On the 3-local profiles of graphs

Hao Huang* Nati Linial† Humberto Naves‡ Yuval Peled§ Benny Sudakov¶

Abstract

For a graph G, let $p_i(G), i = 0, ..., 3$ be the probability that three distinct random vertices span exactly i edges. We call $(p_0(G), ..., p_3(G))$ the *3-local profile* of G. We investigate the set $\mathcal{S}_3 \subset \mathbb{R}^4$ of all vectors $(p_0, ..., p_3)$ that are arbitrarily close to the 3-local profiles of arbitrarily large graphs. We give a full description of the projection of $\mathcal{S}_3$ to the (p_0, p_3) plane. The upper envelope of this planar domain is obtained from cliques on a fraction of the vertex set and complements of such graphs. The lower envelope is Goodman's inequality $p_0 + p_3 \geq \frac{1}{4}$. We also give a full description of the triangle-free case, i.e., the intersection of $\mathcal{S}_3$ with the hyperplane $p_3 = 0$. This planar domain is characterized by an SDP constraint that is derived from Razborov's flag algebra theory.

1 Introduction

For graphs H, G, we denote by $d(H; G)$ the induced density of the graph H in the graph G. Namely, the probability that a random set of $|H|$ vertices in G induces a copy of the graph H.

Many important problems and theorems in graph theory can be formulated in the framework of graph densities. Most of the emphasis so far has been on edge counts, or what is the same, on maximizing $d(K_2; G)$ subject to some restrictions. Thus Turán's theorem determines $\max d(K_2; G)$ under the assumption $d(K_s; G) = 0$ for some $s \geq 3$. The theorem further says that the optimal graph is the complete balanced $(s-1)$-partite graph. This was substantially extended by Erdős and Stone [6] who determined $\max d(K_2; G)$ under the assumption that the H-density (not induced) of G is zero for some fixed graph H. Their theorem also shows that the answer depends only on the chromatic number of H. Ramsey's theorem shows that for any two integers $r, s \geq 2$, every sufficiently large graph G has either $d(K_s, G) > 0$ or $d(\overline{K_r}, G) > 0$. The Kruskal-Katona Theorem [15, 16], can be stated as saying that $d(K_r; G) = \alpha$ implies that $d(K_s; G) \leq \alpha^{s/r}$ for $r \leq s$. Finding $\min d(K_s; G)$ under the assumption $d(K_r; G) = \alpha$ turns out to be more difficult. The case $r = 2$ of this problem was solved only recently in a series of papers by Razborov [21], Nikiforov [17] and Reiher [23]. A closely related question is to minimize $d(K_s; G)$ given that $d(\overline{K}_r; G) = \alpha$ for some real $\alpha \in [0, 1]$ and integers

*School of Mathematics, Institute for Advanced Study, Princeton 08540. Email: huanghao@math.ias.edu. Research supported in part by NSF grant DMS-1128155.

†School of Computer Science and engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email nati@cs.huji.ac.il. Research supported in part by the Israel Science Foundation and by a USA-Israel BSF grant.

‡Department of Mathematics, UCLA, Los Angeles, CA 90095. Email: hnaves@math.ucla.edu.

§School of Computer Science and engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email yuvalp@cs.huji.ac.il

¶Department of Mathematics, UCLA, Los Angeles, CA 90095. Email: bsudakov@math.ucla.edu. Research supported in part by NSF grant DMS-1101185, by AFOSR MURI grant FA9550-10-1-0569 and by a USA-Israel BSF grant.

$r, s \geq 2$. The case $\alpha = 0$ of this problem was posed by Erdős more than 50 years ago. Although, recently Das et al [3], and independently Pikhurko [18], solved it for certain values of r and s it is still widely open in general.

Numerous further questions concerning the numbers $d(H;G)$ suggest themselves. Thus Goodman [10] showed that $\min_G d(K_3;G) + d(\overline{K}_3;G) = 1/4 - o(1)$. As a random $G(n, \frac{1}{2})$ graph shows, this bound is tight. Erdős [5] conjectured that a $G(n, \frac{1}{2})$ graph also minimizes $d(K_r;G) + d(\overline{K}_r;G)$ for all r, but this was refuted by Thomason [25] for all $r \geq 4$. A simple consequence of Goodman's inequality is that $\min_G \max\{d(K_3;G), d(\overline{K}_3;G)\} = 1/8$. The analogue statement for $r = 4$ is not true as can be shown using an example of Franek and Rödl [8] (see [14] for the details). On the other hand, the max-min version of this problem is now solved. As we have recently proved [14], $\max_G \min\{d(K_r;G), d(\overline{K}_r;G)\}$ is obtained by a clique on a properly chosen fraction of the vertices.

