & Algorithms*, 51(1):52–89, 2017.

[7] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.*, 372(5):3019–3062, 2019.

[8] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, and Kati Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[9] Jason I. Brown and Alexander Sidorenko. The inducibility of complete bipartite graphs. *J. Graph Theory*, 18(6):629–645, 1994.

[10] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011.

[11] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *Ann. Probab.*, 45(2):1190–1217, 2017.

[12] Paul Erdös, László Lovász, and Joel Spencer. Strong independence of graphcopy functions. In *Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977)*, pages 165–172. Academic Press, New York-London, 1979.

[13] Francesco Guerra and Fabio Lucio Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.

[14] Hamed Hatami, László Lovász, and Balázs Szegedy. Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.*, 24(1):269–296, 2014.

[15] Hamed Hatami and Serguei Norine. Undecidability of linear inequalities in graph homomorphism densities. *J. Amer. Math. Soc.*, 24(2):547–565, 2011.

[16] James Hirst. The inducibility of graphs on four vertices. *Journal of Graph Theory*, 75(3):231–243, 2014.

[17] James Hirst. Optimal couplings for sparse block models. *Proc. Amer. Math. Soc*, in press, 2020.

[18] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA*, 113(16):E2218–E2223, 2016.

[19] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[20] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 814–827. ACM, New York, 2016.

[21] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[22] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, June 2018.

[23] Vladimir Nikiforov. The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.*, 363(3):1599–1618, 2011.

[24] Oleg Pikhurko and Alexander Razborov. Asymptotic structure of graphs with the minimum number of triangles. *Combinatorics, Probability and Computing*, 26(1):138–160, 2017.

[25] Alexander A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[26] Alexander A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[27] Christian Reiher. The clique density theorem. *Annals of Mathematics*, 184(3):683–707, 2016.

[28] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[29] Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures Algorithms*, 53(3):504–536, 2018.

[30] J. Michael Steele. An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.*, 14(2):753–758, 1986.