The Erdős Pentagon Problem

by

Kris Siy

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The Erdős pentagon problem asks about the maximum number of copies of C_5 that one can find in a triangle-free graph. This problem was posed in 1984, but was not resolved until 2012. In this thesis, we aim to capture the story of solving the pentagon problem as completely as possible. We provide a detailed exposition of the motivation behind and construction of flag algebras, and give a complete solution to the pentagon problem using flag algebra techniques. All known solutions of the pentagon problem to date have been computer-assisted; with the question of whether an elementary solution can be found in mind, we also investigate the pentagon problem and modifications of the problem from different elementary perspectives.

Acknowledgements

First and foremost, I would like to thank my supervisor Penny Haxell for being willing to take me on as a student. Her guidance and mentorship over the course of this project has made me a stronger mathematical thinker and writer, and this thesis would not exist without her patience and encouragement.

I would also like to thank my readers, Peter Nelson and Kevin Purbhoo, for taking the time to read and provide insightful comments on my thesis.

Finally, I would like to thank all of my friends and family who have shown unwavering support and encouragement throughout this endeavour.

Table of Contents

Chapter 1

Introduction

1.1 The Erdős Pentagon Problem

One of the first results in extremal graph theory was given as the solution to a problem posed by Mantel in the journal *Wiskundige Opgaven* in 1907 [24]: "Given are some points, no four of which lie on the same plane. At most how many lines can one draw between the points without forming triangles?"

This is formulated as a geometric problem (as many early graph-theoretic problems were), and the restriction of having no four points on the same plane is to ensure that any line drawn between two points will only pass through those two points. This naturally leads to a graph-theoretic formulation, though, and in the language of graph theory, one can equivalently ask about the maximum number of edges a graph without triangles can have.

Extremal graph theory has grown since then to more generally study maximal or minimal graphs satisfying certain properties, or equivalently, the idea that graphs must carry certain types of structural properties beyond some maximal or minimal point. For example, Mantel's question concerns graphs which carry the property of being triangle-free, and asks about the maximal such graph with respect to the number of edges in the graph. Equivalently, this asks us to find the best possible lower bound for the number of edges in a graph such that it must contain a triangle. Perhaps the most important generalisation of this was proved by Erdős and Stone in 1946 [9], which gave an asymptotic upper bound for the number of edges in a graph that does not contain a given subgraph H, for any non-bipartite graph H.

Along the same lines as these questions, Erdős asked the following question in a 1984 paper [7], which we will henceforth refer to as the *Erdős pentagon problem*:

"Is it true that a graph on $5n$ vertices with no triangle contains at most n^5 pentagons?"

A *pentagon* here is defined to be a copy of C_5. He noted in that paper that if this is true, this bound is tight: One can consider a blow-up of C_5 obtained by replacing each vertex with an independent set of n vertices and replacing each edge with a complete bipartite graph $K_{n,n}$. Note that this graph is triangle-free. Then, by selecting one vertex in each independent set, one can find n^5 pentagons in this graph.

The pentagon problem is simple to state - Erdős took only one sentence to ask the question and one additional sentence to discuss the tightness of this bound - and could easily be explained to someone taking a first course in graph theory. Yet, despite looking very much like other problems in extremal graph theory (such as Mantel's theorem) which were resolved with elegant proofs, the pentagon problem remained open for nearly 30 years after it was first published.

Several attempts got close to solving the problem along the way: In 1989, Győri [14] showed that a triangle-free graph with n vertices contains no more than $c(\frac{n+1}{5})^5$ copies of C_5, where $c < 1.03$. Füredi was later able to refine this approach and improve this bound by showing the same result for $c < 1.001$ (personal communication; see, e.g., [15]).

No further progress was made until 2012, when Grzesik [11] and Hatami et al. [15] independently resolved the conjecture asymptotically in the affirmative using flag algebras. Additionally, Hatami et al. showed that for all n divisible by 5, a triangle-free graph with n vertices maximising the number of 5-cycles in the graph must be a balanced blow-up of C_5, where each vertex of a C_5 is replaced with an independent set of size $\frac{n}{5}$ (see Theorem 48). In 2017, Lidický and Pfender [21] extended these results using flag algebra and graphon methods to show that with the exception of one sporadic graph where $n = 8$, for all $n \geq 5$, a triangle-free graph with n vertices maximising the number of 5-cycles in the graph must be an "almost balanced blow-up" of a C_5, where each vertex of a C_5 is replaced with an independent set of size $\lfloor \frac{n}{5} \rfloor$ or $\lceil \frac{n}{5} \rceil$ (see Theorem 49). Noting that all previous approaches to resolve the pentagon problem were assisted by a computer, a 2018 paper by Grzesik and Kielak [13] gave a non-computer-assisted proof of a modification of the pentagon problem (see Theorem 50) - notably, though, this did not yield an elementary solution to the pentagon problem.

In this thesis, we aim to capture as fully as possible the story of solving the pentagon problem. In the remainder of Chapter 1, we will provide some preliminaries for the remainder of the exposition. In Chapter 2, we will provide several perspectives from which one can examine the pentagon problem, including Győri's approach, elements of Füredi's approach, and an algebraic approach to the problem. In Chapter 3, we will provide a detailed exposition of the motivation behind and the construction of flag algebras, and show how flag algebra techniques can be applied to asymptotically resolve the pentagon problem. Finally, in Chapter 4, we will survey further results on the pentagon problem and a modification of the pentagon problem.

1.2 Preliminaries

In this section, we will establish some mathematical preliminaries that will be used throughout the rest of our exposition.

We will assume that the reader is familiar with elementary graph theory (for a reference, see [6]). We establish some graph-theoretic and notational conventions that we will use throughout our exposition:

- If S is a set, we will use $|S|$ to denote the size of S. For $k \in \mathbb{N}$, by the set $[k]$ we mean the set $\{1, 2, 3, \cdots, k\}$. By convention, let $[0]$ be the empty set.
- We will assume that all graphs are finite and simple (no loops or multiedges).
- If G is a graph, we will let $V(G)$ be the set of vertices of G and $E(G)$ be the set of edges of G.
- Later in our exposition, we will need to work with labelled graphs. To deal with this formally, we can let a graph be defined as $G = (V, E, k, \theta)$, where $k \in \mathbb{N} \cup \{0\}$ and $\theta : [k] \to V(G)$ is an injective map. When $k = 0$, we will call G an *unlabelled graph* and normally write $G = (V, E)$; otherwise we will call G a *labelled graph*.
- If $G = (V(G), E(G), k_G, \theta_G)$ and $H = (V(H), E(H), k_H, \theta_H)$ are graphs, they are said to be *isomorphic* if $k_G = k_H$ and there exists a bijection $\rho : V(G) \to V(H)$ such that $vw \in E(G)$ iff $f(v)f(w) \in E(H)$ for every $v, w \in V(G)$, and $\rho(\theta_G(i)) = \theta_H(i)$ for every $i \in [k]$. We will denote graph isomorphism by writing $G \cong H$.
- If $v \in V(G)$, we will denote the degree of v by $\deg(v)$ (or $\deg_G(v)$, when the former is ambiguous) and the set of vertices adjacent to v by $N(v)$.
- If $v, w \in V(G)$ and there exists a v, w-path, we will denote the minimal distance between v and w in G by $d(v, w)$.
- By an *induced subgraph* of G, we mean a subset $S \subseteq V(G)$ together with all edges $e \in E(G)$ where both endpoints of e are in S. Such a subgraph will be denoted $G[S]$.

- If G and H are graphs and $S \subseteq V(G)$, let $c(H;G)$ be the number of times H appears as an induced subgraph of G, and let $c(H;G,S)$ be the number of times H appears as an induced subgraph of G containing every vertex in S. That is, $c(H;G)$ is the number of ways we could select $|V(H)|$ vertices (without respect to order) from $V(G)$ so that the resulting induced subgraph on these vertices is isomorphic to H, and $c(H;G,S)$ is the number of ways we could select $|V(H)| - |V(S)|$ vertices (without respect to order) from $V(G)$ so that the resulting induced subgraph on these vertices and S is isomorphic to H. Furthermore, observe that in a triangle-free graph, every copy of C_5 is induced, so that if G is triangle-free, $c(C_5;G)$ is the number of copies of C_5 in G.
- If G is a graph with vertices $v_1, \cdots, v_k$, the *blow-up* of G obtained by replacing v_i with an independent set of n_i vertices for all $1 \le i \le k$ and maintaining adjacencies is denoted by $G^{(n_1,\cdots,n_k)}$. If $n_1 = \cdots = n_k = n$, then this is simply denoted $G^{(n)}$.

We will also assume the reader is familiar with elementary linear algebra. Later in our exposition (for Theorem 41), we will need one fact about positive semidefinite matrices. A $n \times n$ real symmetric matrix M is said to be *positive semidefinite* if $v^T M v \ge 0$ for all vectors $v \in \mathbb{R}^n$. Also, recall that $v \ne 0$ is an eigenvector of M with eigenvalue λ if $Mv = \lambda v$. We will prove an equivalent characterisation of positive semidefinite matrices:

Theorem 1. Let M be a real symmetric $n \times n$ matrix. The following are equivalent:

(a) $v^T M v \ge 0$ for all $v \in \mathbb{R}^n$.

(b) All the eigenvalues of M are nonnegative.

(c) There exists a matrix B such that $M = BB^T$.

Proof. (a) $\Rightarrow$ (b): Let λ be an eigenvalue of M. By definition, there exists an eigenvector $v \in \mathbb{R}^n$ such that $Mv = \lambda v$. Left-multiplying by v^T, it follows that $v^T M v = \lambda v^T v$. Since $v^T M v \ge 0$ for every $v \in \mathbb{R}^n$ and $v^T v \ge 0$ for all v as the square of any real number is nonnegative, it follows that λ is nonnegative.

(b) $\Rightarrow$ (c): Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of M with multiplicity considered. As M is symmetric, it follows that it has a Jordan decomposition $M = \sum_{i=1}^{n} \lambda_i x_i x_i^T$. Now, for all $1 \le i \le n$, let $y_i = \sqrt{\lambda_i} x_i$; this is well-defined as every λ_i is nonnegative. Rewriting, we get that $M = \sum_{i=1}^{n} y_i y_i^T$. If we let A be the matrix defined by letting y_i be the ith column of A, it follows from the definition of matrix multiplication that $A^T A = M$, and letting $B = A^T$ we get the desired result.

(c) $\Rightarrow$ (a): Let $v \in \mathbb{R}^n$ and suppose that there exists a B such that $M = B^T B$. Let $A = B^T$, so that $M = A^T A$. Then $v^T M v = v^T A^T A v = (Av)^T (Av) \geq 0$ as the square of any real number is nonnegative, as desired. $\square$

We hope that the remainder of this exposition will be largely self-contained and accessible to someone with a background in elementary graph theory, probability, linear algebra, and abstract algebra. That being said, several of our proofs will make use of miscellaneous theorems which are not otherwise covered in this exposition. We will cover these briefly now:

Theorem 2 (AM-GM Inequality)**.** Let $x_1, x_2, \cdots, x_n$ be non-negative real numbers. Then the arithmetic mean of $x_1, x_2, \cdots, x_n$ is greater than or equal to the geometric mean of $x_1, x_2, \cdots, x_n$; namely, the following inequality holds:

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}.$$

Furthermore, equality only holds when $x_1 = x_2 = \cdots = x_n$.

Theorem 3 (Tychonoff's Theorem)**.** Let I be an index set, and let $\{X_i\}_{i \in I}$ be an indexed family of non-empty topological spaces. Let $X = \prod_{i \in I} X_i$ be the corresponding product space endowed with the product topology. Then X is compact if and only if every X_i is compact.

Theorem 4 (Chebyshev's Inequality)**.** Let X be a random variable with a finite expected value μ and finite non-zero variance σ^2. Then if $k > 0$ is a positive real number,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Theorem 5 (Borel-Cantelli Lemma)**.** Let $\{A_i\}_{i=1}^{\infty}$ be a sequence of events in some probability space. If the sum of the probabilities of these events is finite, that is, if

$$\sum_{i=1}^{\infty} P(A_i) < \infty,$$

then the probability that infinitely many of these events occur is zero. In other words,

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{i>n}^{\infty}A_i\right)=0.$$

Theorem 2 will be used in several proofs throughout this exposition. Theorem 3 will be used in the proof of Theorem 27, and Theorems 4 and 5 will be used in the proof of Theorem 34.

Chapter 2

Perspectives on the Pentagon Problem

2.1 Győri's Elementary Upper Bound

The first published approach to the Erdős pentagon problem was given by Győri in 1989; in this section, we will retrace the approach given in his paper [14]. Győri managed to use elementary methods to find an upper bound on the number of copies of C_5 in an n-vertex triangle-free graph which got very close to, but didn't quite reach, Erdős's conjectured bound. More precisely, he proved the following theorem:

Theorem 6. A triangle-free graph on n vertices cannot have more than $c(\frac{n+1}{5})^5$ C_5's where $c = \frac{3^3 \cdot 5^4}{2^{14}} < 1.03$.

To begin, we establish the following numerical lemma. (We remark that we do not need the conclusion about the equality case in our proof of Theorem 6, but it is immediate from the proof.)

Lemma 7. Suppose that $\gamma, \alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n \in \mathbb{R}^+$ and that $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i = M$. Then $\min\limits_{1 \le i \le n} \alpha_i(\gamma - \beta_i) \le \dfrac{M}{n}(e - \dfrac{M}{n})$. Moreover, equality holds if and only if $\alpha_i = \beta_i = \frac{M}{n}$ for all $1 \le i \le n$.

Proof. Suppose that $\alpha_i(\gamma - \beta_i) \ge \frac{M}{n}(\gamma - \frac{M}{n})$ for all $1 \le i \le n$. By taking the product of these n inequalities and taking the nth root of both sides, we get that

$$\prod_{i=1}^{n} \alpha_i^{1/n}(\gamma - \beta_i)^{1/n} \geq \frac{M}{n}(\gamma - \frac{M}{n}).$$

However, the AM-GM inequality (Theorem 2) tells us that

$$\frac{M}{n} = \frac{\sum_{i=1}^{n} \alpha_i}{n} \geq \left(\prod_{i=1}^{n} \alpha_i\right)^{1/n}$$

and

$$\frac{M}{n} = \frac{\sum_{i=1}^{n}(\gamma - \beta_i)}{n} \geq \left(\prod_{i=1}^{n}(\gamma - \beta_i)\right)^{1/n},$$

so the equality conditions of the AM-GM inequality tell us that $\alpha_i(\gamma - \beta_i) \geq \frac{M}{n}(\gamma - \frac{M}{n})$ for all $1 \leq i \leq n$ if and only if $\alpha_i = \beta_i = \frac{M}{n}$ for all i, as desired. □

This numerical lemma can help us establish Győri's theorem:

Proof of Theorem 6. We proceed by induction on n. For $n = 1$, we have that $c(\frac{2}{5})^5 < 1$, so the statement is clearly true. Now, suppose that $|V(G)| = n$. Let $v \in V(G)$; first, we show that we can find an upper bound on $c(C_5; G, \{v\})$:

Claim. $c(C_5; G, \{v\}) \leq \dfrac{\deg(v)^2}{4} \cdot \left(|E(G)| - \displaystyle\sum_{uv \in E(G)} \deg(u)\right).$

Proof. To pick a C_5 containing v, it suffices to pick the edge opposite to v in the C_5 (call this edge $f = xy$), then two vertices, one adjacent to both v and x, and the other adjacent to both v and y.

If $u_1, u_2 \in N(v)$, then u_1u_2 cannot be an edge of G, or else v, u_1, u_2 would form a triangle; thus, the set of edges incident to u_1 and the set of edges incident to u_2 must be disjoint. The edge opposite to v in a C_5 must only be incident to vertices at distance two or more away from v, so no edge incident to any neighbour of v can be f; thus there are at most $|E(G)| - \displaystyle\sum_{uv \in E(G)} \deg(u)$ ways to choose f.

Observe that for any $f = xy$, if $u \in N(v)$, then ux and uy cannot both be edges, otherwise u, x, y would form a triangle. Thus, if we can choose the vertex incident to both v and x in

k ways, we can choose the vertex incident to both v and y in at most $\deg(v)-k$ ways, so we can choose the remaining two vertices of the C_5 in at most $k(\deg(v)-k) \le \frac{\deg(v)^2}{4}$ ways and thus we can choose a C_5 containing v in at most $\frac{\deg(v)^2}{4} \cdot \left(|E(G)| - \sum_{uv \in E(G)} \deg(u)\right)$ ways. $\square$

Now, let $M = \sum_{v \in V(G)} \deg(v)^2$. Observe that by counting $\deg(v)$ once for each of its neighbours, we also have that $M = \sum_{v \in V(G)} \sum_{uv \in E(G)} \deg(u)$. By the Cauchy-Schwarz inequality,

$$\left(\sum_{v \in V(G)} \deg(v)\right)^2 \le \sum_{v \in V(G)} \deg(v)^2 \sum_{v \in V(G)} 1^2 = Mn,$$

so by the Handshaking Lemma, $|E(G)| \le \frac{\sqrt{Mn}}{2}$. By letting $\alpha_v = \deg(v)^2$ and $\beta_v = \sum_{uv \in E(G)} \deg(u)$ for each $v \in V(G)$ and letting $\gamma = |E(G)|$, by Lemma 7, it follows that there must exist a vertex v so that $c(C_5; G, \{v\}) \le \frac{M}{4n}(\frac{\sqrt{Mn}}{2} - \frac{M}{n})$.

When n is fixed, by taking derivatives, the right-hand side of this inequality is maximised when $M = \frac{9n^3}{64}$, and so by substituting, it follows that $c(C_5; G, \{v\}) \le \frac{3^3 n^4}{2^{14}}$. By the induction hypothesis, we also have that $c(C_5; G - v) \le \frac{3^3}{5 \cdot 2^{14}} n^5$. Thus, G has at most

$$\frac{3^3}{5 \cdot 2^{14}} n^5 + \frac{3^3}{2^{14}} n^4 \le \frac{3^3}{5 \cdot 2^{14}} (n+1)^5$$

C_5's, and so the statement is true for all n, as desired. $\square$

Surprisingly, this argument is mostly tight. For $G = C_5^{(n/5)}$, a balanced blow-up of C_5 with n vertices, each vertex $v \in V(G)$ satisfies $\deg(v) = \frac{2n}{5}$, and $|E(G)| = \frac{n^2}{5}$. Thus, for every vertex $v \in V(G)$,

$$\frac{\deg(v)^2}{4} \cdot \left(|E(G)| - \sum_{uv \in E(G)} \deg(u)\right) = \frac{n^2}{25} \cdot \frac{n^2}{25} = \frac{n^4}{625}.$$

But $M = \sum_{v \in V(G)} \deg(v)^2 = \frac{4n^3}{25}$, so

$$\frac{M}{4n}\left(|E(G)| - \frac{M}{n}\right) = \frac{n^2}{25} \cdot \frac{n^2}{25} = \frac{n^4}{625}.$$

Since $C_5^{(n/5)}$ is regular, the equality conditions on the AM-GM and Cauchy-Schwarz inequalities hold.

The only slackness is in taking derivatives: We showed that the expression $\frac{M}{4n}(\frac{\sqrt{Mn}}{2} - \frac{M}{n})$ is maximised for fixed n when $M = \frac{9n^3}{64}$, but in $C_5^{(n/5)}$, we have $M = \frac{4n^3}{25}$. Because we cannot establish bounds on M for graphs with n vertices maximising $c(C_5; G)$, this argument is not tight as is.

2.2 An Algebraic Perspective

In this section, we will detail a different elementary perspective from which someone might look at the Erdős pentagon problem [16]. The pentagon problem asks about maximising the number of pentagons over all triangle-free graphs, but it seems difficult to do this at first glance, since it's not clear what a triangle-free graph looks like in general.