Closely related to these questions is the notion of inducibility of graphs, first introduced in [19]. The inducibility of a graph H is defined as $\lim_{n\to\infty} \max_G d(H;G)$, where the maximum is over all n-vertex graphs G. This natural parameter has been investigated for several types of graphs H. E.g., complete bipartite and multipartite graphs [1, 2], very small graphs [7, 13] and blow-up graphs [11].

In light of this discussion, the following general concept suggests itself.

Definition 1.1. *For a family of finite graphs $\mathcal{H} = (H_1, ..., H_t)$, let $d(\mathcal{H};G) := (d(H_1;G), \ldots d(H_t;G))$. Define $\Delta(\mathcal{H})$ to be the set of all $\bar{p} = (p_1, ..., p_t) \in [0,1]^t$ for which there exists a sequence of graphs G_n, such that $|G_n| \to \infty$ and $d(\mathcal{H};G_n) \to \bar{p}$. We likewise define $\Delta_{\mathcal{G}}(\mathcal{H})$ where we require that $G_n \in \mathcal{G}$, an infinite families of graphs of interest (e.g. K_s-free graphs).*

The initial discussion suggests that it may be a very difficult task to fully describe $\Delta(\mathcal{H})$ or $\Delta_{\mathcal{G}}(\mathcal{H})$. Indeed, it was shown by Hatami and Norine [12] that in general it is undecidable to determine the linear inequalities that such sets satisfy. In this paper we solve two instances of this question.

We denote by $p_i(G)$ the probability that three distinct random vertices in the graph G span exactly i edges. The first theorem describes the possible distributions of 3-cliques and 3-anticliques in graphs (i.e., of (p_0, p_3)). We have Goodman's inequality [10] as a lower bound, and an upper bound from [14]. We show that these bounds fully describe all possible (p_0, p_3).

Theorem 1.2. *For $p_0 \in [0,1]$, let β be the unique root in $[0,1]$ of $\beta^3 + 3\beta^2(1-\beta) = p_0$. Then, $(p_0, p_3) \in \Delta(\overline{K_3}, K_3)$ iff*

$$p_0 + p_3 \geq \frac{1}{4} \text{ and } p_3 \leq \max\{(1 - {p_0}^{1/3})^3 + 3{p_0}^{1/3}(1 - {p_0}^{1/3})^2, (1-\beta)^3\}.$$

The analogous question concerning $\Delta(\overline{K_r}, K_r)$ for $r > 3$ is widely open. While the analogues upper bound is proved in [14], the situation with respect to the lower bounds is still poorly understood [9, 24].

The second theorem in this paper is proved using the theory of flag algebras [20]. This theory provides a method to derive upper bounds in asymptotic extremal graph theory. This is accomplished by generating certain semidefinite programs (=SDP) that pertain to the problem at hand. By passing to the dual SDP we derive necessary conditions for membership in $\Delta(\mathcal{H})$ or $\Delta_{\mathcal{G}}(\mathcal{H})$. Section 4 contains a self contained discussion, covering this perspective of the theory of flag algebras.

The theorem below demonstrates the special role that bipartite graphs play in the study of triangle-free graphs. As the theorem shows, all 3-local profiles of triangle-free graphs are realizable as well by bipartite graph. Moreover, the theory of flag algebras provides a complete answer to this question. This yields a different perspective to the fact that almost all triangle-free graphs are bipartite [4]. We denote the 3-vertex path by P_3 and its complement by $\overline{P_3}$. Also, as usual, $A \succeq 0$ means that the matrix A is positive semi-definite (=PSD). The class of bipartite (resp. triangle-free) graphs is denoted by $\mathcal{BP}$ (resp. $\mathcal{TF}$).

Theorem 1.3. *For $p_0, p_1, p_2 \geq 0$ s.t. $p_0 + p_1 + p_2 = 1$, the following conditions are equivalent:*

I. $(p_0, p_1, p_2) \in \Delta_{\mathcal{TF}}(\overline{K_3}, \overline{P_3}, P_3)$

II. $p_0 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + p_1 \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \succeq 0$

III. $(p_0, p_1, p_2) \in \Delta_{\mathcal{BP}}(\overline{K_3}, \overline{P_3}, P_3)$

The remainder of this paper is organized as following. In Section 2 we use random graphs to show the realizability of $\Delta_{\mathcal{G}}(\mathcal{H})$. In section 3 we prove Theorem 1.2. In Section 4 we use the theory of flag algebras to derive SDP constraints on membership in $\Delta_{\mathcal{G}}(\mathcal{H})$. In section 5 we prove Theorem 1.3. We close with some concluding remarks and several open problems.