Hence, we can think instead about how we might transform this problem to work over a different space that we might understand more easily. In particular, there is a very nice correspondence between graphs and certain types of matrices, and having the tools of linear algebra at our disposal could be useful. If G is a graph with n vertices, we can define an $n \times n$ matrix $A(G)$, called the *adjacency matrix* of G, through labelling the vertices of G $v_1, \cdots, v_n$, and then letting

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is adjacent to } v_j; \\ 0 & \text{if } v_i \text{ is not adjacent to } v_j. \end{cases}$$

Note that by construction, $A(G)$ is symmetric.

It turns out that we can relate the pentagon problem optimising over triangle-free graphs to another problem optimising over the eigenvalues of the adjacency matrices corresponding to triangle-free graphs, as follows: Let $A = (A_{ij})$ be an $n \times n$ matrix. Recall that $v \neq 0$ is an eigenvector of A with eigenvalue λ if $Av = \lambda v$. By induction, we must have that λ^n is an eigenvalue of A^n for all n. Recall also that the trace of A is $tr(A) = \sum_{i=1}^{n} A_{ii}$. In fact, the sum of the eigenvalues of A is the trace of A:

Lemma 8. If A has eigenvalues $\lambda_1, \cdots, \lambda_n$, then $tr(A) = \sum_{i=1}^{n} \lambda_i$.

Proof. Suppose that A had eigenvalues $\lambda_1, \cdots, \lambda_n$. Recall that the characteristic polynomial of A is $p(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$, which has the eigenvalues of A as its zeroes. Thus $p(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$, so by expanding and equating coefficients, it follows that $c_{n-1} = -(\lambda_1 + \cdots + \lambda_n)$.

On the other hand, using the Leibniz formula to compute the determinant of $B = \lambda I - A$, the only permutation $\sigma \in S_n$ which will yield a λ^{n-1} term in the product $\displaystyle\prod_{i=1}^{n} B_{\sigma(i),i}$ is the identity permutation, so $p(\lambda) = (\lambda - A_{11})\cdots(\lambda - A_{nn}) - q(\lambda)$ for some degree $n-2$ polynomial $q(\lambda)$. Hence $c_{n-1} = -(A_{11} + \cdots + A_{nn}) = -(tr(A))$, and so $tr(A) = \lambda_1 + \cdots + \lambda_n$, as desired. □

With this in mind, we can use the eigenvalues of an adjacency matrix to count the number of closed walks of length k in a graph, as follows:

Lemma 9. If A is the adjacency matrix of a graph G, then A_{ij}^k is the number of walks of length k starting at v_i and ending at v_j.

Proof. We prove this by induction. For $k = 1$, this is true by the definition of an adjacency matrix. Consider that a walk of length k from vertex i to vertex j could have any neighbour of v_j as the penultimate vertex in the walk, so by induction, the number of such walks is $\displaystyle\sum_{l=1}^{n} A_{il}^{k-1} A_{lj}$. But this is just A_{ij}^k by the definition of matrix multiplication. □

In particular, it follows that A_{ii}^k is the number of closed walks of length k from vertex i to itself. Putting Lemmas 8 and 9 together, it follows that if G is a simple, triangle-free graph with eigenvalues $\lambda_1, \cdots, \lambda_n$, we can make the following conclusions:

- $\lambda_1 + \cdots + \lambda_n = 0$. This follows as the sum of the eigenvalues of a matrix is the trace of the matrix, and if G is simple, then all diagonal entries in the corresponding adjacency matrix will be zero.
- $\lambda_1^2 + \cdots + \lambda_n^2 = 2|E(G)|$. This follows as the sum of the squares of the eigenvalues of a matrix A is equal to the trace of A^2, and if G is a graph, then each diagonal entry in the corresponding adjacency matrix will count the number of closed walks

of length 2 in G. Thus, if G is simple, the trace of the adjacency matrix will count each edge exactly twice.

- $\lambda_1^3 + \cdots + \lambda_n^3 = 0$. This follows as the sum of the cubes of the eigenvalues of a matrix A is equal to the trace of A^3, and if G is a graph, then each diagonal entry in the corresponding adjacency matrix will count the number of closed walks of length 3 in G. Thus, if G is simple, the trace of the adjacency matrix will count each triangle exactly 6 times, but since G is triangle-free, this number is zero.

- Recalling that $c(C_5; G)$ is the number of copies of C_5 in G, we see that $\lambda_1^5 + \cdots + \lambda_n^5 = 10c(C_5; G)$. This follows as the sum of the fifth powers of the eigenvalues of a matrix A is equal to the trace of A^5, and if G is a graph, then each diagonal entry in the corresponding adjacency matrix will count the number of closed walks of length 5 in G. A closed walk of length 5 in a triangle-free graph must be a cycle of length 5, and each of these cycles correspond to ten different closed walks of length 5, starting at five possible vertices and moving in one of two possible directions. Hence, $tr(A^5) = 10 \cdot c(C_5; G)$.

Thus, maximising $c(C_5; G)$ over all triangle-free graphs is equivalent to maximising $tr(A^5)$ over all A which are both adjacency matrices of a simple graph and satisfy $tr(A^3) = 0$. It follows Erdős's conjecture is equivalent to the following statement:

> If A is the adjacency matrix of a simple graph G for which $tr(A^3) = 0$, then $tr(A^5) \leq \frac{1}{10}\left(\frac{n}{5}\right)^5$.

Erdős proposed that this was maximised by the balanced-blow up of a pentagon, and indeed, we can verify that the sum of the fifth powers of the eigenvalues of the adjacency matrix for $C_5^{(n/5)}$ is indeed $10 \cdot \left(\frac{n}{5}\right)^5$ by directly computing the eigenvalues as follows: Suppose that $\mathbb{F}$ is a field and $A = (A_{ij}) \in M_{m,n}(\mathbb{F})$, $B = (B_{ij}) \in M_{p,q}(\mathbb{F})$. Then the *Kronecker product* of A and B, denoted $A \otimes B$, is the block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \in M_{mp,nq}(\mathbb{F}).$$

One can verify the following lemma by the definition of the Kronecker product:

Lemma 10. If $A \in M_{m,n}(\mathbb{F}), B \in M_{p,q}(\mathbb{F}), C \in M_{n,k}(\mathbb{F}), D \in M_{q,j}(\mathbb{F})$, then $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

This allows us to characterise the eigenvalues and eigenvectors of the Kronecker product of two square matrices in terms of the eigenvalues and eigenvectors of each of these matrices:

Lemma 11. Suppose that $A \in M_{n,n}(\mathbb{F})$ and $B \in M_{t,t}(\mathbb{F})$, and that A has eigenvalues $\{\lambda_i \mid 1 \leq i \leq n\}_{i=1}^{n}$ and B has eigenvalues $\{\mu_j \mid 1 \leq j \leq t\}$. Then the eigenvalues of $A \otimes B$ are $\{\lambda_i \mu_j \mid 1 \leq i \leq n, 1 \leq j \leq t\}$.

Proof. Suppose that λ_i is an eigenvalue of A with eigenvector x and μ_j is an eigenvalue of B with eigenvector y, so that $Ax = \lambda_i x$ and $By = \mu_j y$. Then by Lemma 10 we must have that $(A \otimes B)(x \otimes y) = (Ax) \otimes (By) = (\lambda_i x) \otimes (\mu_j y) = \lambda_i \mu_j (x \otimes y)$, so it follows that $\lambda_i \mu_j$ is an eigenvalue of $A \otimes B$ with eigenvector $x \otimes y$. □

It turns out that the adjacency matrix of the balanced blow-up of a graph can be written as the Kronecker product of the adjacency matrix of the graph and the all-ones matrix, so its eigenvalues can be easily characterised:

Lemma 12. Suppose that G is a graph where $A(G)$ has eigenvalues $\lambda_1, \cdots, \lambda_n$. Then the eigenvalues of $G^{(t)}$ are $t\lambda_1, \cdots, t\lambda_n$, and 0 with multiplicity $n(t-1)$.

Proof. Suppose that G is a graph on n vertices. Let J_t the $t \times t$ all-ones matrix. Firstly, we claim that $A(G^{(t)}) = A(G) \otimes J_t$: Consider the corresponding graph to the adjacency matrix $A(G) \otimes J_t$. By considering the columns $k + in$ of this matrix for a fixed $1 \leq k \leq n$ and all $0 \leq i \leq n-1$, we see that the respective vertices of the graph form an independent set with the same adjacencies. Furthermore, each vertex in this independent set is adjacent to every vertex in every independent set induced by the neighbours of vertex k in G. Thus, this is the adjacency matrix of $G^{(t)}$, as desired. Since the eigenvalues of J_t are t with multiplicity 1 and 0 with multiplicity $t-1$, the result follows from Lemma 11. □

A simple computation gives us that the eigenvalues of C_5 are 2 (with multiplicity 1), $\frac{\sqrt{5}-1}{2}$ (with multiplicity 2), and $\frac{-\sqrt{5}-1}{2}$ (with multiplicity 2). Thus, by Lemma 12, the eigenvalues of $C_5^{(t)}$ are $2t$ (with multiplicity 1), $t \cdot \frac{\sqrt{5}-1}{2}$ (with multiplicity 2), $t \cdot \frac{-\sqrt{5}-1}{2}$ (with multiplicity 2), and 0 (with multiplicity $5(t-1)$). In particular, if $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of

$C_5^{(n/5)}$, we can now verify that $\lambda_1^5 + \cdots + \lambda_n^5 = 10c(C_5; C_5^{(n/5)}) = 10\left(\frac{n}{5}\right)^5$: Indeed,

$$\begin{aligned}\lambda_1^5 + \cdots + \lambda_n^5 &= \left(2 \cdot \frac{n}{5}\right)^5 + \left(\frac{n}{5} \cdot \frac{\sqrt{5}-1}{2}\right)^5 + \left(\frac{n}{5} \cdot \frac{-\sqrt{5}-1}{2}\right)^5 \\ &= \left(\frac{n}{5}\right)^5 \cdot \left[2^5 + \left(\frac{\sqrt{5}-1}{2}\right)^5 + \left(\frac{-\sqrt{5}-1}{2}\right)^5\right] \\ &= \left(\frac{n}{5}\right)^5 \cdot 10,\end{aligned}$$

as desired.

2.3 Structure in Extremal Graphs

Define a *pentagon-extremal graph* to be a triangle-free graph on n vertices containing the maximum number of C_5's for an n-vertex, triangle-free graph. Another perspective one might take to approach the Erdős pentagon problem is that of trying to show that every pentagon-extremal graph must indeed be the balanced blow-up of C_5, so that the bound proposed by Erdős holds. The graph $C_5^{(n/5)}$ has a number of nice properties, and we can show that some of these properties must also hold in any pentagon-extremal graph.

With this in mind, Füredi observed that one can use Zykov symmetrisation to show that in any pentagon-extremal graph, every vertex is contained in roughly the same number of C_5's, which is one property of a balanced blow-up of C_5:

Proposition 13. Let G be a pentagon-extremal graph on n vertices, and let $x, y \in V(G)$. Then $|c(C_5; G, \{x\}) - c(C_5; G, \{y\})| \leq c(C_5; G, \{x, y\}) < n^3$.

Proof. If $c(C_5; G, \{x\}) = c(C_5; G, \{y\})$, then the conclusion is clearly true. Thus, without loss of generality, suppose that $c(C_5; G, \{x\}) > c(C_5; G, \{y\})$. Construct G' from G by deleting all edges incident to y. Then note that $c(C_5; G') = c(C_5; G) - c(C_5; G, \{y\})$, since all C_5's in G are still present except those containing y, and $c(C_5; G', \{x\}) = c(C_5; G, \{x\}) - c(C_5; G, \{x, y\})$, since all C_5's in G containing x are still present except those containing y.

Now, construct G'' from G' by making y a clone of x through adding edges from y to every vertex in $N_{G'}(x)$, so that $N(x) = N(y)$. Counting the number of C_5's in G'', we get that $c(C_5; G'') = c(C_5; G') + c(C_5; G', \{x\}) = c(C_5; G) - c(C_5; G, \{y\}) + c(C_5; G, \{x\}) -$

$c(C_5; G, \{x, y\})$, as every C_5 in G' is present and y is now a clone of x. Since G is pentagon-extremal, we know that $c(C_5; G'') \leq c(C_5; G)$, so therefore $c(C_5; G, \{x\}) - c(C_5; G, \{y\}) \leq c(C_5; G, \{x, y\}) < n^3$, as if two vertices of a C_5 are fixed, the remaining three vertices of the C_5 can be chosen in at most $\binom{n}{3}$ ways. $\square$

As a consequence of Füredi's observation, we can actually deduce something more about the structure of one of the pentagon-extremal graphs, namely that two vertices are not contained in any copy of C_5 only if their neighbourhoods are identical:

Proposition 14. There exists a pentagon-extremal graph G on n vertices such that for all $x, y \in V(G)$, $N_G(x) = N_G(y)$ iff $c(C_5; G, \{x, y\}) = 0$.

Proof. Let G' be an arbitrary graph, and let $v, w, x \in V(G')$. Define an equivalence relation $\sim$ on $V(G')$ so that $v \sim w$ iff $N(v) = N(w)$. This is clearly an equivalence relation as $N(v), N(w), N(x)$ are sets of vertices, and equivalence of sets is an equivalence relation. Now, let G be a pentagon-extremal graph on n vertices such that $V(G)$ has the fewest number of equivalence classes under $\sim$. We claim that for all $x, y \in V(G)$, $N_G(x) = N_G(y)$ iff $c(C_5; G, \{x, y\}) = 0$.

Let $x, y \in V(G)$. Firstly, we claim that if $c(C_5; G, \{x, y\}) > 0$, then $N_G(x) \neq N_G(y)$: Since $c(C_5; G, \{x, y\}) > 0$, x and y must be either adjacent or at distance two in some copy of C_5. If $xy \in E(G)$, then $N_G(x) \neq N_G(y)$, as $y \in N_G(x)$ but $y \notin N_G(y)$. Now suppose that x and y are at distance two in some copy of C_5 where the vertices taken in cyclic order are v, x, w, y, z. If $N_G(x) = N_G(y)$, then that implies that as $v \in N_G(x)$ and $z \in N_G(y)$, we must have that $v \in N_G(y)$ and $z \in N_G(x)$. But that would imply that the three vertices v, x, z form a triangle, contradicting that G is triangle-free.

Now, we claim that if $c(C_5; G, \{x, y\}) = 0$, then $N_G(x) = N_G(y)$. If there are no pairs (x, y) where $c(C_5; G, \{x, y\}) = 0$ and $x \nsim y$, then we are done. If this is not the case, let x and y be two such vertices. For any vertices $w, z \in V(G)$ such that $w \sim x$ and $y \sim z$, it must also be true that $c(C_5; G, \{w, z\}) = 0$ and $w \nsim z$: If $c(C_5; G, \{w, z\}) > 0$, then since $N_G(w) = N_G(x)$ and $N_G(y) = N_G(z)$, one could take a copy of C_5 containing w and z, replace those two vertices with x and y respectively, and thus obtain a copy of C_5 containing x and y, a contradiction. If $w \sim z$, that implies $x \sim y$ by transitivity, a contradiction.

Let $S = \{v \in V(G) \mid v \sim y\}$ be the equivalence class of vertices in G containing y. For every $v \in S$, it is true that $c(C_5; G, \{x, v\}) = 0$, so it follows from Proposition 13 that $c(C_5; G, \{x\}) = c(C_5; G, \{v\})$. Now, we can iteratively perform the construction described in Proposition 13 by making each $v \in S$ a clone of x; observe that the graph we will obtain after each iteration of the construction must also be pentagon-extremal, because if G_0 is the graph obtained after making a fixed v a clone of x, we will have that $c(C_5; G_0) = c(C_5; G) - c(C_5; G, \{v\}) + c(C_5; G, \{x\}) - c(C_5; G, \{x, v\}) = c(C_5; G)$. Furthermore, S is finite, so we will only perform this construction a finite number of times. After doing this, however, we will obtain a pentagon-extremal graph G' such that $V(G')$ has fewer equivalence classes than $V(G)$ under $\sim$, as every vertex in the equivalence classes of x and y in G are now in the same equivalence class in G'. This contradicts our selection of G, and the desired result follows. □

In $C_5^{(n/5)}$, every vertex has degree $\frac{2n}{5}$. It turns out that one can also use Füredi's observation along with another tool from extremal graph theory to show that every sufficiently large pentagon-extremal graph where n is divisible by 5 has this as an upper bound for the degree of any vertex in that graph: By a *hypergraph* we mean a pair $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of vertices and $\mathcal{E}$ is a set of non-empty subsets of $\mathcal{V}$ called *hyperedges*. By convention we will assume that for a hypergraph $\mathcal{H}$, $\mathcal{V}(\mathcal{H}) \cap \mathcal{E}(\mathcal{H}) = \emptyset$. The *degree* $\deg(v)$ of a vertex $v \in \mathcal{V}(\mathcal{H})$ is the number of hyperedges containing v, and the degree $\deg(v, w)$ of two vertices $v, w \in \mathcal{V}(\mathcal{H})$ is the number of hyperedges containing both v and w. A *r-uniform hypergraph* is a hypergraph where every hyperedge contains r vertices.

Let $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ be a hypergraph. The *point covering number* $t(\mathcal{H})$ is the minimum t such that there exist t edges of $\mathcal{H}$ whose union is $\mathcal{V}(\mathcal{H})$. In [10], Frankl and Rödl proved the following:

Theorem 15. Given $\epsilon > 0$, $r \in \mathbb{N}$, and a real number $a > 3$, there exist $\delta, n_0 > 0$ such that for every $n > n_0$ and every r-uniform hypergraph $\mathcal{H}$ on n vertices, if for some D one has $\deg_{\mathcal{H}}(v) \in [(1-\delta)D, (1+\delta)D]$ for all $v \in \mathcal{V}(\mathcal{H})$ and $\deg_{\mathcal{H}}(v, w) < \frac{D}{(\log n)^a}$ for all $v, w \in \mathcal{V}(\mathcal{H})$, then $t(\mathcal{H}) \leq \frac{n(1+\epsilon)}{r}$.

We will not prove Theorem 15 here, but we can use this theorem to obtain the following result [17]:

Proposition 16. Let $\epsilon > 0$. Then there exists $n_1 > 0$ such that for every $n > n_1$ where n is divisible by 5, every pentagon-extremal graph G on n vertices must satisfy that every vertex in G has degree at most $\frac{(2+2\epsilon)n}{5}$, and must contain at most $\frac{(1+\epsilon)n^2}{5}$ edges.

Proof. For any graph G, define a 5-uniform hypergraph $\mathcal{H}(G)$ from this graph by letting every copy of C_5 in G be a hyperedge in $\mathcal{H}(G)$. For this hypergraph, we have that $\deg_{\mathcal{H}(G)}(v) = c(C_5; G, \{v\})$ and $\deg_{\mathcal{H}(G)}(v, w) = c(C_5; G, \{v, w\})$. If $\epsilon > 0$ is given, then by choosing $a = 4$, we know we can find constants $\delta, n_0 > 0$ satisfying the conditions in Theorem 15. Now, choose n_1 to be large enough such that:

- $n_1 > n_0$;
- $n_1 > \frac{5^4}{2\delta}$; and
- $\frac{n_1}{(\log n_1)^4} > 5^4$.

We claim that our choice of n_1 satisfies the condition given in the proposition.

Let G be a pentagon-extremal graph on $n > n_1$ vertices with n divisible by 5, and let $\mathcal{H} = \mathcal{H}(G)$. Then we know that $c(C_5; G) \geq \left(\frac{n}{5}\right)^5$, since this is a lower bound given by the graph $G = C_5^{(n/5)}$. Now, let $k = \max_{x \in V(G)} c(C_5; G, \{x\})$. We claim that $k \geq \frac{n^4}{5^4}$: Observe that

$$c(C_5; G) = \frac{1}{5} \sum_{x \in V(G)} c(C_5; G, \{x\}) = \frac{1}{5} \sum_{x \in V(G)} \deg_{\mathcal{H}}(x).$$

Since $k \geq \deg_{\mathcal{H}}(x)$ for every $x \in V(G)$, it follows that $\frac{1}{5}kn \geq c(C_5; G) \geq \frac{n^5}{5^5}$, and thus $k \geq \frac{n^4}{5^4}$.

By Proposition 13, it follows that for every $x \in V(G)$, we have that $c(C_5; G, \{x\}) \geq k - n^3$. Thus, it follows that there exists a $D \geq \frac{n^4}{5^4} - \frac{n^3}{2}$ such that $\deg_{\mathcal{H}}(x) \in [D - \frac{n^3}{2}, D + \frac{n^3}{2}]$. Since $n_1 > \frac{5^4}{2\delta}$, it then follows that $\frac{n^3}{2} < \delta D$, and so we must have that $\deg_{\mathcal{H}}(x) \in [(1-\delta)D, (1+\delta)D]$ for all $x \in \mathcal{V}(\mathcal{H})$. Proposition 13 also tells us that for every $x, y \in V(G)$, we have that $c(C_5; G, \{x, y\}) = \deg_{\mathcal{H}}(x, y) < n^3$. As $\frac{n_1}{(\log n_1)^4} > 5^4$, it also follows that $\deg_{\mathcal{H}}(x, y) < \frac{D}{(\log n)^4}$ for this choice of D. By Theorem 15, it now follows that $t(\mathcal{H}) \leq \frac{(1+\epsilon)n}{5}$, which means that $\frac{(1+\epsilon)n}{5}$ hyperedges cover $\mathcal{H}$, and thus one can select $\frac{(1+\epsilon)n}{5}$ copies of C_5 in G such that every vertex in $V(G)$ is in some copy of C_5.

If $v \in V(G)$, then v cannot be incident to three vertices of a single copy of C_5, or else G would contain a triangle with v and two adjacent vertices of the C_5. Thus, v is incident to at most two vertices on every copy of C_5, and so it follows that every vertex in G has degree at most $\frac{(2+2\epsilon)n}{5}$; if not, it would be adjacent to three vertices on one of our $\frac{(1+\epsilon)n}{5}$ copies of C_5 that cover $V(G)$. By the Handshake Lemma, the maximum number of edges in G is thus $\frac{(1+\epsilon)n^2}{5}$. $\square$

Chapter 3

Flag Algebras and the Erdős Pentagon Problem

3.1 Mantel, Erdős, and Razborov

Recall that Mantel's problem asks what the maximum number of edges a graph without triangles (that is, a graph that does not contain K_3 as a subgraph) can have. If one experiments with Mantel's problem for small cases, one may find that a triangle-free graph on n vertices with $\lfloor \frac{n^2}{4} \rfloor$ edges can be constructed by building a complete bipartite graph between independent sets of size $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$ respectively. In the same journal issue as Mantel's problem, Mantel, along with Wythoff and several others, showed that the bound induced by this example is tight, establishing the following theorem:

Theorem 17. Suppose that G is a triangle-free graph on n vertices. Then G has at most $\lfloor \frac{n^2}{4} \rfloor$ edges.

Mantel's theorem has many elegant proofs, one of which is as follows (see, for example, [1]):

Proof 1 of Mantel's Theorem. Let G be a triangle-free graph on n vertices, and let S be a largest independent set in G. The neighbourhood of every vertex $v \in V(G)$ must be an independent set, or else G would contain a triangle. Thus, we must have $\deg(v) \leq |S|$ for all $v \in V(G)$. Also, since S is independent, every edge is incident to some vertex in

$V(G) \setminus S$. Thus, we must have that

$$|E(G)| \leq \sum_{v \in V(G) \setminus S} \deg(v) \leq |S||V(G) \setminus S| \leq \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil \leq \frac{n^2}{4}$$

by the AM-GM inequality (Theorem 2), and in fact, since the number of edges in a graph must be an integer, it is true that $|E(G)| \leq \left\lfloor \frac{n^2}{4} \right\rfloor$. Furthermore, the second inequality is only tight if every vertex in S is adjacent to every vertex in $V(G) \setminus S$, and the third inequality is only tight if $|S| = \left\lfloor \frac{n}{2} \right\rfloor$ or $|S| = \left\lceil \frac{n}{2} \right\rceil$, so the only extremal example is the balanced complete bipartite graph described above. □

Mantel's problem bears a resemblance to Erdős's pentagon problem, in that it asks about the maximal number of some subgraphs in an n-vertex graph (edges in one case, pentagons in the other) subject to the non-existence of some other subgraphs (triangles). The method of flag algebras, introduced by Razborov in [26], allows us to computationally tackle such problems through providing a systematic framework generalising counting techniques in graph theory. Though this method is relatively new, it has already been successfully used to approach or solve many problems in extremal combinatorics (see the survey paper by Razborov [27] for more details), and indeed, it was the method that was first used to resolve the Erdős pentagon problem.

Razborov's original paper was framed in the language of model theory, and the method of flag algebras as it is outlined there works for an arbitrary universal first-order logic theory without constants or function symbols. This is powerful because it can be applied to any combinatorial structure possessing the hereditary property - namely, the property that any subset of vertices of such a combinatorial structure gives rise to another such "induced" structure. Notably, we can apply the method of flag algebras to graph-theoretic problems where we forbid some set of induced subgraphs $\mathcal{H}$: Call a graph G $\mathcal{H}$-*free* if no induced subgraph of G is isomorphic to a graph in $\mathcal{H}$. If G is $\mathcal{H}$-free, it is clear that the subgraph induced by any vertex subset of $V(G)$ must also be $\mathcal{H}$-free as a graph.

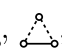
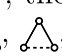
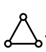
In this chapter, we will introduce the method of flag algebras specifically as it applies to $\mathcal{H}$-free graphs, drawing from [5, 12, 20, 26] as references. We will use Mantel's problem and Mantel's theorem as guiding examples, illustrating ideas from the method as they apply to triangle-free graphs and providing several proofs of Mantel's theorem. We will first employ an asymptotic, density-based proof of Mantel's theorem to motivate the study of flag algebras. This will guide us through the construction of flag algebras; along the way, we will see several important results and constructions found in [26]. Finally, we will introduce the semidefinite method as a way that one can use flag algebras to computationally solve problems in extremal graph theory, and use the semidefinite method to present a solution to the Erdős pentagon problem, as given by Grzesik [11] in 2012.

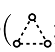
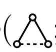
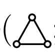
Note that throughout this chapter, we will be using pictures to represent graphs. Dashed lines will be used to represent non-edges, and solid lines will be used to represent edges. For example, the picture represents the empty graph on two vertices, and the picture  represents the path P_3 on three vertices.

3.2 An Asymptotic Proof of Mantel's Theorem

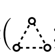
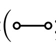
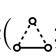
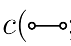
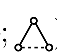
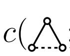
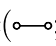
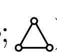
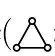
In this section, we will introduce the notion of density and use this to provide an asymptotic proof of Mantel's theorem.

Recall that if G and F are graphs, we defined $c(F;G)$ to be the number of times F appears as an induced subgraph of G. Define the *density* of F in G to be $p(F;G) = c(F;G)\binom{|V(G)|}{|V(F)|}^{-1}$; namely, this is the probability that when $|V(F)|$ vertices are selected uniformly at random from $V(G)$, we obtain an induced subgraph of G isomorphic to F.

If G is an arbitrary graph, we can state some "density identities" on G that will enable us to prove Mantel's theorem: Firstly, if we choose a subset of three vertices from G, the corresponding induced subgraph will be isomorphic to one of the four graphs , , , or . Thus, the following equation holds for all graphs G:

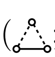
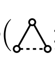
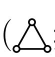
$p($  $;G) + p($  $;G) + p($  $;G) + p($  $;G) = 1.$ (I1)

Consider also that we can compute the number of edges in G by first counting the number of edges in each three-vertex subgraph of G, and then counting the number of appearances of each three-vertex subgraph in G:

$c($ $;$  $)c($  $;G) + c($  $;$  $)c($  $;G) + c($  $;$  $)c($  $;G) + c($  $;$  $)c($  $;G)$

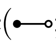
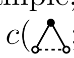
$= (|V(G)|-2)c($ $;G),$

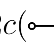
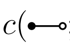
as each edge is in $|V(G)|-2$ subgraphs of size three. Then, we can turn this into a density identity through dividing through by $\binom{|V(G)|}{2}\cdot(|V(G)|-2) = 3\binom{|V(G)|}{3}$, so that we obtain an expression for $p($ $;G)$:

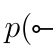
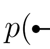
$\frac{0}{3}p($  $;G) + \frac{1}{3}p($  $;G) + \frac{2}{3}p($  $;G) + \frac{3}{3}p($  $;G) = p($  $;G).$ (I2)

Alternately, one can interpret this identity from a probability standpoint: If we want to find the edge density of G, we can consider the edge density in each possible three-vertex subgraph, then factor this through the density of each three-vertex subgraph in G.

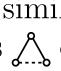
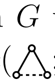
Unfortunately, these identities alone are not enough - to build the other identities needed in this proof, we need to extend the notion of density to labeled graphs. Let G be a graph, and let G^v be the graph obtained from G by labeling the vertex v with the label 1. (In the formalism of section 1.2, we have chosen $k_G = 1$ and let $\theta_G(1) = v$.) In our pictures, we will represent a graph with a vertex labeled 1 by filling in the corresponding vertex. For example, the picture  represents the path on three vertices where the vertex of degree two is labeled.

Now, if F is a graph with a vertex labeled 1, let $c(F; G^v)$ be the number of times F appears as an induced subgraph of G where v is the distinguished vertex of F. In other words, this is the number of sets $U \subseteq V(G) \setminus \{v\}$ with $|U| = |V(F)| - 1$ such that $G[U \cup \{v\}]$ is isomorphic to F via an isomorphism that preserves the label 1. For example, for all triangle-free graphs G and an arbitrary $v \in V(G)$, $c($$; G^v) = \deg(v)$ and $c($$; G^v) = \binom{\deg(v)}{2}$. Then, analogously to our definition of density above, let $p(F; G^v) = c(F; G^v)\binom{|V(G)|-1}{|V(F)|-1}^{-1}$ be the probability that a set $U \subseteq V(G) \setminus \{v\}$ with $|U| = |V(F)| - 1$ chosen uniformly at random is isomorphic to F via a label-preserving isomorphism.

Firstly, note that we can obtain several density identities by relating counts of subgraphs (hence densities) in unlabeled graphs and that of graphs with one labeled vertex: Let G be a graph with $|V(G)| = n$. By fixing each vertex v in turn and looking at the edges with this fixed vertex as one endpoint, we will consider each edge in G twice, so that $2c($$; G) = \sum_{v \in V(G)} c($$; G^v)$. (This is a different formulation of what is usually known as the Handshake Lemma.) We can obtain a density result from this by dividing both sides by $n(n-1)$, so that

$p($$; G) = \frac{1}{n} \sum_{v \in V(G)} p($$; G^v)$. (I3)

Interpreting this from a probability standpoint, to obtain the edge density of G, we can average over the edge densities of G where one vertex of the edge is fixed.

We can obtain similar results for other subgraphs. For example, each three-vertex subgraph with two edges  only has one vertex of degree two, so by fixing each vertex v in turn and looking at subgraphs on three vertices in G with this fixed vertex, we will consider each copy of this subgraph exactly once, and $c($$; G) = \sum_{v \in V(G)} c($$; G^v)$. Through dividing by

$3\binom{|V(G)|}{3}$, we obtain a density result:

$$\frac{1}{3}p(\overset{\circ}{\wedge}; G) = \frac{1}{n}\sum_{v\in V(G)} p(\overset{\bullet}{\wedge}; G^v). \tag{I4}$$

Again, we can interpret this from a probability standpoint: By taking the average density of $\overset{\bullet}{\wedge}$ in G over all vertices in G, we only obtain one-third of the density of $\overset{\circ}{\wedge}$ in G, because if we pick a vertex in $\overset{\circ}{\wedge}$ at random, there's only a $\frac{1}{3}$ chance that we pick the vertex of degree two.

Finally, for a fixed vertex v, consider the product $p(\bullet\!\!-\!\!\circ; G^v)^2$. This is the probability that, when two vertices in $V(G)\setminus\{v\}$ (not necessarily distinct) are picked, both of these vertices are adjacent to v. By considering the possible induced subgraphs on v and two other vertices, if the two vertices picked are distinct, then the probability that both vertices are adjacent to v is $p(\overset{\bullet}{\wedge}; G^v) + p(\overset{\bullet}{\triangle}; G^v)$. Otherwise, the probability that the two vertices picked are identical and adjacent to v is $\frac{\deg(v)}{(n-1)^2} = \frac{c(\bullet\!\!-\!\!\circ;G^v)}{(n-1)^2}$. Thus, it follows that

$$p(\bullet\!\!-\!\!\circ; G^v)^2 = p(\overset{\bullet}{\wedge}; G^v) + p(\overset{\bullet}{\triangle}; G^v) + \frac{c(\bullet\!\!-\!\!\circ; G^v)}{(n-1)^2}. \tag{I5}$$

We now have the tools we need to prove Mantel's theorem in the density sense. We will prove the following theorem:

Theorem 18. Suppose that G is a graph on n vertices with $p(\triangle; G) = 0$. Then $p(\circ\!\!-\!\!\circ; G) \le \frac{n-1}{2n-6}$.

Clearly, this implies Mantel's theorem asymptotically.

Proof 2 of Mantel's Theorem. Let G be a graph with $|V(G)| = n$ and $p(\triangle; G) = 0$. For any vertex $v \in V(G)$, it is true that $0 \le (1 - 2p(\bullet\!\!-\!\!\circ; G^v))^2$. By averaging this over all vertices in $V(G)$, we get that

$$0 \le \frac{1}{n}\sum_{v\in V(G)} (1 - 2p(\bullet\!\!-\!\!\circ; G^v))^2.$$

Then, by expanding and applying identities I3, I4, and I5, we get that

$$
\begin{aligned}
0 &\le \frac{1}{n} \sum_{v \in V(G)} (1 - 2p(\bullet\!\!-\!\!\circ; G^v))^2 \\
&= \frac{1}{n} \sum_{v \in V(G)} \big(1 - 4p(\bullet\!\!-\!\!\circ; G^v) + 4p(\bullet\!\!-\!\!\circ; G^v)^2\big) \\
&= \frac{1}{n} \sum_{v \in V(G)} \left(1 - 4p(\bullet\!\!-\!\!\circ; G^v) + 4p(\Lambda^{\bullet}; G^v) + 4p(\triangle^{\bullet}; G^v) + 4 \cdot \frac{c(\bullet\!\!-\!\!\circ; G^v)}{(n-1)^2}\right) \\
&= 1 - 4p(\circ\!\!-\!\!\circ; G) + \frac{4}{3} p(\Lambda; G) + \frac{4}{n} \sum_{v \in V(G)} \frac{c(\bullet\!\!-\!\!\circ; G^v)}{(n-1)^2}.
\end{aligned}
$$

But since $p(\triangle, G) = 0$, identity I2 tells us that

$$
2p(\circ\!\!-\!\!\circ; G) = \frac{2}{3} p(\underline{\overset{\circ}{\;\;}}; G) + \frac{4}{3} p(\Lambda; G).
$$

Furthermore, by the Handshaking Lemma, we have that

$$
\frac{4}{n} \sum_{v \in V(G)} \frac{c(\bullet\!\!-\!\!\circ; G^v)}{(n-1)^2} = \frac{8c(\circ\!\!-\!\!\circ; G)}{n(n-1)^2} = \frac{4p(\circ\!\!-\!\!\circ; G)}{n-1}.
$$

Thus, it follows that

$$
\begin{aligned}
0 &\le 1 - 4p(\circ\!\!-\!\!\circ; G) + \frac{4}{3} p(\Lambda; G) + \frac{4}{n} \sum_{v \in V(G)} \frac{c(\bullet\!\!-\!\!\circ; G^v)}{(n-1)^2} \\
&= 1 - 2p(\circ\!\!-\!\!\circ; G) - \frac{2}{3} p(\underline{\overset{\circ}{\;\;}}; G) + \frac{4}{n-1} p(\circ\!\!-\!\!\circ; G) \\
&\le 1 - 2p(\circ\!\!-\!\!\circ; G) + \frac{4}{n-1} p(\circ\!\!-\!\!\circ; G),
\end{aligned}
$$

and so it follows that $p(\circ\!\!-\!\!\circ; G) \le \frac{n-1}{2n-6}$, as desired. $\square$

This proof leaves something to be desired. It follows easily from the density identities we built earlier, but how did we know that these were the right identities to derive? Furthermore, the first inequality in the proof helps us get where we want, but seems contrived; is there a more natural way this inequality could arise? Or, to avoid this altogether, is there a more systematic way that we could work with these density identities?

In the next few sections, we will work to define *flag algebras*, structures that will enable us to deal with density identities and density computations more easily. We will see how all of the identities used in this proof arise naturally through working with different flag algebras, and demonstrate how we can use flag algebras to generate proofs of results such as Mantel's theorem in an essentially mechanical manner.

3.3 Flags and Density

In this section, we will set up some of the machinery required to proceed with constructing flag algebras. In the process, we will investigate a generalisation of the notion of density defined in the previous section.

We are interested in looking at extremal graph theory problems concerning classes of graphs which forbid given subgraphs - in the case of both Mantel's theorem and the Erdős pentagon problem, for example, we are interested in triangle-free graphs. For the remainder of this chapter, fix $\mathcal{H}$, a set of forbidden induced subgraphs. For example, in the case of both Mantel's theorem and the Erdős pentagon problem, we will have that $\mathcal{H} = \{\triangle\}$; in general, though, we can forbid more than one subgraph.

We will now set up several definitions, using the formalism of labelled graphs defined in section 1.2:

Definition 19 (Types and flags)**.** Recall that $[k] = \{1, 2, 3, \cdots, k\}$ for $k \in \mathbb{N}$, and $[0]$ is the empty set. Then:

- A *type* of size k is a labelled $\mathcal{H}$-free graph σ with $|V(\sigma)| = k_\sigma = k$.
- Let σ be a type of size k, and F be an unlabelled $\mathcal{H}$-free graph with $|V(F)| \geq k$. An *embedding* of σ into F is an injective map $\theta : [k] \to V(F)$ such that $\sigma \cong \mathrm{Im}(\theta)$ as labeled graphs.
- Let σ be a type of size k. A *σ-flag* of size n is a pair (F, θ) where F is an $\mathcal{H}$-free unlabelled graph with $n = |V(F)| \geq k$ and θ is an embedding of σ into F. (Alternately, a σ-flag of size n is a labelled graph $(V(F), E(F), k, \theta)$ where $|V(F)| = n$.)
- Let (F, θ) and (G, η) be σ-flags. We say these are *isomorphic* (and denote flag isomorphism by the symbol $\cong$) if there is a labelled graph isomorphism $\rho : V(F) \to V(G)$ with $\rho(\theta(i)) = \eta(i)$ for all $i \in [|V(\sigma)|]$.

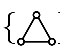
Note that when the embedding θ of σ into a σ-flag is irrelevant, we may write the σ-flag (F, θ) as simply F. If it is clear that σ is a fixed type, we may also refer to a σ-flag as simply a "flag".

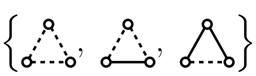
To rephrase these definitions more colloquially: A type is a labeled graph on k vertices where every vertex has a distinct label, and we can embed types into a larger graph by injectively mapping $[k]$ into the vertex set of the larger graph to yield a labelled graph so that the graph induced by the image of the map is isomorphic to the type as a labeled graph. For a fixed type σ, a σ-flag is a labelled graph with an embedded copy of σ, and two σ-flags are isomorphic if there is a labelled graph isomorphism preserving the labels of σ.

As the title of this chapter suggests, flags will be our central object of study for the remainder of this chapter. To that extent, one might wonder how the term "flag" was chosen, and whether the term "flag" is meant to bear any connection to other mathematical objects labeled with the word "flag" in different fields. In a footnote of [28], Razborov clarifies that there is no such connection:

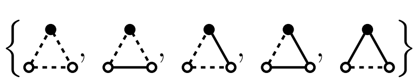
> "The choice of the term 'flag'... is admittedly somewhat arbitrary. It is largely suggested by a visual association: A few vertices are fixed rigidly while many more are 'free' and 'waving' through the model we are studying. It has very little to do with other usages of this term in mathematics... incidentally, I have never seen a good explanation of what [flags in linear algebra] have to do with corporeal flags, either."