2 Random Constructions

Let $\mathcal{H} = (H_1, ..., H_t)$ be a collection of graphs, and $\bar{p} = (p_1, ..., p_t) \in [0,1]^t$. In order to prove that $\bar{p} \in \Delta_{\mathcal{G}}(\mathcal{H})$, we need arbitrarily large graphs G for which $\|d(\mathcal{H}; G) - \bar{p}\|$ is negligible. We accomplish this using appropriately designed random G's. Let Π be a symmetric $n \times n$ matrix with entries in $[0,1]$ and zeros along the diagonal. Corresponding to Π is a distribution $G(\Pi)$ on n-vertex graphs where ij is an edge with probability $\Pi_{i,j}$ and the choices are made independently for all $n \geq i > j \geq 1$. We say that a graph G is *supported* on Π if G is chosen from $G(\Pi)$ with positive probability.

Lemma 2.1. *For every list of graphs $H_1, ..., H_t$ there exists an integer N_0 such that if $n > N_0$ and Π is an $n \times n$ matrix as above, then there exists an n-vertex graph G^* supported on Π such that*

$$\forall\, i = 1, ..., t \qquad \left| d(H_i; G^*) - \mathop{\mathbb{E}}_{G \sim G(\Pi)} [d(H_i; G)] \right| \leq \frac{1}{\sqrt{n}}$$

We note that the statement need not hold if $G(\Pi)$ is replaced by an arbitrary distribution on n-vertex graphs.

Proof. Fix a graph H. Let us view an n-vertex graph G as a $\binom{n}{2}$-dimensional binary vector. The mapping $G \mapsto d(H; G)$ has Lipschitz constant $\binom{|H|}{2} / \binom{n}{2}$. We can therefore apply Azuma's inequality and conclude that

$$\Pr_{G^* \sim G(\Pi)} \left[\left| d(H; G^*) - \mathop{\mathbb{E}}_{G \sim G(\Pi)} [d(H; G)] \right| \geq \frac{1}{\sqrt{n}} \right] \leq 2 \exp\left(- \frac{\binom{n}{2}}{2n \binom{|H|}{2}^2} \right).$$

Using the union bound and denoting $h = \max|H_i|$, we get

$$\Pr_{G^* \sim G(\Pi)} \left[\left\| d(\mathcal{H}; G^*) - \mathop{\mathbb{E}}_{G \sim G(\Pi)} [d(\mathcal{H}; G)] \right\|_\infty \le \frac{1}{\sqrt{n}} \right] \ge 1 - 2t \exp\left(-\frac{\binom{n}{2}}{2n\binom{h}{2}^2} \right) = 1 - o_n(1).$$

□

This lemma is easily generalized for hypergraphs of greater uniformity.

3 Distribution of 3-cliques and 3-anticliques

In this section we prove Theorem 1.2 and produce a full description of the set $\Delta(\overline{K_3}, K_3)$. We state the known lower and upper bounds and show that they fully describe this set.

Theorem 3.1 (Goodman [10])**.** *For every n-vertex graph G*

$$p_0(G) + p_3(G) \ge \frac{1}{4} - O\left(\frac{1}{n}\right).$$

Theorem 3.2 ([14])**.** *Let $r, s \ge 2$ be integers and suppose that $d(\overline{K}_r; G) \ge \alpha$ where G is an n-vertex graph and $1 \ge \alpha \ge 0$. Let β be the unique root of $\beta^r + r\beta^{r-1}(1-\beta) = \alpha$ in $[0,1]$. Then*

$$d(K_s; G) \le \max\{(1-\alpha^{1/r})^s + s\alpha^{1/r}(1-\alpha^{1/r})^{s-1}, (1-\beta)^s\} + o(1)$$

Namely, given $d(\overline{K}_r; G)$, the maximum of $d(K_s; G)$ is attained upto a negligible error-term either by a clique on some subset of the n vertices, or by the complement of such a graph. In particular, for every G

$$p_3(G) \le \max\{(1-p_0(G)^{1/3})^3 + 3p_0(G)^{1/3}(1-p_0(G)^{1/3})^2, (1-\beta)^3\},$$

where β is the unique root of $\beta^3 + 3\beta^2(1-\beta) = p_0(G)$ in $[0,1]$.

Proof of Theorem 1.2. Let C_1, C_2 be the (p_0, p_3) curves induced by cliques and complements of cliques resp.

$$C_1 = \left\{ \left((1-x)^3 + 3(1-x)^2x, x^3\right) \mid x \in [0,1] \right\}$$
$$C_2 = \left\{ \left(x^3, (1-x)^3 + 3(1-x)^2x\right) \mid x \in [0,1] \right\}$$

For $i = 1, 2$ let $B_i \subset [0,1]^2$ be the region bounded by $p_0 \ge 0,\ p_3 \ge 0,\ p_0 + p_3 \ge \frac{1}{4}$, and by C_i. We need to prove that

$$\Delta(\overline{K_3}, K_3) = B_1 \cup B_2.$$

By Theorems 3.1 and 3.2 $\Delta \subseteq B_1 \cup B_2$.