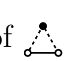
For a fixed type σ, let $\mathcal{F}^\sigma$ be the set of all σ-flags up to isomorphism, and $\mathcal{F}^\sigma_n \subseteq \mathcal{F}^\sigma$ be all the σ-flags of size n up to isomorphism. We now present a brief example to illustrate these definitions:

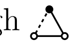
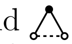
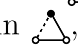
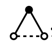
Example 20. Let $\mathcal{H} = \{$$\}$. If we let $\sigma = \varnothing$ be the empty type, then $\mathcal{F}^\varnothing_3$, the set of $\varnothing$-flags of size 3, consists of all unlabeled triangle-free graphs on three vertices. There are three of them:

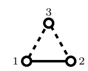
$\mathcal{F}^\varnothing_3 =$  .

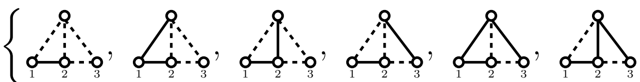
If we let $\sigma = \bullet$ be the type on one vertex, then $\mathcal{F}^\bullet_3$, the set of $\bullet$-flags of size 3, consists of all labeled triangle-free graphs on three vertices where a label has been assigned to exactly one vertex. There are five of these:

$\mathcal{F}^\bullet_3 =$  .

Note that even though the underlying unlabeled graphs of  and  are isomorphic, as

•-flags they are not isomorphic, because in , the vertex of degree zero has been labeled, whereas in , a vertex of degree one has been labeled. Similarly, even though  and  are isomorphic as unlabeled graphs, as •-flags they are not isomorphic, because in , a vertex of degree one has been labeled, whereas in , the vertex of degree two has been labeled.

Finally, we can consider a larger example: If we let $\sigma =$ , namely the graph with $V(\sigma) = [3]$ and $E(\sigma) = \{12\}$, then $\mathcal{F}_4^\sigma$ consists of the six labeled triangle-free graphs on four vertices where three vertices have been labeled with the labels $\{1, 2, 3\}$ in a way that admits an embedding of σ:

$\mathcal{F}_4^\sigma = \Bigg\{$  $\Bigg\}.$

(The edge $e = 13$ is not drawn into these pictures because e is not an edge of σ, and hence cannot be an edge of any σ-flag.)

Note that $\mathcal{F}_n^\sigma$ is always a finite set, but $\mathcal{F}^\sigma$ could be finite or infinite. For example, if $\mathcal{H}$ consists of all graphs on three vertices, then the one-vertex type $\sigma = \bullet$ and the two two-vertex •-flags  and are all $\mathcal{H}$-free, but there are no •-flags of size three or larger. If $\mathcal{F}^\sigma$ is finite, we call σ a *degenerate type*. Henceforth, we will assume that all types are nondegenerate; that is, $\mathcal{F}^\sigma$ is infinite.

Recall that in our asymptotic proof of Mantel's theorem, we had to employ identities considering densities of graphs with one labeled vertex - namely, •-flags. In our proof, we mainly considered the density of one •-flag inside another •-flag, but notice that in constructing identity I5, we asymptotically considered the density of fitting *two* •-flags inside another •-flag so that the embedding of • is preserved. Motivated by this, we now work to extend the idea of density to fitting a number of σ-flags inside a larger σ-flag.

Fix a type σ. First, let us try to find the right idea for what it means to fit more than one flag in a larger flag. Imagine trying to map two flags F_1, F_2 into a very large flag G so that the embedding of σ is preserved. If G is sufficiently large, the chance that the images of F_1, F_2 in G will overlap is small enough that we can disregard it in the limit. (We will see a more formal explanation of this shortly.) Thus, in looking at how we can construct the density of F_1 and F_2 in G, it makes sense that we largely want to consider the ways in which we can map F_1 and F_2 into G in such a way that they only overlap in their embeddings of σ.

This bears a resemblance to a sunflower in extremal set theory: A collection $S_1, \cdots, S_t$ of finite sets is said to be a sunflower with centre C if $S_i \cap S_j = C$ for every two distinct

$i, j \in [t]$. (In fact, in [26], Razborov defined density more generally in terms of sunflowers.) As an analogy, one can compare this to a real-life sunflower, where each set is represented by the union of a petal and the centre of the flower. In this situation, we essentially want to build a "sunflower" of graphs rather than sets, where the graphs given by the embeddings of the flags F_1, F_2 only overlap in their embeddings of σ in G, and indeed, the definition of density that we want to build is the probability that the sets generated by a randomly chosen sunflower of appropriate size in $V(G)$ yield sets whose induced subgraphs are isomorphic to F_1, F_2:

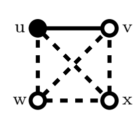
Definition 21 (Density)**.** Let σ be a type of size k, and let $F_1, \cdots, F_t, G$ be σ-flags. Define the *density* $p(F_1, \cdots, F_t; G)$ of $F_1, \cdots, F_t$ in G to be as follows:

- We say that $F_1, \cdots, F_t$ *fit* in G if $(|V(F_1)| - k) + \cdots + (|V(F_t)| - k) \leq |V(G)| - k$.
- If $F_1, \cdots, F_t$ fit in G, then $p(F_1, \cdots, F_t; G)$ is the probability that, if we choose pairwise disjoint sets $S_1, \cdots, S_t \subseteq V(G) \setminus \mathrm{Im}(\theta)$ of unlabeled vertices uniformly at random such that $|S_i| = |V(F_i)| - k$ for all $i \in [t]$, then $G[S_i \cup \mathrm{Im}(\theta)] \cong F_i$ for all $i \in [t]$.
- If $F_1, \cdots, F_t$ do not fit in G, then $p(F_1, \cdots, F_t; G) = 0$.

Note that the definition of fitting σ-flags into a larger flag only depends on the numbers $|V(G)|$ and $|V(F_i)|$ rather than the flags themselves; it tells us whether or not we can find a "sunflower" in G with a "centre" of size $|\sigma|$ and "petals" corresponding to the sizes of $|F_1 \setminus \sigma|, \cdots, |F_t \setminus \sigma|$. Then, if we choose this sunflower randomly by fixing σ and picking sets S_i corresponding to the "petals" one at a time, the density of the flags $F_1, \cdots, F_t$ is the probability that for all i, S_i along with σ induces a σ-flag isomorphic to F_i. Note that if $t = 1$, this definition is precisely the definition of density we used previously.

However, there is one aspect of precision in this definition that we should be careful of. In particular, we choose $S_1, \cdots, S_t$ in order, and this does make a difference, as demonstrated in the following example:

Example 22. Let $\sigma = \bullet$ be the one-vertex type. Consider the $\bullet$-flag

$G =$ 

,

the graph on four vertices $\{u, v, w, x\}$ with one edge $\{uv\}$ where one endpoint u of the edge is labeled. Then, we can fit the $\bullet$-flags $F_1 = \bullet\!-\!\circ$, $F_2 = \bullet\cdots\circ$, $F_3 = \bullet\cdots\circ$ in G; let us compute $p(\bullet\!-\!\circ, \bullet\cdots\circ, \bullet\cdots\circ; G)$. By considering the three two-vertex subgraphs of G induced by the vertex sets $\{u,v\},\{u,w\},\{u,x\}$, we see one of the subgraphs is a copy of $\bullet\!\!-\!\!\circ$ and two of the subgraphs are copies of $\bullet\cdots\circ$.

However, this does not mean that $p(\bullet\!\!-\!\!\circ, \bullet\cdots\circ, \bullet\cdots\circ; G) = 1$. Of the 3! possible ways to choose the one-vertex sets S_1, S_2, S_3 in order from $\{v,w,x\}$, only two of these will assign S_1 to be v, so that $G[S_1 \cup \{u\}] \cong \bullet\!\!-\!\!\circ$, $G[S_2 \cup \{u\}] \cong \bullet\cdots\circ$, and $G[S_3 \cup \{u\}] \cong \bullet\cdots\circ$. Thus, it follows that $p(\bullet\!\!-\!\!\circ, \bullet\cdots\circ, \bullet\cdots\circ; G) = \frac{1}{3}$.

As seen from this example, computing the density of flags inside another small flag is not difficult with careful counting. However, is there a way that one can easily compute the density of flags in a much larger flag? Or, as is the case in many problems in extremal combinatorics, what if we don't know what this larger flag might look like, but know the density of certain flags inside this larger flag? For example, in an extremal graph theory problem working with triangle-free graphs, we may not know what a triangle-free graph G looks like, but we will know that G would satisfy $p(\triangle; G) = 0$.

Recall that in our derivation of identity I2 in the previous section, we were able to relate the edge density of a graph G to the density of three-vertex subgraphs in G by counting the number of appearances of an edge in each three-vertex subgraph, then counting the number of appearances of each three-vertex subgraph in G. In an analogous way, we can relate the density of smaller σ-flags $p(F;G)$ to the density of larger σ-flags as follows: If we knew the densities of all the flags of a fixed size n in G, where $|V(F)| \leq n \leq |V(G)|$, we could obtain another expression for $p(F;G)$ by considering the densities of F in each of these flags of size n, then factoring this through the density of each flag of size n in G. That is, if $|V(F)| \leq n \leq |V(G)|$, it follows that

$$p(F;G) = \sum_{F' \in \mathcal{F}_n^\sigma} p(F;F')p(F';G).$$

We can more rigorously prove a generalisation of this accounting for the densities of multiple flags ([26], Lemma 2.2):

Theorem 23. Let σ be a type of size k. Let $F_1, \cdots, F_t, G$ be σ-flags, and let $1 \leq s \leq t$ and $n \in \mathbb{N}$ be such that

- $F_1, \cdots, F_s$ fit in a σ-flag of size n, and
- A σ-flag of size n along with $F_{s+1}, \cdots, F_t$ fit in G.

Then

$$p(F_1, \cdots, F_t; G) = \sum_{F \in \mathcal{F}_n^\sigma} p(F_1, \cdots, F_s; F)p(F, F_{s+1}, \cdots, F_t; G).$$

Proof. Let $l = |\mathcal{F}_n^\sigma|$, and fix an ordering of $\mathcal{F}_n^\sigma$. Suppose that we first choose pairwise disjoint sets $S^*, S_{s+1}, \cdots, S_t \subseteq V(G) \setminus \mathrm{Im}(\theta)$ of unlabeled vertices uniformly at random such that $|S^*| = n-k$ and $|S_i| = |V(F_i)|-k$ for $i \in \{s+1, \cdots, t\}$, and then choose pairwise disjoint sets $S_1, \cdots, S_s \subseteq S^*$ uniformly at random such that $|S_i| = |V(F_i)| - k$ for $i \in [s]$. Let A be the event that $G[S_i \cup \mathrm{Im}(\theta)] \cong F_i$ for all $i \in [t]$. For $i \in [l]$, let B_i be the event that both $G[S^* \cup \mathrm{Im}(\theta)]$ is isomorphic to the ith element of $\mathcal{F}_n^\sigma$ and $G[S_i \cup \mathrm{Im}(\theta)] \cong F_i$ for all $i \in \{s+1, \cdots, t\}$.

Observe that $B_1, \cdots, B_l$ are disjoint events. Furthermore, observe that the event A will not occur if none of the events B_i occur, since any choice of $n-k$ unlabeled vertices in G along with $\mathrm{Im}(\theta)$ will induce some σ-flag of size n. Thus, it follows that $P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_l)P(B_l)$.

Let F' be the ith element of $\mathcal{F}_n^\sigma$. By definition, we have that $P(A) = p(F_1, \cdots, F_t; G)$, $P(B_i) = p(F', F_{s+1}, \cdots, F_t; G)$, and $P(A|B_i) = p(F_1, \cdots, F_s; F')$. The result follows by substitution. $\square$

This is usually referred to as the *chain rule*, owing to the intuition that we can "chain" the density of a flag through the densities of larger flags. We demonstrate this with a brief example:

Example 24. Suppose that $\mathcal{H} = \{\triangle\}$, and let $\sigma = \bullet$ be the one-vertex type. Observe that the two $\bullet$-flags $\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ$ fit in a $\bullet$-flag of size 3. As noted in a previous example, we can enumerate $\mathcal{F}_3^\bullet$ as $\mathcal{F}_3^\bullet = \{\triangle, \triangle, \triangle, \triangle, \triangle\}$. Then for any $\bullet$-flag G, the chain rule tells us that

$$p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; G) = \sum_{F' \in \mathcal{F}_3^\bullet} p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; F')p(F'; G) = p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; \triangle)p(\triangle; G),$$

as the only $\bullet$-flag of size 3 in which we can embed two copies of $\bullet\!\!-\!\!\circ$ overlapping only in the one vertex $\bullet$ is the flag $\triangle$, but it follows that

$$p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; \triangle)p(\triangle; G) = p(\triangle; G),$$

since both possible ways to choose two disjoint sets of one non-$\bullet$ vertex in $\triangle$ generate two flags isomorphic to $\bullet\!\!-\!\!\circ$. Thus, we get that $p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; G) = p(\triangle; G)$ for every $\bullet$-flag G.

Recalling the construction of identity I5 in our asymptotic proof to Mantel's Theorem, we showed that when $\mathcal{H} = \{\triangle\}$, it is true that $p(\bullet\!\!-\!\!\circ; G)^2 = p(\triangle; G) + \frac{c(\circ\!\!-\!\!\circ; G)}{(n-1)^2}$, so that asymptotically, $p(\bullet\!\!-\!\!\circ; G)^2 \approx p(\triangle; G)$. This example shows us that $p(\triangle; G) = p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; G)$ exactly, so the difference between $p(\bullet\!\!-\!\!\circ; G)^2$ and $p(\bullet\!\!-\!\!\circ, \bullet\!\!-\!\!\circ; G)$ should shrink as the size of G grows larger. Indeed, we interpreted this difference as the probability that two independently picked vertices which were not a fixed vertex $v \in V(G)$ were identical and adjacent to v. Using this idea, we can show in general that working with the density of several flags is essentially equivalent asymptotically to working with the product of the density of these individual flags ([26], Lemma 2.3):

Theorem 25. *Let σ be a type of size k. Then there exists a function $f(n) = O(\frac{1}{n^t})$ such that if $F_1, \cdots, F_t, G$ are σ-flags,*

$$|p(F_1, \cdots, F_t; G) - \prod_{i=1}^{t} p(F_i; G)| \leq f(|V(G)|).$$

Proof sketch. Suppose that we choose sets $S_1, \cdots, S_t \subseteq V(G) \setminus \mathrm{Im}(\theta)$ of unlabeled vertices uniformly at random such that $|S_i| = |V(F_i)| - k$; note that these sets do not need to be pairwise disjoint. Let A be the event that $G[S_i \cup \mathrm{Im}(\theta)] \cong F_i$ for all $i \in [t]$, and let B be the event that $S_1, \cdots, S_t$ are pairwise disjoint.

Now, by the definition of conditional probability, $P(A)+P(B)-P(A|B) \leq P(A)+P(B)-P(A \cap B) = P(A \cup B) \leq 1$. Thus, $P(A) - P(A|B) \leq 1 - P(B)$. By construction, we also know that $P(A) = \prod_{i=1}^{t} p(F_i; G)$ and $P(A|B) = p(F_1, \cdots, F_t; G)$. Finally, we can also see that

$$1 - P(B) = 1 - \frac{\binom{|V(G)|-k}{|V(F_1)|-k, |V(F_2)|-k, \cdots, |V(F_t)|-k, |V(G)|+(t-1)k-\sum_{i=1}^{t}|V(F_i)|}}{\prod_{i=1}^{t}\binom{|V(G)|-k}{|V(F_i)|-k}}$$

is the probability that $S_1, \cdots, S_t$ are not pairwise disjoint, and it can be shown that $1 - P(B) \leq O\left(\frac{1}{|V(G)|^t}\right)$. □

3.4 A Historical Aside

At this point, we want to provide the reader with some historical context in order to situate the study of flag algebras in the graph theory landscape. Along the way, we will introduce some definitions necessary for a more thorough discussion of flag algebras.

Compared to other fields of mathematics, graph theory is a relatively new field. The foundational paper of graph theory was only published in 1736 when Euler published his solution to the problem of the bridges of Königsberg, and throughout most of the next two centuries, the development of what we now see as "graph theory" was largely focused on the exploration of specific graphs, specific types of graphs, or seen as an application to another field. For example, Vandermonde, Kirkman, and Hamilton worked on the question of the existence of a Hamiltonian circuit in the specific context of the knight's tour graph, graphs of polyhedra, and the icosahedron graph respectively, and the four-colour theorem was first posed in the context of colouring maps. In fact, even the term "graph" stemmed from applications of graph theory to chemistry - it was first penned by Sylvester in 1878, and was derived from chemistry's "graphical notation"! (The interested reader can find more about the early history of graph theory in [2].)

It still took several decades and many publications for graph theory to be recognised by the greater mathematical community as a field of mathematics in its own right, though; arguably, this happened when Kőnig published the first textbook on graph theory in 1936. As graph theory continued to blossom throughout the mid-20th century, mathematicians working in the field wondered what connections could be made between graph theory and other branches of mathematics, and how these connections could be exploited. For example, would it be possible to find a way to discuss algebraic maps and structures in relation to graphs in a way that one could employ the more powerful algebraic theories being developed at the time?

Recall that a homomorphism, in general, is a map between two objects of the same structure preserving the operations in that structure. For example, if Γ_1, Γ_2 are groups with binary operation $*$, a map $f : \Gamma_1 \to \Gamma_2$ is a homomorphism if for all $x, y \in \Gamma_1$, $f(x*y) = f(x)*f(y)$. In a 1961 paper [29], Sabidussi laid out an analogous definition for homomorphisms between graphs: If G_1, G_2 are graphs, a map $\phi : V(G_1) \to V(G_2)$ is a *graph homomorphism* if for all $u, v \in V(G_1)$, $f(u)f(v) \in E(G_2)$ when $uv \in E(G_1)$. In this way, the "operation" that is preserved between the graphs is edge adjacency - note, though, that this definition does not require that non-adjacency be preserved, and does not require that the map ϕ is injective.

Graph homomorphisms, defined in this sense, have remained an object of study since then. There has been a lot of development in the study of graph homomorphisms in the last fifty years, and a more thorough account of the progress in this field can be found in [18]. That being said, one can view the efforts we are undertaking here as part of this study of graph homomorphisms: In looking at the ways that a graph F appears as an induced subgraph of G, we are really considering *induced homomorphisms* from F to G - homomorphisms which are injective and preserve both adjacency and non-adjacency.

Accordingly, what we have previously defined as simply "density" for the one-graph case is sometimes also known as *induced homomorphism density*, and in this sense represents the proportion of possible choices of vertex sets which can yield induced homomorphisms. This view of being able to "normalise" to see homomorphisms as a proportion rather than as a count, though, did not arise for two decades after homomorphisms started to be studied;

the first recorded instance of this was in a 1978 paper [8] by Erdős, Lovász, and Spencer. This paper was interested in obtaining results relating limit points of sequences of different types of homomorphism densities of small graphs in increasingly large graphs. These density parameters always lie between 0 and 1, so looking at limit points was a sensible question framed from a density perspective, but not from a counting perspective.

If we want to study these limit points of sequences, it would be nice for these limit points to exist in first place. Thus, we can try and define sequences for which we know these limit points will exist. Let $\mathcal{G}$ be the set of all $\mathcal{H}$-free graphs.

Definition 26 (Convergent sequence of graphs)**.** Let $(G_k)_{k\geq 1}$ be a sequence of graphs in $\mathcal{G}$. We say that this sequence is *convergent* if, for every graph $F \in \mathcal{G}$, $\lim\limits_{k\to\infty} p(F; G_k)$ exists.

Interestingly, every increasing sequence of graphs (that is, a sequence of graphs (G_k) for which $|V(G_i)| > |V(G_{i-1})|$, $\forall i > 1$) has a convergent subsequence, so given that we want to look at the homomorphism densities of small graphs in increasingly large graphs, it would really suffice to study convergent sequences of graphs.