We show that every point in this domain can be approximated arbitrarily well by $(p_0(G), p_3(G))$ for arbitrarily large G. We define the following parameterized family of random graphs:

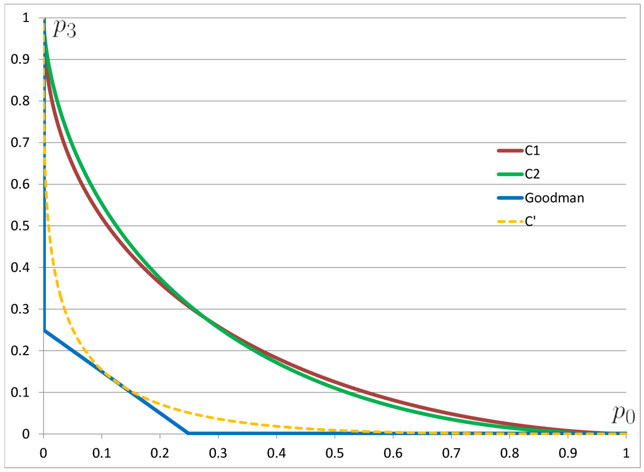

Figure 1: The curves that bound the region of $\Delta(\overline{K_3}, K_3)$, and the auxiliary curve C' used in the proof

Definition 3.3. *For every $x, a, b, c \in [0,1]$, $G_{x,a,b,c}$, is the class of random graphs (V,E), where $V = A\dot{\cup}B$ with $|A| = x|V|$ and $|B| = (1-x)|V|$. Adjacencies are chosen independently among pairs, with*

$$Pr(ij \in E) = \begin{cases} a & i,j \in A \\ b & i,j \in B \\ c & i \in A,\ j \in B \text{ or vice versa} \end{cases}$$

A simple computation shows that

$$\mathbb{E}p_0(G_{x,a,b,c}) = x^3(1-a)^3 + (1-x)^3(1-b)^3 + 3x^2(1-x)(1-a)(1-c)^2 + 3x(1-x)^2(1-b)(1-c)^2 + o(1)$$

and

$$\mathbb{E}p_3(G_{x,a,b,c}) = x^3a^3 + (1-x)^3b^3 + 3x^2(1-x)ac^2 + 3x(1-x)^2bc^2 + o(1)$$

By Lemma 2.1, $(\mathbb{E}p_0(G_{x,a,b,c}), \mathbb{E}p_3(G_{x,a,b,c})) \in \Delta(\overline{K_3}, K_3)$ for every $(x,a,b,c) \in [0,1]^4$. The following curve is used in the proof.

$$C' = \left\{ \left(t^3, (1-t)^3\right) \mid t \in [0,1] \right\}$$

Consider the following continuous map,

$$H : [0,1] \times [0,1] \to \Delta(\overline{K_3}, K_3)$$

$$H(x,a) = (\mathbb{E}[p_0(G_{x,a,1-a,1-a})], \mathbb{E}[p_3(G_{x,a,1-a,1-a})]) \,.$$

The following claims are immediate.

1. $H(x,0) = \left(x^3, (1-x)^3 + 3(1-x)^2x\right)$.
2. $H(x,1) = \left((1-x)^3 + 3(1-x)^2x, x^3\right)$
3. $H(1,a) = ((1-a)^3, a^3)$

4. $H(x, \frac{1}{2}) = (\frac{1}{8}, \frac{1}{8})$.
5. $H(0, a) = (a^3, (1-a)^3)$

$H \mid_{[0,1]\times[0,\frac{1}{2}]}$ is a continuous map from a topological 2-disc. The boundary of this disk is mapped to a path encircling $C_2 \cup C'$. Therefore, $C_2 \cup C'$ is contractible in $im(H)$, and consequently the region bounded by C_2 and C' is contained in $im(H)$, and also in $\Delta(\overline{K_3}, K_3)$. A similar argument for $H \mid_{[0,1]\times[\frac{1}{2},1]}$ shows that the region bounded by C_1 and C' is contained in $\Delta(\overline{K_3}, K_3)$.
The remaining area in $\Delta(\overline{K_3}, K_3)$ will be covered similarly. Consider the following continuous map,

$$H_1 : [0,1] \times [0,1] \to \Delta(\overline{K_3}, K_3)$$

$$H_1(x, a) = (\mathbb{E}[p_0(G_{x,a,a,1-a})], \mathbb{E}[p_3(G_{x,a,a,1-a})]) .$$

Again, the following claims are immediate.