Theorem 27. Let $(G_k)_{k\geq 1}$ be an increasing sequence of graphs in $\mathcal{G}$. Then (G_k) has a convergent subsequence.

Proof sketch. For a fixed F and G_i, we have that $p(F; G_i) \in [0, 1]$. Note that $[0, 1]$ is compact and $\mathcal{G}$ is countable; by Tychonoff's theorem (Theorem 3), $[0, 1]^{\mathcal{G}}$ with the product topology is a compact metrisable space. To see this as an explicit construction, fix an arbitrary enumeration $\{F_1, F_2, \cdots\}$ of $\mathcal{G}$ and let v, w be vectors in $[0, 1]^{\mathcal{G}}$ where v_i, w_i are the ith entries of v, w. Then, D defined as follows is a metric on $[0, 1]^{\mathcal{G}}$ whose induced topology is equivalent to the product topology:

$$D(v, w) := \sum_{n=1}^{\infty} \frac{|v_n - w_n|}{2^n}.$$

Then, for a fixed G_i, we can identify the function $p(\cdot; G_i)$ with the vector $(p(F_1; G_i), p(F_2; G_i), \cdots) \in [0, 1]^{\mathcal{G}}$. In a metrisable space, every sequence has a convergent subsequence; thus the sequence $(p(\cdot; G_k))_{k\geq 1}$ has a convergent subsequence in $[0, 1]^{\mathcal{G}}$, and hence (G_k) has a convergent subsequence, as desired. $\square$

An analogous definition of convergence and argument can be made for sequences of σ-flags, using density as we have defined it for flags instead of graphs. (Such a definition will be given shortly.) Observing that $\mathcal{G} = \mathcal{F}^{\varnothing}$, we can work more generally in $\mathcal{F}^{\sigma}$.

In studying convergent sequences of graphs, we are not particularly interested in the densities of graphs in a fixed G_i, but rather in the limiting behaviour of homomorphism densities in this sequence of graphs. To this end, we can explore the notion of "limiting behaviour" in two different ways, depending on which objects we'd like to explore the limits of: For one, we could ask whether there are limit objects for sequences of graphs for which we can "read off" the information we'd like to know about homomorphism densities. This was first answered by Lovász and Szegedy in their 2004 paper [23], who concluded that such limit objects do exist, as symmetric measurable functions $W : [0,1]^2 \to [0,1]$. These are now known as *graphons* (short for "graph functions"), and a more complete theory of graphons and limits of graphs that has developed in the years since then can be found in Lovász's book [22].

On the other hand, we could look at "limiting behaviour" from the perspective of small graphs rather than the increasing sequence of graphs: Consider that for a convergent sequence of σ-flags (G_k), we can build a well-defined function $\phi : \mathcal{F}^\sigma \to \mathbb{R}$ such that $\phi(F) = \lim_{k\to\infty} p(F; G_k)$. We can then read off the density of fixed flags in the limit of a convergent sequence of flags by evaluating this function. More generally, though, if we can understand what the *entire space* of these functions looks like, we can obtain results that link back to extremal graph theory: For example, if we look at the maximum value of $\phi(\circ\!\!-\!\!\circ)$ over all convergent sequences of triangle-free graphs, we can obtain an asymptotic result about edge density in triangle-free graphs. We can give a more precise definition:

Definition 28 (Limit functionals)**.** Call $\phi : \mathcal{F}^\sigma \to \mathbb{R}$ a *limit functional* if there exists a convergent sequence $(G_k)_{k\geq 1}$ of flags in $\mathcal{F}^\sigma$ such that $\phi(F) = \lim_{k\to\infty} p(F; G_k)$, for all $F \in \mathcal{F}^\sigma$.

Define Φ to be the set of all limit functionals. In a nutshell, Razborov's theory of flag algebras gives a characterisation of Φ that lets us see more easily the relationships between homomorphism densities of small graphs in large $\mathcal{H}$-free graphs. By optimising over the space generated by Φ under this characterisation, we can then obtain asymptotic results about the density of graphs - in particular, this will help us provide a solution to the Erdős pentagon problem.

3.5 Flag Algebras

In this section, we will guide the reader through the construction of flag algebras.

We would eventually like to use the tools of mathematical optimisation to help us derive results about extremal graph theory, and to that extent, it would help if we could view the domain of a limit functional ϕ as a vector space rather than a set. Define $\mathbb{R}\mathcal{F}^\sigma$ to be the free real vector space generated by all σ-flags - that is, the set of all formal, real, finite linear combinations of σ-flags. We can now redefine limit functionals over these vector spaces by extending the function linearly: That is, if ϕ is defined on $\mathcal{F}^\sigma$, we can define ϕ on $\mathbb{R}\mathcal{F}^\sigma$ by letting $\phi(F_1 + F_2) = \phi(F_1) + \phi(F_2)$ and $\phi(cF) = c\phi(F)$, for $F_1, F_2 \in \mathcal{F}^\sigma$ and $c \in \mathbb{R}$. Thus, we can write the following modified definitions:

Definition 29. Let $(G_k)_{k\geq 1}$ be a sequence of graphs in $\mathcal{F}^\sigma$. We say that this sequence is *convergent* if, for every flag $F \in \mathcal{F}^\sigma$, $\lim\limits_{k\to\infty} p(F; G_k)$ exists. Then, call $\phi : \mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}$ a *limit functional* if ϕ is linear and there exists a convergent sequence $(G_k)_{k\geq 1}$ of flags in $\mathcal{F}^\sigma$ such that $\phi(F) = \lim\limits_{k\to\infty} p(F; G_k)$, for all $F \in \mathcal{F}^\sigma$.

It is easy to understand what linear functionals $\mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}$ look like (they are uniquely defined by their values on each element of $\mathcal{F}^\sigma$), but how can we go from here to try and understand what a limit functional on $\mathbb{R}\mathcal{F}^\sigma$ looks like? Recall that the chain rule (Theorem 23) in its simplest form tells us that for flags $F, G \in \mathcal{F}^\sigma$ and $|V(F)| \leq n \leq |V(G)|$, $p(F; G) = \sum\limits_{F' \in \mathcal{F}^\sigma_n} p(F; F')p(F'; G)$. For example, identity I2 from our asymptotic proof of Mantel's theorem was a particular example of this, stating for $\sigma = \varnothing$ and $\mathcal{H} = \{\ \}$ that $p(\ ; G) = \frac{1}{3}p(\ ; G) + \frac{2}{3}p(\ ; G)$.

Since identity I2 holds true for every graph G, it follows that if we had a convergent sequence of graphs (G_k) with an associated limit functional ϕ, it would be true that $p(\ ; G_i) = \frac{1}{3}p(\ ; G_i) + \frac{2}{3}p(\ ; G_i)$ for each $G_i \in (G_k)$ and hence, taking limits, $\phi(\ ) = \frac{1}{3}\phi(\ ) + \frac{2}{3}\phi(\ )$. By linearity, this means that $\phi(\ - \frac{1}{3}\ - \frac{2}{3}\ ) = 0$. Interestingly, this implies that the linear combination $\ - \frac{1}{3}\ - \frac{2}{3}\ $ is in the kernel of *every* limit functional ϕ.

More generally, if (G_k) is a convergent sequence of flags associated with a limit functional ϕ, it follows that for each $F \in \mathcal{F}^\sigma$, a sufficiently large $G_i \in (G_k)$, and $|V(F)| \leq n \leq |V(G_i)|$, $p(F; G_i) = \sum\limits_{F' \in \mathcal{F}^\sigma_n} p(F; F')p(F'; G)$. By taking limits, we get that

$$\phi(F) = \phi\left(\sum_{F' \in \mathcal{F}^\sigma_n} p(F; F')F'\right),$$

so that $F - \sum\limits_{F' \in \mathcal{F}^\sigma_n} p(F; F')F' \in \mathrm{Ker}(\phi)$ for every limit functional ϕ.

This is some progress towards trying to understand what limit functionals $\phi : \mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}$ look like - not every linear functional will be a limit functional, because we require that the kernel of ϕ contains particular vectors in $\mathbb{R}\mathcal{F}^\sigma$. Instead of trying to restrict our space of linear functionals by enforcing each of these infinitely many relations, though, it would be useful to look at the quotient space induced by these relations. Let $\mathcal{K}^\sigma$ be the set of all finite linear combinations of vectors in the set $\left\{ F - \sum_{F' \in \mathcal{F}_n^\sigma} p(F; F') \cdot F' : F \in \mathcal{F}^\sigma, n \geq 1 \right\}$. This is a subspace of $\mathbb{R}\mathcal{F}^\sigma$, so we can quotient it out; accordingly, let $\mathcal{A}^\sigma$ be the quotient $\mathbb{R}\mathcal{F}^\sigma / \mathcal{K}^\sigma$. Now, any vector in $\mathcal{K}^\sigma$ is in the kernel of every limit functional, so every limit functional $\mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}$ is still a linear functional $\mathcal{A}^\sigma \to \mathbb{R}$.

At this point, it seems like our job has gotten more difficult, since it's less easy to see what vectors in this new vector space look like. The benefit of passing to the quotient $\mathcal{A}^\sigma$, though, is that there is now a natural way to equip this vector space with a bilinear product so that we can take advantage of the structure of an algebra.

How might we go about defining this product? Recall that in Example 24 above, we showed that when $\mathcal{H} = \{\ \}$ and $\sigma = \bullet$, we had that $p(\ ,\ ; G) = p(\ ; G)$ for every $\bullet$-flag G. In this way, there is a relation of sorts between the two flags , and the single flag that is not encoded in $\mathcal{K}^\sigma$ (as the "basis" of relations in $\mathcal{K}^\sigma$ only relate one flag - rather than multiple flags - to a linear combination of other flags). Ideally, we'd like to encode in the definition of our product any similar relation that we could generate via the chain rule, and we can do so as follows.

Firstly, define a product $\cdot : \mathcal{F}^\sigma \times \mathcal{F}^\sigma \to \mathcal{A}^\sigma$, through supposing that F_1, F_2 are σ-flags that fit in a σ-flag of size n, and defining

$$F_1 \cdot F_2 := \sum_{F \in \mathcal{F}_n^\sigma} p(F_1, F_2; F) F + \mathcal{K}^\sigma.$$

We want to extend this to a product from $\mathcal{A}^\sigma$ to itself, but before we do this, we should make sure this product is indeed well-defined, through showing that the definition of the product does not depend on our choice of n.

Theorem 30. If F_1, F_2 are σ-flags that fit in a σ-flag of size n and $m \geq n$, then

$$\sum_{F \in \mathcal{F}_n^\sigma} p(F_1, F_2; F) F - \sum_{F \in \mathcal{F}_m^\sigma} p(F_1, F_2; F) F \in \mathcal{K}^\sigma.$$

In other words, the definition of the product does not depend on our choice of n.

Proof. By applying the chain rule (Theorem 23), it follows that

$$\begin{aligned}\sum_{F\in\mathcal{F}_n^\sigma} p(F_1,F_2;F)F &= \sum_{F\in\mathcal{F}_n^\sigma}\sum_{F'\in\mathcal{F}_m^\sigma} p(F_1,F_2;F')p(F';F)F\\ &= \sum_{F'\in\mathcal{F}_m^\sigma} p(F_1,F_2;F')\sum_{F\in\mathcal{F}_n^\sigma} p(F';F)F,\end{aligned}$$

and since $F' - \sum_{F\in\mathcal{F}_n^\sigma} p(F';F)F \in \mathcal{K}^\sigma$ for each $F'\in\mathcal{F}_m^\sigma$, it follows that

$$\sum_{F\in\mathcal{F}_n^\sigma} p(F_1,F_2;F)F = \sum_{F'\in\mathcal{F}_m^\sigma} p(F_1,F_2;F')F',$$

as desired. $\square$

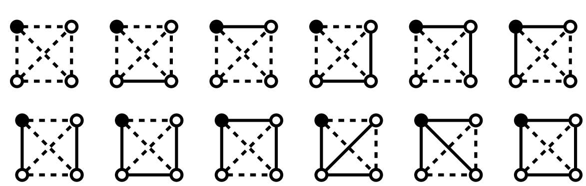
Before we continue, we can illustrate the definition of the product and the previous theorem with an example:

Example 31. Firstly, we can use our new machinery to reframe Example 24: Suppose that $\mathcal{H} = \{\ \}$, and let $\sigma = \bullet$ be the one-vertex type. The two $\bullet$-flags , fit in a $\bullet$-flag of size 3. Recall that $\mathcal{F}_3^\bullet = \{\ ,\ ,\ ,\ ,\ \}$. Then by setting $n = 3$ and using the same reasoning as in Example 24, we get that

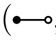
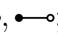
$$\cdot = \sum_{F\in\mathcal{F}_3^\bullet} p(\ ,\ ;F)F = \ ,$$

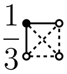
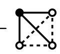
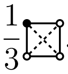
aligning with our earlier result that $p(\ ,\ ;G) = \sum_{F'\in\mathcal{F}_3^\bullet} p(\ ,\ ;F')p(F';G) = p(\ ;G)$.

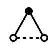
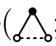
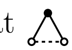
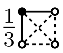
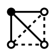
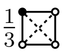
Furthermore, consider that the two $\bullet$-flags , also fit in a $\bullet$-flag of size 4, and $\mathcal{F}_4^\bullet$ consists of twelve flags, as follows.



By setting $n = 4$ and expanding the product, we then get that

 $\cdot$ $= \sum_{F \in \mathcal{F}_4^{\bullet}} p($, $; F)F$

$= \frac{1}{3}$ $+$  $+ \frac{1}{3}$.

But now observe that by definition, as  $- \sum_{F \in \mathcal{F}_4^{\bullet}} p($$; F)F \in \mathcal{K}^{\bullet}$, it follows that  $-$ $\frac{1}{3}$ $-$  $- \frac{1}{3}$ $\in \mathcal{K}^{\bullet}$, so that our two expressions for $\cdot$ differ by an element of $\mathcal{K}^{\bullet}$.

At this point, we have defined a product $\cdot : \mathcal{F}^{\sigma} \times \mathcal{F}^{\sigma} \to \mathcal{A}^{\sigma}$, but we would like to extend this to a product from $\mathcal{A}^{\sigma}$ to itself. To do this, we can first extend this product bilinearly (that is, linearly in each component), so that if $c \in \mathbb{R}$ and $f, g, h \in \mathcal{F}^{\sigma}$, then $(cf) \cdot g = f \cdot (cg) = c(f \cdot g)$, $(f + g) \cdot h = (f \cdot h) + (g \cdot h)$, and $f \cdot (g + h) = (f \cdot g) + (f \cdot h)$. In this way, this extends the product to the tensor product $(\mathbb{R}\mathcal{F}^{\sigma}) \otimes (\mathbb{R}\mathcal{F}^{\sigma}) \to \mathcal{A}^{\sigma}$, so that the generator $f \otimes g$ of $(\mathbb{R}\mathcal{F}^{\sigma}) \otimes (\mathbb{R}\mathcal{F}^{\sigma})$ would map to $f \cdot g$ under this product and the product is linear in each component. If we considered the induced product on $\mathcal{A}^{\sigma}$ defined on $f + \mathcal{K}^{\sigma}$ and $g + \mathcal{K}^{\sigma}$ instead of f and g, we can show that this induced product turns $\mathcal{A}^{\sigma}$ into a commutative associative algebra ([26], Lemma 2.4):

Theorem 32.

(a) The definition of the product on $(\mathbb{R}\mathcal{F}^{\sigma}) \otimes (\mathbb{R}\mathcal{F}^{\sigma})$ induces a symmetric bilinear product $\mathcal{A}^{\sigma} \otimes \mathcal{A}^{\sigma} \to \mathcal{A}^{\sigma}$, by letting $(f + \mathcal{K}^{\sigma}) \otimes (g + \mathcal{K}^{\sigma}) \mapsto (f \cdot g) + \mathcal{K}^{\sigma}$.

(b) If $F_1, \cdots, F_t$ are σ-flags that fit in a σ-flag of size n, then

$$((F_1 \cdot F_2) \cdot F_3 \cdot \ldots) \cdot F_t = \sum_{F \in \mathcal{F}_n^{\sigma}} p(F_1, \cdots, F_t; F)F + \mathcal{K}^{\sigma}.$$

(c) If σ is a non-degenerate type, then $\mathcal{A}^{\sigma}$ with the induced product $\mathcal{A}^{\sigma} \otimes \mathcal{A}^{\sigma} \to \mathcal{A}^{\sigma}$ is a commutative associative algebra with identity element σ.

The proof of most of these parts relies simply on repeated applications of the chain rule (Theorem 23):

Proof. (a): Firstly, observe that by construction, the density function is symmetric in all of its arguments; in particular, for σ-flags F_1, F_2, F, we will always have that

$p(F_1, F_2; F) = p(F_2, F_1; F)$. Thus, it follows that the operation $\cdot$ is symmetric. Furthermore, since we extended the product bilinearly to $(\mathbb{R}\mathcal{F}^\sigma) \otimes (\mathbb{R}\mathcal{F}^\sigma)$, it suffices to show that if $f \in \mathcal{K}^\sigma$ and $g \in \mathbb{R}\mathcal{F}^\sigma$, then $f \cdot g \in \mathcal{K}^\sigma$. Since the product as we have defined it is bilinear on $(\mathbb{R}\mathcal{F}^\sigma) \otimes (\mathbb{R}\mathcal{F}^\sigma)$, we can assume that f is of the form $F - \sum_{F' \in \mathcal{F}_n^\sigma} p(F; F')F'$ and $g = G$ is a σ-flag; the result will then hold by linearity. Thus, we want to prove that $F \cdot G = \sum_{F' \in \mathcal{F}_n^\sigma} p(F; F')(F' \cdot G) + \mathcal{K}^\sigma$.

Suppose that l is sufficiently large so that F and G fit in a σ-flag of size l, and F' and G fit in a σ-flag of size l. By the definition of the product and Theorem 30, we then have that

$$F \cdot G = \sum_{H \in \mathcal{F}_l^\sigma} p(F, G; H)H + \mathcal{K}^\sigma$$

$$\sum_{F' \in \mathcal{F}_n^\sigma} p(F; F')(F' \cdot G) = \sum_{H \in \mathcal{F}_l^\sigma} \sum_{F' \in \mathcal{F}_n^\sigma} p(F; F')p(F', G; H)H + \mathcal{K}^\sigma,$$

and by the chain rule, the desired equality follows.

(b): We induct on t. For $t = 1$, the construction of $\mathcal{K}^\sigma$ tells us that $F_1 - \sum_{F \in \mathcal{F}_n^\sigma} p(F_1, \cdots, F_t; F)F \in \mathcal{K}^\sigma$. Now suppose we knew that $((F_1 \cdot F_2) \cdot F_3 \cdot \ldots) \cdot F_{t-1} = \sum_{H \in \mathcal{F}_l^\sigma} p(F_1, \cdots, F_{t-1}; H)H + \mathcal{K}^\sigma$, for some $t \geq 2$ and $l \leq n$ where $F_1, \cdots, F_{t-1}$ fit in a σ-flag of size l. (We can assume that $l \leq n$ by Theorem 30.) Then by the definition of the product and the chain rule, it follows that

$$\begin{aligned}
((F_1 \cdot F_2) \cdot F_3 \cdot \ldots) \cdot F_t &= (((F_1 \cdot F_2) \cdot F_3 \cdot \ldots) \cdot F_{t-1}) \cdot F_t \\
&= \left(\sum_{H \in \mathcal{F}_l^\sigma} p(F_1, \cdots, F_{t-1}; H)H + \mathcal{K}^\sigma \right) \cdot F_t \\
&= \sum_{H \in \mathcal{F}_l^\sigma} p(F_1, \cdots, F_{t-1}; H)(H \cdot F_t) + \mathcal{K}^\sigma \\
&= \sum_{F \in \mathcal{F}_n^\sigma} \sum_{H \in \mathcal{F}_l^\sigma} p(F_1, \cdots, F_{t-1}; H)p(H, F_t; F)F + \mathcal{K}^\sigma \\
&= \sum_{F \in \mathcal{F}_n^\sigma} p(F_1, \cdots, F_t; F)F + \mathcal{K}^\sigma,
\end{aligned}$$

and hence induction is complete.