1. $H_1(x, 0) = (x^3 + (1-x)^3, 0)$
2. $H_1(\frac{1}{2}, a) = \frac{1}{8}\left(1 - (2a-1)^3, 1 + (2a-1)^3\right)$
3. $H_1(x, 1) = (0, x^3 + (1-x)^3)$
4. $H_1(0, a) = ((1-a)^3, a^3)$

$H_1 \mid_{[0,\frac{1}{2}]\times[0,1]}$ is a continuous map from a topological 2-disc, mapping its boundary to a path encircling C', $[\frac{1}{4}, 1] \times \{0\}, \{\frac{t}{4}, \frac{1-t}{4} \mid t \in [0,1]\}$ and $\{0\} \times [\frac{1}{4}, 1]$. Therefore, as before, the region between these curves is contained in $\Delta(\overline{K_3}, K_3)$. Altogether, $B_1 \cup B_2 \subseteq \Delta(\overline{K_3}, K_3)$ is obtained.

□

4 Flag algebras - a dual perspective

Let $\mathcal{G}$ be a infinite family of graphs closed under taking induced subgraphs, let $\mathcal{H} = (H_1, ..., H_t)$ a collection of graphs. We formulate necessary conditions for membership in the set $\Delta_{\mathcal{G}}(\mathcal{H})$ which are stated in terms of feasibility of some SDP. This part is self-contained, and concentrates on the connections between the theory of flag algebras and standard arguments in discrete optimization.

Definition 4.1. *An* (s, k)-flagged graph $F = (H, U)$ *consists of an* s*-vertex graph* H *and a* flag $U = (u_1, ..., u_k)$, *an* ordered *set of* k *vertices in* H. *An isomorphism* $F \cong F'$ *between flagged graphs* $F = (H, U)$ *and* $F' = (H', U')$ *is a graph isomorphism*

$$\varphi : V(H) \longrightarrow V(H') \text{ such that } \varphi(u_i) = u'_i \quad \forall i.$$

Definition 4.2. *Let* G *be a graph and* F_1, F_2 *be* (s, k)*-flagged graphs. Choose uniformly at random two subsets* (V_1, V_2) *of* $V(G)$ *of size* s *with intersection* $U = V_1 \cap V_2$ *of cardinality* k *and choose random ordering of* U*. Define*

$$p(F_1, F_2; G) = \Pr\left[F_i \cong (G|_{V_i}, U) \quad i = 1, 2\right].$$

Associated with every list $F_1, ..., F_l$ of (s,k)-flagged graphs is the $l \times l$ matrix $A^G = A^G(F_1, ..., F_l)$

$$\forall i,j \quad A^G(F_1, ..., F_l)_{i,j} = p(F_i, F_j; G).$$

Note that A^G is a symmetric matrix.

Example 4.3. *Denote by e (resp. $\bar{e}$) the edge (its complement) with one flagged vertex. Also, P_3 denotes the path on 3 vertices. Then*

$$A^{P_3}(\bar{e}, e) = \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Proof. Let $V_1, V_2 \subset V(P_3)$ be chosen randomly with $|V_1| = |V_2| = 2$ and $|V_1 \cap V_2| = 1$. First, $p(\bar{e}, \bar{e}; P_3) = 0$ since either V_1 or V_2 spans an edge. Also, $p(e, e; P_3) = \frac{1}{3}$ since both sets spans an edge iff their common vertex has degree 2. Finally, $p(\bar{e}, e; P_3) = \frac{1}{3}$ since the common vertex has degree 1 with probability 2/3, and conditioned on that, the first set V_1 spans an edge with probability 1/2. $\square$

We denote by $PSD(l)$ the cone of $l \times l$ positive semi-definite matrices.

Theorem 4.4. *Let F_i, $i = 1, ..., l$, be (s,k)-flagged graphs. For an n-vertex graph G,*

$$dist\left(A^G(F_1, ..., F_l), PSD(l)\right) = O\left(\frac{1}{n}\right)$$

where dist stands for distance in l_2.