(c): By part (b) and because the operation $\cdot$ is symmetric, it follows that $\cdot$ is commutative and associative. Furthermore, σ is the identity element of $\mathcal{A}^\sigma$: Let F be a σ-flag. Then since the probability that the embedding of σ in F with zero other vertices induces a flag isomorphic to σ will always be 1, it follows that $p(\sigma; F) = 1$ and $p(\sigma, F_1, \cdots, F_t; F) = p(F_1, \cdots, F_t; F)$. Thus, if F fits in a flag of size n, then

$$F \cdot \sigma = \sigma \cdot F = \sum_{F' \in \mathbb{F}_n^\sigma} p(\sigma, F; F')F' = \sum_{F' \in F_n^\sigma} p(F; F')F' = F + \mathcal{K}^\sigma.$$

Finally, to show that $\mathcal{A}^\sigma$ is a nontrivial space, we need to show that $\sigma \notin \mathcal{K}^\sigma$. Let S be a finite set of relations of the form $F - \sum_{F' \in \mathcal{F}_n^\sigma} p(F, F')F'$, and let l be an upper bound for the size of all the flags that appear in all these relations. As σ is non-degenerate, we can assume that $\mathcal{F}_l^\sigma$ is nonempty, so that we can choose a $G \in \mathcal{F}_l^\sigma$. But now, if f is a linear combination of elements of S, then by considering the linear extension of the density function $p(\cdot, G)$, we have that $p(f; G) = 0$ while $p(\sigma; G) = 1$. This tells us that $\sigma \notin \operatorname{span}(S)$ and hence $\sigma \notin \mathcal{K}^\sigma$, as desired.

□

The algebra $\mathcal{A}^\sigma$ equipped with the product $\cdot$ is called the *flag algebra of type* σ.

There is a natural map between an element $f \in \mathbb{R}\mathcal{F}^\sigma$ and an element $f + \mathcal{K}^\sigma \in \mathcal{A}^\sigma$; so far, we have taken care to distinguish between cosets in $\mathcal{A}^\sigma$ and their representatives in $\mathbb{R}\mathcal{F}^\sigma$, but henceforth we will generally write elements of $\mathcal{A}^\sigma$ by writing their representatives.

Recall that our original goal was to find a way to characterise Φ, the set of all limit functionals, in such a way that they would be easier to work with. We've previously managed to show that all limit functionals are linear functionals from $\mathcal{A}^\sigma$ to $\mathbb{R}$, but we haven't managed to make much more progress beyond that. Now that we understand that $\mathcal{A}^\sigma$ carries an algebra structure (rather than just a vector space structure), we can wonder whether limit functionals are well-behaved with respect to the product we defined. Indeed, they are; in fact, we can show that limit functionals preserve the product operation and hence are homomorphisms:

Theorem 33. Let $f, g \in \mathcal{A}^\sigma$, and let $\phi : \mathcal{A}^\sigma \to \mathbb{R}$ be a limit functional associated with a convergent sequence of graphs $(G_k)_{k\geq 1}$. Then ϕ is a homomorphism from $\mathcal{A}^\sigma$ to $\mathbb{R}$: That is, it satisfies that $\phi(\sigma) = 1$ and $\phi(f \underset{\mathcal{A}^\sigma}{\cdot} g) = \phi(f) \underset{\mathbb{R}}{\cdot} \phi(g)$. Furthermore, this homomorphism is *positive*, that is, $\phi(F) \geq 0$ for every σ-flag F.

Proof. If $F \in \mathcal{F}^\sigma$, then by definition, $\phi(F) = \lim_{k\to\infty} p(F; G_k)$ must exist and be in $[0, 1]$ by the definition of density, so $\phi(F) \geq 0$ for any σ-flag F.

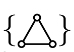
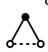
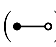
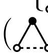
By linearity, it suffices to assume that $f = F$ and $g = G$ satisfy $F, G \in \mathbb{R}\mathcal{F}^\sigma$. Now, observe that for each $G_i \in (G_k)$, $p(\sigma, G_i) = 1$; thus it follows from taking limits that $\phi(\sigma) = 1$. Now, consider that by the definition of a limit functional,

$$\phi(F) \underset{\mathbb{R}}{\cdot} \phi(G) = \lim_{k\to\infty} p(F; G_k)p(G; G_k),$$

and by the definition of the product in $\mathcal{A}^\sigma$ and the chain rule (Theorem 23), for any n such that F and G fit in a σ-flag of size n, it follows that

$$\begin{aligned}
\phi(F \cdot G) &= \phi\left(\sum_{H\in\mathcal{F}^\sigma_n} p(F, G; H)H\right) \\
&= \sum_{H\in\mathcal{F}^\sigma_n} p(F, G; H)\phi(H) \\
&= \lim_{k\to\infty} \sum_{H\in\mathcal{F}^\sigma_n} p(F, G; H)p(H; G_k) \\
&= \lim_{k\to\infty} p(F, G; G_k).
\end{aligned}$$

By Theorem 25, though, we know that $\lim_{k\to\infty} |p(F_1, F_2; G) - p(F_1; G)p(F_2; G)| = 0$, so it follows that $\phi(f \underset{\mathcal{A}^\sigma}{\cdot} g) = \phi(f) \underset{\mathbb{R}}{\cdot} \phi(g)$ for any $f, g \in \mathcal{A}^\sigma$, as desired. $\square$

So, continuing Example 31, suppose that $\mathcal{H} = \{$$\}$ and $\sigma = \bullet$. Since we know that  $\cdot$ $=$  in $\mathcal{A}^\bullet$, it follows that $\phi($$)^2 = \phi($$)$ for every limit functional ϕ. Observe that in a way, this is a "limit formulation" of identity I5 that we constructed in our asymptotic proof of Mantel's theorem!

In general, it will be useful to develop notation for the set of homomorphisms from $\mathcal{A}^\sigma$ to the reals and the set of positive homomorphisms from $\mathcal{A}^\sigma$ to the reals: Denote the former set by $\mathrm{Hom}(\mathcal{A}^\sigma, \mathbb{R})$ and the latter set by $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$. Thus, the above theorem showed that every limit functional ϕ is in $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$. While this is certainly a nice property for limit functionals to have, it doesn't give us a complete characterisation of Φ. Surprisingly, though, the converse is also true ([23], Theorem 2.2; [26], Theorem 3.3):

Theorem 34. A linear functional $\phi : \mathcal{A}^\sigma \to \mathbb{R}$ is a limit functional if and only if it is in $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$. That is, every limit functional is a positive homomorphism, and every positive homomorphism is a limit functional.

Proof. The forward direction is shown in Theorem 33, so it suffices to show that every positive homomorphism is a limit functional. Let $\phi \in \mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$, and let $n \in \mathbb{N}$. We claim that the quantities $\phi(F)$, for $F \in \mathcal{F}_n^\sigma$, define a probability measure over $\mathcal{F}_n^\sigma$: Observe that $\sum_{F \in \mathcal{F}_n^\sigma} \phi(F) = 1$; one can view this as a generalisation of identity I1. As ϕ is linear, and therefore countably additive (note that $\mathcal{F}_n^\sigma$ is finite), it follows that ϕ is a probability measure over $\mathcal{F}_n^\sigma$. Thus, we can consider the corresponding product measure on $\prod_{n=|V(\sigma)|}^{\infty} \mathcal{F}_n^\sigma$.

Now, choose a sequence of flags $S = \{F_k\}_{k \geq 1}$, where $F_k \in \mathcal{F}_{k^2}^\sigma$ for each $k \in \mathbb{N}$, at random according to this measure. We claim that with probability 1, S converges, and ϕ is a limit functional for S. Since $\mathcal{F}^\sigma$ is countable, it suffices to prove that for every fixed $F \in \mathcal{F}_n^\sigma$ and fixed $\epsilon > 0$, with probability 1 there exists a l_0 such that for all $l \geq l_0$ and $F_k \in S$, $|p(F; F_k) - \phi(F)| \leq \epsilon$. This will show that $\phi(F) = \lim_{k \to \infty} p(F; F_k)$ with probability 1. Fix a $F \in \mathcal{F}_n^\sigma$ and an $\epsilon > 0$. We will now prove a claim:

Claim. For every $F_l \in S$ satisfying $l^2 \geq n$, $p(F; F_l)$ is a discrete random variable with expected value $\phi(F)$ and variance $O\left(\frac{1}{l^2}\right)$.

Proof. Let $F_l \in S$ satisfy $l^2 \geq n$. Since F_l was randomly chosen from $\mathcal{F}_{l^2}^\sigma$ according to our probability measure, it follows that $p(F; F_l)$ is a discrete random variable over the sample space $\mathcal{F}_{l^2}^\sigma$; label this random variable X. Every outcome $G \in \mathcal{F}_{l^2}^\sigma$ has a probability of $\phi(G)$ of occurring and a value of $p(F; G)$. Thus, we can calculate the expected value of this random variable:

$$\mathbb{E}(X) = \sum_{G \in \mathcal{F}_{l^2}^\sigma} p(F; G)\phi(G) = \phi(F),$$

where the last equality follows from the construction of $\mathcal{A}^\sigma$ (since $F - \sum_{G \in \mathcal{F}_{l^2}^\sigma} p(F; G)G \in \mathcal{K}^\sigma$).

Furthermore, through linearity of expectation and ϕ, we can calculate the variance of this random variable:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \sum_{G \in \mathcal{F}_{l^2}^\sigma} p(F; G)^2 \phi(G) - (\phi(F))^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{G \in \mathcal{F}^\sigma_{l^2}} p(F;G)^2 \phi(G) - \phi(F \cdot F) \\
&= \sum_{G \in \mathcal{F}^\sigma_{l^2}} p(F;G)^2 \phi(G) - \sum_{G \in \mathcal{F}^\sigma_{l^2}} p(F,F;G)\phi(G) \\
&\leq O\left(\frac{1}{l^2}\right) < \frac{c}{l^2},
\end{aligned}
$$

for some positive real number c, where the last line follows from Theorem 25. $\square$

Recalling that we fixed a $F \in \mathcal{F}^\sigma_n$ and an $\epsilon > 0$, our claim above says that for every $k \geq \sqrt{n}$, $p(F;F_k)$ is a discrete random variable with expected value $\phi(F)$ and variance less than $\frac{c}{k^2}$. Thus, for every $k \geq \sqrt{n}$, let X_k be the random variable given by $p(F;F_k)$ and define the event A_k by $P(|X_k - \phi(F)| > \epsilon)$.

Recall that Chebyshev's inequality (Theorem 4) states that if a random variable X has expected value μ and variance τ^2, then $P(|X - \mu| \geq m\tau) \leq \frac{1}{m^2}$. Considering $Var(X_k) \leq \frac{c}{k^2}$ for each k and substituting $m = \frac{\epsilon}{\sqrt{Var(X_k)}}$ for each X_k, it follows that $P(A_k) \leq \frac{c}{\epsilon^2 k^2}$; observing that c and ϵ are fixed constants, let $c' = \frac{c}{\epsilon^2}$. Furthermore, observe that our events are indexed over a subset of $\mathbb{N}$. Thus, considering the infinite sequence of events $\{A_k\}_{k \geq \sqrt{n}}$, it follows that

$$
\sum_{k \geq \sqrt{n}} P(A_k) < \sum_{k \geq \sqrt{n}} \frac{c'}{k^2} \leq \sum_{k=1}^{\infty} \frac{c'}{k^2} = \frac{c'\pi^2}{6} < \infty,
$$

and by the Borel-Cantelli Lemma (Theorem 5), it follows that the probability that infinitely many of these events occur is 0. Thus, it follows that there exists an l_0 such that for all $l \geq l_0$ and $F_k \in S$, $|p(F;F_k) - \phi(F)| \leq \epsilon$, and we are done. $\square$

In particular, this tells us that the linear extension of Φ is precisely $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$. Thus, we've achieved our goal of finding a nice characterisation of the set Φ, and providing another answer to the question of what limit objects of convergent sequences of graphs or flags look like.

3.6 The Downward Operator

In this section, we will introduce an operator relating flag algebras of different types.

Recall that in our asymptotic proof of Mantel's theorem, in order to obtain a result about $\varnothing$-flags (or unlabeled graphs), we had to think about the density of $\bullet$-flags (or graphs where we labeled one vertex), and relate this to the density of the corresponding $\varnothing$-flags. It turns out this is no fluke: To obtain results about unlabeled graphs, we often need to think about flags of different types.

At this point, we have found generalisations for several of the density identities we originally constructed in our asymptotic proof of Mantel's theorem, but we have not yet found generalisations of identities I3 and I4, where we considered the relationship between flags of two different types. How might we go about doing this?

Recall that identities I3 and I4 stated that

$$p(\circ\!\!-\!\!\circ; G) = \frac{1}{n} \sum_{v \in V(G)} p(\bullet\!\!-\!\!\circ; G^v)$$

and

$$\frac{1}{3} p(\ ; G) = \frac{1}{n} \sum_{v \in V(G)} p(\ ; G^v)$$

respectively. The process of deriving these two identities was nearly identical: We wanted another way to think about the subgraph count (and hence density) of a given unlabeled graph in a larger graph G. An alternate way to count the number of such subgraphs is to look at a labeled version of that subgraph (where one vertex is labeled), see how many times that labeled version appears in a labeled G^v (where a vertex $v \in V(G)$ is selected to be labeled), sum this over all choices of v, and then divide to correct for overcounting. To obtain a density relation instead of a counting relation, we would simply need to take an average instead of a sum. Still, these two identities seem quite different in one respect: The left-hand side of identity I3 has a coefficient of 1 in front of the density, and the left-hand side of identity I4 has a coefficient of $\frac{1}{3}$ in front of the density. Where does this difference come from?

From a counting perspective, one could say that this comes about as we "correct for overcounting": By fixing one vertex of G at a time and looking at the counts of the subgraph $\bullet\!\!-\!\!\circ$ over all fixed vertices, we'll count each such subgraph twice, one for each end. On the other hand, if we look at the counts of the subgraph over all fixed vertices, we'll only count each such subgraph once, because each such subgraph only has one vertex of degree two. From a density perspective, though, one could look at the numbers 1 and $\frac{1}{3}$ as the probabilities that labeling a vertex of $\circ\!\!-\!\!\circ$ and respectively induce flags isomorphic to $\bullet\!\!-\!\!\circ$ and , since we need vertices to be specifically labeled in this way in order for the original graphs to be "counted" in the right-hand side of the identities!

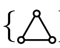
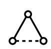
With this in mind, we can construct a generalisation of the above identities by relating elements in $\mathcal{F}^\sigma$ (and hence $\mathcal{A}^\sigma$) with their corresponding unlabeled graphs in $\mathcal{F}^\varnothing$:

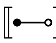
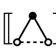
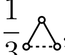
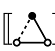
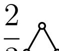
Definition 35 (Downward operator)**.** Define the *downward operator* $[\![\cdot]\!]_\sigma : \mathcal{A}^\sigma \to \mathcal{A}^\varnothing$ as the linear extension of the operator defined on $F \in \mathcal{F}^\sigma$ by

$$[\![F]\!]_\sigma = q_\sigma(F)(\downarrow F),$$

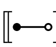
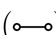
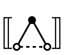
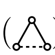
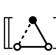
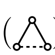
where $\downarrow F$ is the unlabeled version of F in $\mathcal{F}^\varnothing$ and $q_\sigma(F)$ is the probability that a random injective mapping $\theta : V(\sigma) \to V(\downarrow F)$ is an embedding of σ into $\downarrow F$ yielding a σ-flag isomorphic to F.

This is also sometimes called the *averaging operator*, with good reason - roughly speaking, we can see this as a generalisation of averaging the density of a σ-flag $p(F; G)$ over all possible ways to embed σ in F. We will show momentarily that $[\![\cdot]\!]_\sigma$ is indeed a linear operator, but first, we introduce an example to show what "generalised density identities" look like as relations in flag algebras:

Example 36. Let $\mathcal{H} = \{$$\}$ and $\sigma = \bullet$ be the one-vertex type. If we select a vertex of the $\varnothing$-flag at random, the probability that we will obtain a flag isomorphic to is 1. If we select a vertex of the $\varnothing$-flag  at random, the probability that we will obtain a flag isomorphic to  is $\frac{1}{3}$, and the probability that we will obtain a flag isomorphic to  is $\frac{2}{3}$. Thus, it follows that

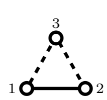
$[\![$$]\!]_\bullet =$ $,\qquad [\![$$]\!]_\bullet = \frac{1}{3}$$,\qquad [\![$$]\!]_\bullet = \frac{2}{3}$.

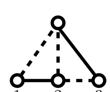
Since these relations hold between the flag algebras $\mathcal{A}^\bullet$ and $\mathcal{A}^\varnothing$, it follows that for all limit functionals ϕ,

$\phi([\![$$]\!]_\bullet) = \phi($$),\qquad \phi([\![$$]\!]_\bullet) = \frac{1}{3}\phi($$),\qquad \phi([\![$$]\!]_\bullet) = \frac{2}{3}\phi($$)$.

One can view the first two of these equations as the natural generalisations in $\mathcal{A}^\bullet$ of the density identities we constructed earlier.

Here is a slightly less trivial example. Suppose that σ and F are as follows:

$\sigma =$ 

$,\ F =$ 

Then, out of $4 \cdot 3 \cdot 2 = 24$ possible injective maps from $V(\sigma)$ to $V(\downarrow F)$, there are 8 injective maps which yield a σ-flag isomorphic to F (there are 4 ways to map 1 and 2 to an edge, and 2 ways to select vertex 3), and hence $[\![F]\!]_\sigma = \frac{1}{3}(\downarrow F)$. Accordingly, for every limit functional ϕ, it follows that $\phi([\![F]\!]_\sigma) = \frac{1}{3}\phi(\downarrow F)$.

Note that we did gloss over one technicality in the definition of the downward operator. While it is easy to extend the operator defined on $\mathcal{F}^\sigma$ linearly to $\mathbb{R}\mathcal{F}^\sigma$, it is not entirely clear why this then defines a linear mapping on the quotient $\mathcal{A}^\sigma$. We can remedy this by showing that elements in $\mathcal{K}^\sigma$ will map to elements in $\mathcal{K}^\varnothing$ ([26], Theorem 2.5):

Theorem 37. If $f \in \mathcal{K}^\sigma$, then $[\![f]\!]_\sigma \in \mathcal{K}^\varnothing$. Thus, $[\![\cdot]\!]_\sigma$ is indeed a linear mapping $\mathcal{A}^\sigma \to \mathcal{A}^\varnothing$.

Proof. By linearity, it suffices to assume that f is of the form $F - \sum_{F' \in \mathcal{F}_n^\sigma} p(F;F')F'$, where $F \in \mathcal{F}^\sigma$. Applying the downward operator to f, we get that

$$\left[\!\!\left[F - \sum_{F' \in \mathcal{F}_n^\sigma} p(F;F')F'\right]\!\!\right]_\sigma = q_\sigma(F)(\downarrow F) - \sum_{F' \in \mathcal{F}_n^\sigma} q_\sigma(F')p(F;F')(\downarrow F').$$

If we consider that we can also expand $\downarrow F$ as a linear combination of $\varnothing$-flags, namely by writing $\downarrow F = \sum_{\overline{F} \in \mathcal{F}_n^\varnothing} p(F;\overline{F})\overline{F}$, by comparing coefficients of $\overline{F}$, it then suffices to show that for a fixed $F \in \mathcal{F}^\sigma$ and $\overline{F} \in \mathcal{F}_n^\varnothing$,

$$q_\sigma(F)p(F;\overline{F}) = \sum_{F' \in \mathcal{F}_n^\sigma,\ \downarrow F' = \overline{F}} q_\sigma(F')p(F;F').$$

Indeed, suppose we choose a random injective mapping $\theta : V(\sigma) \to V(\overline{F})$, and then choose a set S of $|V(\overline{F})| - |V(\sigma)|$ vertices uniformly at random from $V(\overline{F}) \setminus \mathrm{Im}(\theta)$. The probability that $\overline{F}[\mathrm{Im}(\theta) \cup S] \cong F$ as σ-flags is then equal to both sides of the above equation, where the terms on the right side are split up by the isomorphism type of F' as a σ-flag. $\square$

3.7 Motivating the Semidefinite Method

In this section, we will provide motivation for the *semidefinite method*, a systematic way to approach problems such as Mantel's problem and the Erdős pentagon problem.