Corollary 4.5. *Let $\mathcal{G}$ be a class of graphs that is closed under taking induced subgraphs. Let $\mathcal{G}_n$ be the set of n-vertex members of $\mathcal{G}$. Let $\mathcal{H} = (H_1, ..., H_t)$ be a complete list of all the isomorphism types of graphs in $\mathcal{G}_r$. Let F_i, $i = 1, ..., l$ be (s,k)-flagged graphs. Then for every $(p_1, ..., p_t) \in \Delta_{\mathcal{G}}(\mathcal{H})$,*

$$\sum_{\alpha=1}^{t} p_\alpha \cdot A^{H_\alpha}(F_1, ..., F_l) \succeq 0\,.$$

Let us illustrate how this corollary helps us derive an upper bound on $\lim_{n\to\infty} \max_{G\in\mathcal{G}_n} d(H;G)$ for some fixed graph H. (The limit exists since $\max_{G\in\mathcal{G}_n} d(H;G)$ is a non-increasing function of n). Note that

$$d(H;G) = \sum_{\alpha=1}^{t} d(H;H_\alpha) d(H_\alpha;G).$$

Therefore the following SDP yields an upper bound

$$\max \quad \sum_{\alpha=1}^{t} d(H;H_\alpha) p_\alpha \quad s.t.$$

$$\text{all } p_\alpha \geq 0 \text{ and } \sum_{\alpha} p_\alpha = 1$$

$$\sum_{\alpha=1}^{t} p_\alpha \cdot A^{H_\alpha}(F_1, ..., F_l) \succeq 0$$

By SDP duality, this maximum can be also upper-bounded by

$$\min_{Q \in PSD(l)} \left(\max_{1 \le \alpha \le t} \left[d(H; H_\alpha) + \mathrm{Tr}(Q \cdot A^{H_\alpha}) \right] \right)$$

which is the more familiar form of SDP used in the literature on applications of flag algebras (see, e.g., [22, 3]). The proofs of the above two statements are based on standard arguments.

Theorem 4.4 $\implies$ *Corollary 4.5.* First we prove that for every graph G on at least r vertices,

$$A^G(F_1, ..., F_l) = \sum_{\alpha=1}^{t} d(H_\alpha; G) A^{H_\alpha}(F_1, ..., F_l).$$

Namely, that for every $1 \le i, j \le l$,

$$p(F_i, F_j; G) = \sum_{\alpha=1}^{t} d(H_\alpha; G) p(F_i, F_j; H_\alpha).$$

This is just an application of the law of total probability. On the LHS we sample uniformly two sets V_1, V_2 of size s with $|V_1 \cap V_2| = k$ from $V(G)$ together with a random ordering of $V_1 \cap V_2$, and on the RHS we first sample a random set V' of size r from $V(G)$, and then uniformly sample V_1, V_2 as above from V'. To finish the proof, let $\bar{p} = (p_1, ..., p_t) \in \Delta_{\mathcal{G}}(\mathcal{H})$. By the definition of $\Delta_{\mathcal{G}}(\mathcal{H})$ and Theorem 4.4, for every $\epsilon > 0$ there is a sufficiently large graph $G \in \mathcal{G}$ such that both

$$|p_\alpha - d(H_\alpha; G)| < \epsilon \quad \forall \alpha$$

and

$$\mathrm{dist}\left(\sum_{\alpha} d(H_\alpha; G) A^{H_\alpha}, PSD(l) \right) < \epsilon.$$

Therefore, $\mathrm{dist}\left(\sum_\alpha p_\alpha A^{H_\alpha}, PSD(l)\right) = 0$. □

Proof of Theorem 4.4. Let G be an n-vertex graph. Consider the following equivalent description of the underlying distribution in the definition of the matrix $A^G = A^G(F_1, ..., F_l)$. Choose uniformly at random an ordered set $U \subset V(G)$ of size k, two *disjoint* sets $S_1, S_2 \subset V(G) \setminus U$ of size $s-k$ and let $V_i = S_i \dot\cup U$, $i = 1, 2$. Thus $A^G_{i,j}$ is the probability that $F_i \cong (G|_{V_i}, U)$, for $i = 1, 2$. Note that for every fixed U, two sets $S_1, S_2 \subset V(G) \setminus U$ of size $s-k$ chosen uniformly and independently at random are disjoint with probability $1 - O(1/n)$. Therefore, it suffices to prove that the matrix B^G, defined exactly as A^G except that S_1, S_2 are chosen *independently*, is PSD.

Consider the matrix Q with l rows and $\frac{n!}{(n-k)!}$ columns indexed by ordered sets $U \subset V(G)$ of size k, defined as following. Choose a random subset $S \subset V(G) \setminus U$ of size $s-k$, and let $Q_{i,U} = \Pr[F_i \cong (G|_{S \cup U}, U)]$. Then,

$$B^G = \frac{(n-k)!}{n!} Q Q^T \succeq 0$$

□

5 Triangle-free graphs

In this section we prove Theorem 1.3, by showing that the set $\Delta_{\mathcal{TF}}(\overline{K_3}, \overline{P_3}, P_3)$ is characterized by the quadratic constraints deduced from the flag algebra theory.