Now that we've built up some of the machinery behind flag algebras, we can return to some of the questions we first asked after giving our asymptotic proof of Mantel's theorem.

Recall that the first line of this proof stated that $0 \leq \dfrac{1}{n} \displaystyle\sum_{v \in V(G)} (1 - 2p(\bullet\!\!-\!\!\circ; G^v))^2$, and that we wondered at the time where this could have naturally come from. We can now recognise the averaging happening on the right-hand side of this inequality as corresponding to a downward operator in the respective flag algebra $\mathcal{A}^\bullet$, but what about the squaring? Razborov proved a flag-algebraic version of the Cauchy-Schwarz inequality in relation to the downward operator which allows us to generate many inequalities for limit functionals ([26], Lemma 3.14):

Theorem 38. Let $f, g \in \mathcal{A}^\sigma$. For every positive homomorphism $\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$,

$$\phi([\![f^2]\!]_\sigma)\phi([\![g^2]\!]_\sigma) \geq \phi([\![f \cdot g]\!]_\sigma)^2.$$

In particular, letting $g = \sigma$, it follows that

$$\phi([\![f^2]\!]_\sigma) \geq \phi([\![f^2]\!]_\sigma)\phi(\downarrow\!\sigma) \geq \phi([\![f]\!]_\sigma)^2 \geq 0.$$

Henceforth, our use of the phrase "Cauchy-Schwarz inequality" will reference this theorem. We will demonstrate the use of this theorem as a source of limit functional inequalities while providing another example of some of our flag algebra machinery through rephrasing our asymptotic proof of Mantel's theorem through the lens of flag algebras:

Theorem 39. Let $\mathcal{H} = \{\triangle\}$. For all positive homomorphisms $\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$,

$$\phi(\circ\!\!-\!\!\circ) \leq \frac{1}{2}.$$

Furthermore, there exists a ϕ with $\phi(\circ\!\!-\!\!\circ) = \frac{1}{2}$, so this bound is tight.

Proof 3 of Mantel's Theorem. Let $\varnothing$ be the empty flag; note that $p(\varnothing; G) = 1$ for any flag G by the definition of density and hence $\phi(\varnothing) = 1$. Also, the Cauchy-Schwarz inequality tells us that

$$\phi([\![(\varnothing - 2\bullet\!\!-\!\!\circ)^2]\!]_\bullet) \geq 0.$$

By the definition of the product in $\mathcal{A}^\bullet$, we have that

$$(\varnothing - 2\bullet\!\!-\!\!\circ)^2 = \varnothing - 4\bullet\!\!-\!\!\circ + 4\triangle,$$

and the definition of the downward operator tells us that

$$[\![\varnothing - 4\,\bullet\!\!-\!\!\circ + 4\,\overset{\bullet}{\Lambda}]\!]_{\bullet} = \varnothing - 4\,\circ\!\!-\!\!\circ + \frac{4}{3}\Lambda,$$

so by linearity it follows that

$$1 - 4\phi(\circ\!\!-\!\!\circ) + \frac{4}{3}\phi(\Lambda) \geq 0.$$

However, note that $\circ\!\!-\!\!\circ - \frac{1}{3}\dot{\underline{\ \ }} - \frac{2}{3}\Lambda \in \mathcal{K}^{\bullet}$, so that $2\phi(\circ\!\!-\!\!\circ) - \frac{2}{3}\phi(\dot{\underline{\ \ }}) - \frac{4}{3}\phi(\Lambda) = 0$. By adding this to the previous equation and noting that ϕ is positive, we see that

$$\begin{aligned} 1 - 2\phi(\circ\!\!-\!\!\circ) - \frac{2}{3}\phi(\dot{\underline{\ \ }}) &\geq 0 \\ 1 - 2\phi(\circ\!\!-\!\!\circ) &\geq 0 \\ \phi(\circ\!\!-\!\!\circ) &\leq \frac{1}{2}. \end{aligned}$$

Let $(G_k)_{k\geq 1}$ be the sequence of graphs defined by the balanced complete bipartite graphs on $2k$ vertices, that is $G_n = K_{n,n}$ for every $n \in \mathbb{N}$. This sequence is convergent: Every induced subgraph of $K_{n,n}$ is itself complete bipartite, so that the limiting density of every non-complete bipartite graph in (G_k) is zero and the limiting density of the complete bipartite graph $K_{l,m}$ is $\binom{l+m}{l}/2^{l+m-1}$ (this can be seen through assuming we have a complete bipartite graph with a large number of vertices and choosing, in turn, which part of the bipartition to place each of the $l+m$ vertices in). Then ϕ, the limit functional associated with this convergent graph sequence, satisfies $\phi(\circ\!\!-\!\!\circ) = \frac{1}{2}$, since by definition,

$$\phi(\circ\!\!-\!\!\circ) = \lim_{k\to\infty} p(\circ\!\!-\!\!\circ; G_k) = \lim_{k\to\infty} \frac{\binom{\frac{k^2}{4}}{}}{\binom{k}{2}} = \frac{1}{2}.$$

□

Though we have now translated our asymptotic, density-based proof of Mantel's theorem into a much more compact flag algebra-based proof, one element of the proof still remains a mystery: It's clear how the Cauchy-Schwarz inequality generates the first inequality in this proof, but how did we know that we were supposed to pick this particular inequality to work with? Is there a way we could arrive at such a proof which requires "less ingenuity"?

Surprisingly, it turns out that there is indeed a way to generate proofs of results such as

Mantel's theorem in an essentially mechanical manner, through viewing extremal graph theory problems as optimisation problems over the space of limit functionals. This is called the *semidefinite method*, as these optimisation problems reduce to semidefinite programming problems.

If we don't want to use the clever inequality used to start our previous two proofs of Mantel's theorem, we should go back and rethink our proof strategy. Let ϕ be a limit functional, and let $\mathcal{H} = \{\triangle\}$. Then by taking the limits of identity I2 or considering elements of the kernel of ϕ, we see that

$$\begin{aligned}
\phi(\circ\!\!-\!\!\circ) &= \frac{0}{3}\phi(\triangle) + \frac{1}{3}\phi(\triangle) + \frac{2}{3}\phi(\triangle) \\
\phi(\circ\!\!-\!\!\circ) &\leq \frac{2}{3}\phi(\triangle + \triangle + \triangle + \triangle) \\
\phi(\circ\!\!-\!\!\circ) &\leq \frac{2}{3},
\end{aligned}$$

where the last line comes as a result of taking limits in identity I1. This immediately tells us the edge density in a triangle-free graph must be less than or equal to $\frac{2}{3}$, but it isn't quite enough to get us Mantel's theorem.

Could we adapt this idea to work, though? What if, more generally, we managed to find constants $c_1, c_2, c_3 \in \mathbb{R}$ such that for all positive homomorphisms ϕ, we had that $0 \leq c_1\phi(\triangle) + c_2\phi(\triangle) + c_3\phi(\triangle)$? Then it would follow that

$$\begin{aligned}
\phi(\circ\!\!-\!\!\circ) &\leq c_1\phi(\triangle) + \left(c_2 + \frac{1}{3}\right)\phi(\triangle) + \left(c_3 + \frac{2}{3}\right)\phi(\triangle) \\
\phi(\circ\!\!-\!\!\circ) &\leq \max\left(c_1, c_2 + \frac{1}{3}, c_3 + \frac{2}{3}\right),
\end{aligned}$$

and if we can get this to be $\frac{1}{2}$ (and show, as above, that this is tight) we'd arrive at Mantel's theorem asymptotically.

It seems tough at first glance to figure out how to do this. All we've got at the moment that allows us to generate inequalities for positive homomorphisms is the Cauchy-Schwarz inequality, which tells us that a positive homomorphism applied to the square of an element in $\mathcal{A}^\sigma$ will yield a positive number. By linearity, this is also true for sums of squares of elements in $\mathcal{A}^\sigma$.

Sums of squares are nice objects that we see in a lot of other places in mathematics, and we can ask whether there's a way to use tools from other fields to approach these particular sums of squares. Considering in particular that we can really rephrase Mantel's question (asymptotically) as trying to find the maximum possible value of $\phi(\circ\!-\!\circ)$ over all possible positive homomorphisms, we might see whether sums of squares come up in mathematical optimisation, so we can use tools from that field.

Indeed, they do, in that there's a strong connection between positive semidefinite matrices and sums of squares of vectors, and solving optimisation problems over positive semidefinite matrices has been well studied. Recall that a $n \times n$ real symmetric matrix M is said to be *positive semidefinite* if $v^T M v \geq 0$ for all vectors $v \in \mathbb{R}^n$; we usually denote that M is positive semidefinite by writing $M \succeq 0$. It is well known (see, for example, [3]) that a multivariate polynomial $p(x)$ in n variables and of even degree $2d$ is a sum of squares if and only if there exists a positive semidefinite matrix Q (the Gram matrix) such that $p(x) = z^T Q z$, where z is the vector of all monomials up to degree d. It turns out a similar relation holds in $\mathcal{A}^\sigma$, in that a vector x can be written as a sum of squares if and only if there's a positive semidefinite matrix M such that $x = v^T M v$, where v is a finite vector of elements in $\mathcal{F}^\sigma$. (We will prove this in the next section, as Theorem 41.)

How could we exploit this? Given that we want to eventually say that some linear combination of , , will be nonnegative under the image of any limit functional ϕ and that sums of squares satisfy this property, we can try and find a vector v in $\mathcal{A}^\sigma$ where $v^T M v$ will be a linear combination of σ-flags for which the underlying $\varnothing$-flags are , , and . If $\sigma = \bullet$, we can generate such flags by taking various products of $\bullet\cdots\circ$ and $\bullet\!-\!\circ$. So let's try this out and see what happens when we try and reprove the first part of Theorem 39:

Proof 4 of Mantel's Theorem. Let $\sigma = \bullet$ be the one-vertex type, and let $v = (\bullet\cdots\circ, \bullet\!-\!\circ)^T$ be the vector consisting of all elements of $\mathcal{F}_2^\bullet$. By Theorem 41, we know that if $M = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$ is a 2×2 positive semidefinite matrix, then $v^T M v$ can be written as a sum of squares in $\mathcal{A}^\bullet$, and thus by the Cauchy-Schwarz inequality (Theorem 38), we will have that $\phi([\![v^T M v]\!]_\bullet) \geq 0$ for every positive homomorphism ϕ. We claim that by finding the maximum value of $\phi([\![v^T M v]\!]_\bullet)$ over all possible M, we can show that $\phi(\circ\!-\!\circ) \leq \frac{1}{2}$ for every positive homomorphism ϕ.

Consider that by the definition of the (commutative) product on $\mathcal{A}^\bullet$ and the definition of the downward operator, we can rewrite the inequality $\phi(\llbracket v^T M v \rrbracket_\bullet) \geq 0$ as

$$\begin{aligned}
\phi(\llbracket v^T M v \rrbracket_\bullet) &= \phi(\llbracket a(\ \cdot\ ) + 2c(\ \cdot\ ) + b(\ \cdot\ )\rrbracket_\bullet) \\
&= \phi(\llbracket a + a + c + c + b \rrbracket_\bullet) \\
&= \phi\left(a + \frac{a}{3} + \frac{2c}{3} + \frac{2c}{3} + \frac{b}{3} \right) \\
&= a\phi(\ ) + \frac{a+2c}{3}\phi(\ ) + \frac{b+2c}{3}\phi(\ ) \geq 0
\end{aligned}$$

for every positive homomorphism ϕ. This is an inequality of the form described in our work above, so it follows that

$$\phi(\ ) \leq \max\left(a, \frac{1+a+2c}{3}, \frac{2+b+2c}{3}\right)$$

over all possible $M \succeq 0$. But we can rewrite this as a semidefinite program:

$$\begin{aligned}
\text{minimise} \quad & \lambda \\
\text{subject to} \quad & a \leq \lambda \\
& \frac{1+a+2c}{3} \leq \lambda \\
& \frac{2+b+2c}{3} \leq \lambda \\
& M = \begin{pmatrix} a & c \\ c & b \end{pmatrix} \succeq 0.
\end{aligned}$$

This program can easily be solved using standard semidefinite programming techniques (which we will not detail in this exposition) to yield $\lambda = \frac{1}{2}$ with $M = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$, showing that $\phi(\ ) \leq \frac{1}{2}$ for all positive homomorphisms ϕ, as desired. $\square$

3.8 The Semidefinite Method

In this section, we will delve more into the theory behind the semidefinite method, and show how the previous proof of Mantel's theorem can essentially be generated automatically using the semidefinite method. We note that our exposition in this section in particular closely follows that of [5].

As we have seen in the last few sections, we can reframe extremal graph theory problems such as Mantel's theorem and the Erdős pentagon problem as optimisation problems over the space of limit functionals. For example, fixing $\mathcal{H} = \{\triangle\}$, we can rewrite Mantel's problem as the following mathematical optimisation problem:

$$\begin{array}{ll} \text{maximise} & \phi(\circ\!\!-\!\!\circ) \\ \text{subject to} & \phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R}). \end{array}$$

Any positive homomorphism ϕ will generate a *feasible solution* to this problem, since $\phi(\circ\!\!-\!\!\circ) \geq 0$ is well-defined by construction. Mantel's theorem, however, states that the *optimal solutions* to this problem are given by the positive homomorphisms ϕ for which $\phi(\circ\!\!-\!\!\circ) = \frac{1}{2}$ - that is, $\frac{1}{2}$ is the maximal value or the *optimal value* of the objective function.

More generally, let's say we're interested in looking at the maximum density of a certain fixed graph C among all $\mathcal{H}$-free graphs. In Mantel's problem, we have $\mathcal{H} = \{\triangle\}$ and $C = \circ\!\!-\!\!\circ$; in the case of the pentagon problem, we have $\mathcal{H} = \{\triangle\}$ and $C = C_5$. We can rewrite all such problems as mathematical optimisation problems:

$$\begin{array}{ll} \text{maximise} & \phi(C) \\ \text{subject to} & \phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R}). \end{array} \tag{P1}$$

One key concept in mathematical optimisation is that of *duality*. This says that optimisation problems can be viewed from two equivalent perspectives, through the lens of a *primal problem* and a *dual problem*. The "P" in our labelling of the above problem stands for "primal"; the corresponding dual problem is as follows:

$$\begin{array}{ll} \text{minimise} & \lambda \\ \text{subject to} & \phi(C) \leq \lambda \quad \forall \phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R}). \end{array} \tag{D1}$$

In general, dual problems provide bounds on the optimal solutions of primal problems; in this instance, any feasible solution to (D1) provides an upper bound for the optimal value of (P1). By construction, we also have here that the optimal values to (P1) and (D1) are equal (though this is not true for optimisation problems in general).

In order to formulate the semidefinite method, we want to relate (P1) and (D1) to a second set of optimisation problems. Let us first define a few terms:

Definition 40. Let σ be a fixed type. Define the *semantic cone* of type σ to be the set of all elements of the corresponding flag algebra $\mathcal{A}^\sigma$ which are nonnegative under the image of every positive homomorphism, namely:

$$\mathcal{S}^\sigma := \{f \in \mathcal{A}^\sigma : \phi(f) \geq 0 \quad \forall \phi \in \mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})\}.$$

Let $(\mathcal{A}^\sigma)^*$ be the dual vector space to $\mathcal{A}^\sigma$ - that is, the space of all linear maps from $\mathcal{A}^\sigma$ to $\mathbb{R}$. Then, define the *dual cone* of type σ to be the elements of $(\mathcal{A}^\sigma)^*$ which map every element of $\mathcal{S}^\sigma$ to a nonnegative real number, namely:

$$(\mathcal{S}^\sigma)^* := \{\phi \in (\mathcal{A}^\sigma)^* : \phi(f) \geq 0 \quad \forall f \in \mathcal{S}^\sigma\}.$$

It is worth noting that by Theorem 37, it follows that the image of $\mathcal{S}^\sigma$ under the downward operator is a subset of $\mathcal{S}^\varnothing$. Note that by definition, the dual cone of type σ contains all positive homomorphisms, or elements of $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$. As noted in the proof of Theorem 39, we also have that $\phi(\varnothing) = 1$ for every positive homomorphism ϕ, so it follows that

$$\max_{\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})} \phi(C) \leq \max_{\phi \in (\mathcal{S}^\varnothing)^*, \phi(\varnothing)=1} \phi(C),$$

as every ϕ satisfying the constraints on the right-hand side also satisfies the constraints on the left-hand side. We can also rewrite the right-hand side into a mathematical optimisation problem, as follows:

$$\begin{aligned} \text{maximise} \quad & \phi(C) \\ \text{subject to} \quad & \phi \in (\mathcal{S}^\varnothing)^* \\ & \phi(\varnothing) = 1. \end{aligned} \tag{P2}$$

Then the previous equation tells us that the optimal value of (P1) is less than or equal to that of (P2).

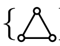
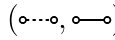
At first glance, (P2) may look like a more difficult problem to analyse than (P1). However, there is a key difference between these two problems, in that (P2) is a *conic programming problem*, as we are optimising over the intersection between a cone $(\mathcal{S}^\sigma)^*$ and a subspace generated by the relation $\phi(\varnothing) = 1$. This is actually good for us, in that we have more tools from mathematical optimisation to be able to work with conic programs than we do to be able to work with limit functionals. Still, though, it's unclear as to how we might go about working with the problem (P2) directly, and how we can reasonably do this when the optimal value of (P2) only provides a bound for the optimal value of (P1). To see this, we need to look at the dual problem to (P2), which is as follows:

$$\begin{array}{ll} \text{minimise} & \lambda \\ \text{subject to} & \lambda\varnothing - C \in \mathcal{S}^{\varnothing}. \end{array} \tag{D2}$$

Every solution to (D2) provides an upper bound to the optimal value of (P2), since if $\lambda\varnothing - C \in \mathcal{S}^{\varnothing}$, then by definition $\phi(\lambda\varnothing - C) \geq 0$, and by linearity, $\phi(C) \leq \lambda$. In fact, the optimal values of (P2) and (D2) are equal (though we will not show this here). More importantly, though, the optimal values of (D1) and (D2) are equal, since $\phi(C) \leq \lambda$ for all positive homomorphisms ϕ if and only if $\phi(\lambda\varnothing - C) \geq 0$ for all ϕ, if and only if $\lambda\varnothing - C \in \mathcal{S}^{\varnothing}$. This means that all four of our optimisation problems - (P1), (D1), (P2), and (D2) - have the same optimal value, and so we can look at our original problem, (P1), through the lens of any of these four problems.

It so happens that (D2) will be the most useful formulation of this problem for us. In its current formulation, though, it seems that solving this problem requires us to understand the global structure of the cone $\mathcal{S}^{\varnothing}$. There's no immediately apparent way to do this, but we could try and look at this cone more locally through taking a look at individual elements of this cone: More generally, in $\mathcal{S}^{\sigma}$, any *conic combination* of σ-flags - that is, a linear combination where all coefficients are nonnegative - is in $\mathcal{S}^{\sigma}$, by the definition of a positive homomorphism. But also, we know that if $f \in \mathcal{A}^{\sigma}$ is a sum of squares of elements in $\mathcal{A}^{\sigma}$ - that is, if f can be written as $g_1^2 + \cdots + g_t^2$ for elements $g_1, \cdots, g_t \in \mathcal{A}^{\sigma}$, then by the Cauchy-Schwarz inequality (Theorem 38), we will have that $\phi(f) = \phi(g_1^2) + \cdots + \phi(g_t^2) = \phi(g_1)^2 + \cdots + \phi(g_t)^2 \geq 0$ for all homomorphisms ϕ (and therefore all positive homomorphisms ϕ) and therefore $f \in \mathcal{S}^{\sigma}$. All of these elements of $\mathcal{S}^{\sigma}$ will generate a subcone of $\mathcal{S}^{\sigma}$, and if we optimise over this subcone instead, we'll get a feasible solution to (D2) which may or may generate the optimal value for (D2); either way, this in turn will provide an upper bound for the optimal value to (P1).