Proof of Theorem 1.3. $(I) \implies (II)$. This implication is a direct application of Corollary 4.5. Let $\bar{e}$,e be (2,1)-flagged graphs. $\bar{e}$ (resp. e) is the empty (complete) graph over 2 vertices with one flagged vertex. By a straightforward computation (See example 4.3),

$$A^{\overline{K_3}}(\bar{e}, e) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A^{\overline{P_3}}(\bar{e}, e) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix}, \quad A^{P_3}(\bar{e}, e) = \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

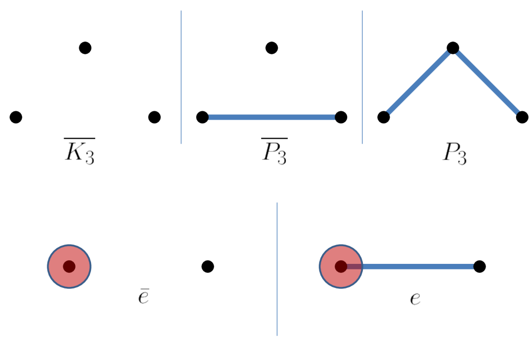

Figure 2: Triangle free graphs and flagged graphs used in the proof

Since these are all the graphs on 3 vertices in the family $\mathcal{TF}$, we may apply Corollary 4.5, and obtain (II).

$(II) \implies (III)$. Suppose p_0, p_1, p_2 satisfy the condition in (II). Since $p_0 + p_1 + p_2 = 1$, this can be reformulated as

$$\begin{pmatrix} 3p_0 + p_1 & 1 - p_0 \\ 1 - p_0 & 1 - p_0 - p_1 \end{pmatrix} \succeq 0,$$

which implies that

$$0 \leq (3p_0 + p_1)(1 - p_0 - p_1) - (1 - p_0)^2, \tag{1}$$

$$p_0 + p_1 \leq 1.$$

Recall Definition 3.3 of $G_{x,a,b,c}$, and denote $G_{\alpha,q} := G_{\alpha,0,0,q}$ a distribution on bipartite graphs, for $\alpha, q \in [0, 1]$. Then,

$$\mathbb{E}[p_0(G_{\alpha,q})] = 1 - 3\alpha(1 - \alpha)q(2 - q) + o(1).$$

$$\mathbb{E}[p_1(G_{\alpha,q})] = 6\alpha(1 - \alpha)q(1 - q) + o(1).$$

By Lemma 2.1, for every α, q, $(\mathbb{E}[p_0(G_{\alpha,q}]), \mathbb{E}[p_1(G_{\alpha,q})], \mathbb{E}[p_2(G_{\alpha,q})]) \in \Delta_{\mathcal{BP}}(\overline{K_3}, \overline{P_3}, P_3)$. Thus, it suffices, given p_0, p_1 that satisfy (1), to find $(\alpha, q) \in [0,1]^2$ such that,

$$p_0 = 1 - 3\alpha(1-\alpha)q(2-q), \text{ and } p_1 = 6\alpha(1-\alpha)q(1-q).$$

This implies that

$$q = \frac{2 - 2p_0 - 2p_1}{2 - 2p_0 - p_1} \in [0,1],$$

and

$$(1-2\alpha)^2 = \frac{(3p_0 + p_1)(1 - p_0 - p_1) - (1-p_0)^2}{3(1 - p_0 - p_1)}$$

Miraculously, $\alpha \in [0,1]$ that satisfies this equation exists iff the quadratic constraint in (1) are satisfied and $p_0 + p_1 < 1$. Indeed it is easy to check that in this case the right hand side is non-negative and is ≤ 1. On the other hand, if $p_0 + p_1 = 1$, then by (1) $p_0 = 1$ and this profile is attained for $q = 0$.

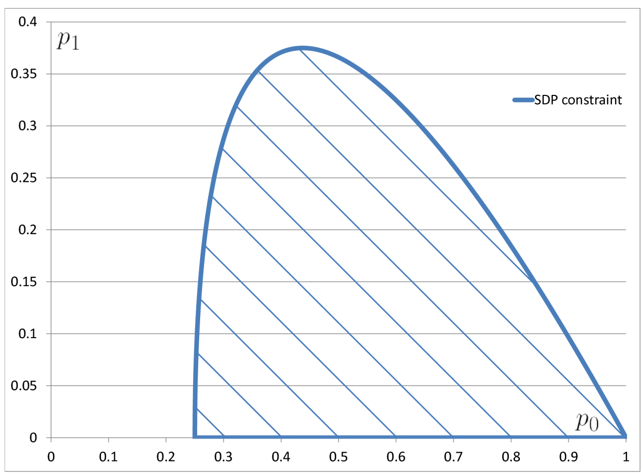

Figure 3: The region of possible p_0, p_1 of triangle-free graphs

$(III) \implies (I)$. Immediate, since every bipartite graph is triangle free.

□

6 Concluding remarks

In this paper we study the set $\mathcal{S}_3 \subset \mathbb{R}^4$ of all vectors $(p_0, ..., p_3)$ that are arbitrarily close to the 3-local profiles of arbitrarily large graphs. We show that the projection of this set to the (p_0, p_3) plane is completely realizable by the graphs that are generated by a model which partitions the vertices into two sets. We also show that the intersection of $\mathcal{S}_3$ with the plane $p_3 = 0$, i.e. triangle-free graphs, is completely realizable by a simple model of random bipartite graphs. We wonder how far these observations can be extended. Razborov's work [21] shows that certain 3 profiles require the use of k-partite models for arbitrarily large k. Also in general, it is not true that a k-local profile of every K_k-free graph can be realized by $(k-1)$-partite graph. Indeed, it was shown in [3], that already for

$k \geq 4$ the minimum density of empty sets of size k in K_k-free graphs is strictly smaller than what can be achieved by $(k-1)$-partite graphs.

It still remains a challenge to get a full description of the set $\mathcal{S}_3$. The analogous questions concerning r-profiles, $r > 3$ seems even more difficult. Even characterizing the profiles of (r-cliques, r-anticliques), which is solved here for $r = 3$, is still widely open.

References

[1] B. Bollobás, C. Nara, and S. Tachibana. The maximal number of induced complete bipartite graphs. *Discrete mathematics*, 62(3):271–275, 1986.

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

[3] S. Das, H. Huang, J. Ma, H. Naves, and B. Sudakov, A problem of Erdős on the minimum of k-cliques, submitted.

[4] P. Erdős, D.J. Kleitman, and B.L. Rothschild. Asymptotic enumeration of K_n-free graphs. *Theorie Combinatorie, Proceedings of the Conference, Vol. II, Rome, 1973, Roma, Acad. Nazionale dei Lincei*, pages 19–27, 1976.

[5] P. Erdos. On the number of complete subgraphs contained in certain graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 7:459–464, 1962.

[6] P. Erdös and A.H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc*, 52:1087–1091, 1946.

[7] G. Exoo. Dense packings of induced subgraphs. *Ars Combin*, 22:5–10, 1986.

[8] F. Franek and V. Rödl. 2-colorings of complete graphs with small number of monochromatic k_4 subgraphs. *Discrete mathematics*, 114(1):199–203, 1993.

[9] G. Giraud. Sur le probleme de goodman pour les quadrangles et la majoration des nombres de ramsey. *Journal of Combinatorial Theory, Series B*, 27(3):237–253, 1979.

[10] A.W. Goodman. On sets of acquaintances and strangers at any party. *American Mathematical Monthly*, pages 778–783, 1959.

[11] H. Hatami, J. Hirst, and S. Norine. The inducibility of blow-up graphs. *arXiv preprint arXiv:1108.5699*, 2011.

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

[13] J. Hirst. The inducibility of graphs on four vertices. *arXiv preprint arXiv:1109.1592*, 2011.

[14] H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov. On the densities of cliques and independent sets in graphs, submitted.

[15] G. Katona. A theorem of finite sets. *Theory of Graphs, Akadémia Kiadó, Budapest*, pages 187–207, 1968.

[16] J. Kruskal. The number of simplicies in a complex. *Mathematical Optimization Techniques, Univ. of California Press*, pages 251–278, 1963.

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

[18] O. Pikhurko, Minimum number of k-cliques in graphs with bounded independence number, submitted.

[19] N. Pippenger and M.C. Golumbic. The inducibility of graphs. *Journal of Combinatorial Theory, Series B*, 19(3):189–203, 1975.

[20] A.A. Razborov. Flag algebras. *Journal of Symbolic Logic*, 72(4):1239–1282, 2007.

[21] A.A. Razborov. On the minimal density of triangles in graphs. *Combinatorics Probability and Computing*, 17(4):603–618, 2008.

[22] A.A. Razborov. On 3-hypergraphs with forbidden 4-vertex configurations. *SIAM Journal on Discrete Mathematics*, 24(3):946–963, 2010.

[23] C. Reiher. Minimizing the number of cliques in graphs of given order and edge density, manuscript.

[24] K. Sperfeld. On the minimal monochromatic k4-density. *arXiv preprint arXiv:1106.1030*, 2011.

[25] A. Thomason. A disproof of a conjecture of erdős in ramsey theory. *Journal of the London Mathematical Society*, 2(2):246–255, 1989.