One might ask why it is useful here to restrict our optimisation problem to a subcone of $\mathcal{S}^{\sigma}$, especially when we are no longer guaranteed to get back the optimal value for (P1). As mentioned in the previous section, this is useful because there is a correspondence between vectors which can be written as sums of squares in $\mathcal{A}^{\sigma}$ and positive semidefinite matrices, and there is a more well-developed theory behind solving semidefinite programming problems.

We first introduce some definitions. Recall that the *size* of a flag is the number of vertices in the flag. Define the *degree* of a vector $f \in \mathbb{R}\mathcal{F}^{\sigma}$ to be the largest size of a flag appearing in the linear combination f, and by convention, define the degree of the zero vector to be -1. (This is to distinguish between the degrees of the vector $\varnothing$ and the zero vector.) We can extend this definition and define the degree of a vector $f + \mathcal{K}^\sigma \in \mathcal{A}^\sigma$ to be the minimum degree of any $g \in f + \mathcal{K}^\sigma$. Finally, for a fixed type σ and $n \geq |V(\sigma)|$, define $v_{\sigma,n}$ to be the $|\mathcal{F}_n^\sigma| \times 1$ vector of elements in $\mathcal{A}^\sigma$ enumerating the elements of $\mathcal{F}_n^\sigma$ as representatives of elements in $\mathcal{A}^\sigma$ in some fixed order. For example, if $\mathcal{H} = \{$$\}$, we can let $v_{\varnothing,2} = ($$)^T$ and $v_{\bullet,3} = ($$)^T$. Now, we can detail this correspondence between sums of squares in $\mathcal{A}^\sigma$ and positive semidefinite matrices more precisely:

Theorem 41. Let σ be a fixed type, $f \in \mathcal{A}^\sigma$, and $n \geq |V(\sigma)|$. Then there are vectors $g_1, \cdots, g_t \in \mathcal{A}^\sigma$ for some $t \geq 1$, each of degree at most n, for which $f = g_1^2 + \cdots + g_t^2$ (that is, f can be written as a sum of squares in $\mathcal{A}^\sigma$) if and only if there exists a $|\mathcal{F}_n^\sigma| \times |\mathcal{F}_n^\sigma|$ positive semidefinite matrix Q with $f = v_{\sigma,n}^T Q v_{\sigma,n}$.

We remark that that the proof of this theorem uses the fact that a symmetric matrix A is positive semidefinite if and only if it can be factored as $A = BB^T$ for some matrix B, which we proved as Theorem 1. By the properties of matrix multiplication, this is true if and only if A can be written as $A = b_1 b_1^T + \cdots + b_t b_t^T$ for some vectors $b_1, \cdots, b_t$.

Proof. ($\Rightarrow$) Suppose that we had vectors $g_1, \cdots, g_t$ of degree at most n in $\mathcal{A}^\sigma$ where $f = g_1^2 + \cdots + g_t^2$. Note that using the chain rule (Theorem 23), any flag of size less than n can be written as a linear combination of flags of size exactly n, and so it follows that there are vectors $h_1, \cdots, h_t \in \mathbb{R}\mathcal{F}^\sigma$ such that $h_i \in g_i + \mathcal{K}^\sigma$ is a linear combination of flags of size exactly n for each $1 \leq i \leq t$. Extracting the coefficients of each element of $\mathcal{F}_n^\sigma$ in the fixed order determined by $v_{\sigma,n}$ from each of the vectors h_i into a $|\mathcal{F}_n^\sigma| \times 1$ vector b_i, we can rewrite each h_i as $h_i = b_i^T v_{\sigma,n}$, for all $1 \leq i \leq t$. Then it follows that

$$\sum_{i=1}^{t} h_i^2 = \sum_{i=1}^{t} (b_i^T v_{\sigma,n})^2 = \sum_{i=1}^{t} v_{\sigma,n}^T b_i b_i^T v_{\sigma,n},$$

so we know that $Q = \displaystyle\sum_{i=1}^{t} b_i b_i^T$ is a positive semidefinite matrix by Theorem 1, and it follows that $f = v_{\sigma,n}^T Q v_{\sigma,n}$ in $\mathcal{A}^\sigma$, as desired.

($\Leftarrow$) Suppose that we had a $|\mathcal{F}_n^\sigma| \times |\mathcal{F}_n^\sigma|$ positive semidefinite matrix Q with $f = v_{\sigma,n}^T Q v_{\sigma,n}$. By Theorem 1, there are some vectors $b_1, \cdots, b_t$ so that we can write $Q = b_1 b_1^T + \cdots + b_t b_t^T$. But letting $g_i = b_i^T v_{\sigma,n}$ for every $1 \leq i \leq t$, it follows that $f = g_1^2 + \cdots + g_t^2$ and every g_i has degree at most n in $\mathcal{A}^\sigma$. $\square$

In light of the work we've done here, let's reformulate our optimisation problem: If we have a fixed $v_{\sigma,n}$ and Q, then $v_{\sigma,n}^T Q v_{\sigma,n}$ is a sum of squares (by Theorem 41) and thus is in $\mathcal{S}^\sigma$; by Theorem 37, it follows that $[\![v_{\sigma,n}^T Q v_{\sigma,n}]\!]_\sigma \in \mathcal{S}^\varnothing$. Also, as stated above, we know that any conic combination γ of $\varnothing$-flags is in $\mathcal{S}^\varnothing$. Therefore, any feasible solution to the following problem yields an upper bound for an optimal value of (P1):

$$\begin{aligned} &\text{minimise} && \lambda \\ &\text{subject to} && \lambda\varnothing - C = \gamma + [\![v_{\sigma,n}^T Q v_{\sigma,n}]\!]_\sigma \\ & && \gamma \text{ is a conic combination of } \varnothing\text{-flags} \\ & && Q \text{ is a } |\mathcal{F}_n^\sigma| \times |\mathcal{F}_n^\sigma| \text{ positive semidefinite matrix.} \end{aligned} \tag{D3}$$

Though we know how to solve semidefinite programs, at the moment, (D3) is not a semidefinite program - to be a semidefinite program, the condition "$\lambda\varnothing - C = \gamma + [\![v_{\sigma,n}^T Q v_{\sigma,n}]\!]_\sigma$", which is currently an identity on elements of $\mathcal{A}^\varnothing$, needs to be a linear constraint (or linear constraints) on λ and the entries of Q. Furthermore, to use our semidefinite programming tools, we need to find some way to not additionally deal with optimising over the space of conic combinations of $\varnothing$-flags.

Let's try and rewrite our condition first. The space of $n \times n$ matrices is endowed with the inner product $\langle A, B\rangle = tr(A^T B) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}$. Considering that Q is a matrix of real numbers, we can then write:

$$[\![v_{\sigma,n}^T Q v_{\sigma,n}]\!]_\sigma = [\![\langle v_{\sigma,n} v_{\sigma,n}^T, Q\rangle]\!]_\sigma = \langle [\![v_{\sigma,n} v_{\sigma,n}^T]\!]_\sigma, Q\rangle,$$

where the downward operator applied to a matrix signifies that the operator is applied pointwise to every entry of the matrix. Thus, our first condition in (D3) now reads $\lambda\varnothing - C = \gamma + \langle [\![v_{\sigma,n} v_{\sigma,n}^T]\!]_\sigma, Q\rangle$. Furthermore, this is a condition in $\mathcal{A}^\varnothing$, and to work with this more concretely we can lift this to $\mathbb{R}\mathcal{F}^\varnothing$ by choosing a representative for C from $C + \mathcal{K}^\varnothing$ and a representative for every element of $\mathcal{A}^\varnothing$ in the matrix $[\![v_{\sigma,n} v_{\sigma,n}^T]\!]_\sigma$ from their respective cosets.

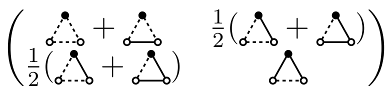
To demonstrate what this looks like in practice, we'll now provide an example that will later help us systematically recreate our previous proof of Mantel's theorem (recalling that we are proving the first part of the statement from Theorem 39 by showing that $\phi(\circ\!\!-\!\!\circ) \leq \frac{1}{2}$ for all positive homomorphisms $\phi \in \mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$):

Example 42. Let $\sigma = \bullet$ be the one-vertex type and $v = v_{\bullet,2} = (\bullet\cdots\circ, \bullet\!\!-\!\!\circ)^T$ be the vector enumerating all elements of $\mathcal{F}_2^\bullet$. Then any feasible solution to the following program yields an upper bound for the maximum possible value of $\phi(\circ\!\!-\!\!\circ)$ over all positive homomorphisms

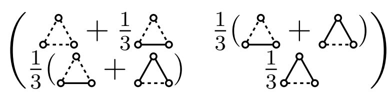
$\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$:

$$\begin{array}{rl} \text{minimise} & \lambda \\ \text{subject to} & \lambda\varnothing - C = \gamma + \langle [\![vv^T]\!]_\bullet, Q\rangle \\ & \gamma \text{ is a conic combination of } \varnothing\text{-flags} \\ & Q \text{ is a } |\mathcal{F}_2^\bullet| \times |\mathcal{F}_2^\bullet| \text{ positive semidefinite matrix.} \end{array}$$

Now, by applying the definition of the product in $\mathcal{A}^\bullet$ (as in Example 31) and choosing the representatives in $\mathbb{R}\mathcal{F}^\sigma$ of smallest size, we get that

$$vv^T =$$  ,

and by applying the downward operator pointwise to each entry of this matrix (following Example 36), we get that

$$[\![vv^T]\!]_\bullet =$$  .

We will now take a pause from this example as we figure out how we might proceed in our recreation of our previous proof of Mantel's theorem. At the moment, we're working with representatives in $\mathbb{R}\mathcal{F}^\varnothing$, but we need the identity $\lambda\varnothing - C = \gamma + \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle$ to hold in $\mathcal{A}^\varnothing$. This can be remedied, though: Firstly, let us extend the density function $p(F;G)$ linearly to $\mathbb{R}\mathcal{F}^\sigma$ (so that $p(F_1 + F_2; G) = p(F_1; G) + p(F_2; G)$ and $p(cF; G) = cp(F; G)$ for flags F_1, F_2 and constant $c \in \mathbb{R}$). Then, we can show that our identity on vectors in $\mathcal{A}^\varnothing$ will be satisfied if a set of analogous relations holds in $\mathbb{R}\mathcal{F}^\varnothing$ for densities over every flag G in a sufficiently large $\mathcal{F}_N^\varnothing$:

Proposition 43. Suppose that $p(\lambda\varnothing - C; G) = p(\gamma; G) + p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$ for every $G \in \mathcal{F}_N^\varnothing$, where some $N > 0$ is fixed. Then $\lambda\varnothing - C = \gamma + \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle$ in $\mathcal{A}^\varnothing$.

Proof. Let $\{H_i\}_{i\geq 1}$ be a convergent sequence of graphs. Then, by expanding using the chain rule (Theorem 23), it follows that

$$p(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; H_i) = \sum_{G \in \mathcal{F}_N^\varnothing} p(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)p(G; H_i)$$

for every $i \in \mathbb{N}$. Since $\{H_i\}$ is convergent, we can take the limit of both sides, yielding that

$$\phi(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle) = \sum_{G\in\mathcal{F}_N^\varnothing} p(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)\phi(G)$$

for every limit functional ϕ. But every term on the sum on the right-hand side is zero, since by hypothesis, $p(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G) = 0$ for every $G \in \mathcal{F}_N^\varnothing$. Furthermore, by construction, we have that

$$\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle - \sum_{G\in\mathcal{F}_N^\varnothing} p(\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)G \in \mathcal{K}^\sigma.$$

Hence it follows that $\lambda\varnothing - C - \gamma - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle \in \mathcal{K}^\sigma$ and thus $\lambda\varnothing - C = \gamma + \langle [\![vv^T]\!]_\sigma, Q\rangle$ in $\mathcal{A}^\varnothing$, as desired. □

Recall that we also wanted to find some way to do away with γ - namely, to avoid dealing with also optimising over the space of conic combinations of $\varnothing$-flags. Working in $\mathbb{R}\mathcal{F}^\sigma$ allows us to do this, because if γ is a conic combination of $\varnothing$-flags, then $p(\gamma; G) \geq 0$ for every $G \in \mathcal{F}^\sigma$, because density is always nonnegative!

Bearing all this in mind, let's rewrite our constraint from (D3) one last time. Recall that the original identity we wanted to satisfy in $\mathcal{A}^\varnothing$ is

$$\lambda\varnothing - C = \gamma + \langle [\![vv^T]\!]_\sigma, Q\rangle,$$

for some conic combination γ of $\varnothing$-flags and some positive semidefinite matrix Q. We proved in Proposition 43 that we can show this identity is satisfied through showing that the relation

$$p(\lambda\varnothing - C; G) = p(\gamma; G) + p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$$

is satisfied for every $G \in \mathcal{F}_N^\varnothing$, where $N > 0$ is fixed. Because $p(\gamma; G) \geq 0$ as γ is a conic combination of $\varnothing$-flags, it follows that if this identity is satisfied, then

$$p(\lambda\varnothing - C; G) \geq p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$$

for every $G \in \mathcal{F}_N^\varnothing$. The converse is also true, but this requires a short proof:

Proposition 44. Let $N > 0$ be fixed. If $p(\lambda\varnothing - C; G) \geq p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$ for every $G \in \mathcal{F}_N^\varnothing$, then there exists a conic combination γ of $\varnothing$-flags such that $p(\lambda\varnothing - C; G) = p(\gamma; G) + p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$ for every $G \in \mathcal{F}_N^\varnothing$.

Proof. Let $F_N^{\varnothing}$ be enumerated as $\{G_1, \cdots, G_k\}$. By hypothesis and linearity, it follows that there exist nonnegative constants $\alpha_1, \cdots, \alpha_k$ such that

$$p(\lambda\varnothing - C - \langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G_i) = \alpha_i \geq 0$$

for every $i \in [k]$. Now, let $\gamma = \sum_{i=1}^{k} \alpha_i G_i$. This is a conic combination of $\varnothing$-flags, and $p(\gamma; G_i) = \alpha_i p(G_i; G_i) = \alpha_i$ for every $i \in [k]$, so the result follows. $\square$

Finally, for a matrix M with entries in $\mathbb{R}\mathcal{F}^\sigma$, let $p(M; G)$ denote the matrix generated by applying the (linearly extended) density function $p(F; G)$ to every entry in M. By linearity and through noting that $p(\varnothing; G) = 1$, we get that

$$p(\lambda\varnothing - C; G) = p(\gamma; G) + p(\langle [\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma, Q\rangle; G)$$

for every $G \in \mathcal{F}_N^{\varnothing}$ if and only if

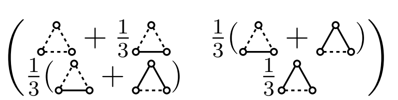
$$\lambda \geq p(C; G) + \langle p([\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma; G), Q\rangle$$

for every $G \in \mathcal{F}_N^{\varnothing}$. This is now a finite number of linear constraints on λ and the entries of Q, which is what we wanted to rewrite (D3) into a semidefinite program! Therefore, any feasible solution to the following semidefinite program yields an upper bound for an optimal value of (P1):

$$\begin{aligned} &\text{minimise} && \lambda \\ &\text{subject to} && \lambda \geq p(C; G) + \langle p([\![v_{\sigma,n}v_{\sigma,n}^T]\!]_\sigma; G), Q\rangle \quad \forall G \in \mathcal{F}_N^{\varnothing}, N > 0 \text{ fixed} \\ & && Q \text{ is a } |\mathcal{F}_n^\sigma| \times |\mathcal{F}_n^\sigma| \text{ positive semidefinite matrix.} \end{aligned} \tag{D4}$$

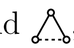
Let's see what this looks like in practice through finishing our recreation of our previous proof of Mantel's Theorem:

Proof 5 of Mantel's Theorem. Recall from Example 42 that we defined $v = v_{\bullet,2} = (,)^T$, and that

$[\![vv^T]\!]_\bullet =$ 

Rewriting our program as described above, any feasible solution to the following program yields an upper bound for the maximum possible value of $\phi( )$ over all positive homomorphisms $\phi \in \mathrm{Hom}^+(\mathcal{A}^{\varnothing}, \mathbb{R})$:

$$\begin{array}{ll} \text{minimise} & \lambda \\ \text{subject to} & \lambda \geq p(\circ\!\!-\!\!\circ; G) + \langle p([\![vv^T]\!]_\bullet; G), Q \rangle \quad \forall G \in \mathcal{F}^{\varnothing}_N, N > 0 \text{ fixed} \\ & Q \text{ is a } |\mathcal{F}^\bullet_2| \times |\mathcal{F}^\bullet_2| \text{ positive semidefinite matrix.} \end{array}$$

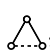
Let us write this program out in more detail. Let $Q = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$ be a positive semidefinite matrix. We will choose $N = 3$ here. There are three flags in $\mathcal{F}^{\varnothing}_3$: , , and , so we will get three linear constraints in our semidefinite program, as follows:

- For $G =$ , we see that $p(\circ\!\!-\!\!\circ; G) = 0$ and $p([\![vv^T]\!]_\bullet; G) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so the linear constraint generated here is

$$\lambda \geq \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, Q \right\rangle = a.$$

- For $G =$ , we see that $p(\circ\!\!-\!\!\circ; G) = \frac{1}{3}$ and $p([\![vv^T]\!]_\bullet; G) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix}$, so the linear constraint generated here is

$$\lambda \geq \frac{1}{3} + \left\langle \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix}, Q \right\rangle = \frac{1 + a + 2c}{3}.$$

- For $G =$ , we see that $p(\circ\!\!-\!\!\circ; G) = \frac{2}{3}$ and $p([\![vv^T]\!]_\bullet; G) = \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}$, so the linear constraint generated here is

$$\lambda \geq \frac{2}{3} + \left\langle \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}, Q \right\rangle = \frac{2 + b + 2c}{3}.$$

This generates the same semidefinite program as before:

$$
\begin{array}{rl}
\text{minimise} & \lambda \\
\text{subject to} & \lambda \geq a \\
& \lambda \geq \dfrac{1+a+2c}{3} \\
& \lambda \geq \dfrac{2+b+2c}{3} \\
& Q = \begin{pmatrix} a & c \\ c & b \end{pmatrix} \succeq 0,
\end{array}
$$

so it follows that $\phi(\circ\!\!-\!\!\circ) \leq \frac{1}{2}$ for all $\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$, as desired. $\square$

In general, we note that we are not restricted to working with a single type. Because $[\![v_{\sigma,n}^T Q v_{\sigma,n}]\!]_\sigma \in \mathcal{S}^\varnothing$ for *any* given σ and n, it follows that we could actually use multiple types $\sigma_1, \cdots, \sigma_t$ with associated $n_i \geq |V(\sigma_i)|$, $1 \leq i \leq t$, to write (D3) as follows:

$$
\begin{array}{rl}
\text{minimise} & \lambda \\
\text{subject to} & \lambda\varnothing - C = \gamma + \displaystyle\sum_{i=1}^{t} [\![v_{\sigma_i,n_i}^T Q_i v_{\sigma_i,n_i}]\!]_{\sigma_i} \\
& \gamma \text{ is a conic combination of } \varnothing\text{-flags, } 1 \leq i \leq t \\
& Q_i \text{ is a } |\mathcal{F}_{n_i}^{\sigma_i}| \times |\mathcal{F}_{n_i}^{\sigma_i}| \text{ positive semidefinite matrix, } 1 \leq i \leq t.
\end{array}
\tag{D3'}
$$

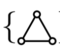
Following the same procedure as above, we can then rewrite (D4) as follows:

$$
\begin{array}{rl}
\text{minimise} & \lambda \\
\text{subject to} & \lambda \geq p(C;G) + \langle p([\![v_{\sigma_i,n_i} v_{\sigma_i,n_i}^T]\!]_{\sigma_i}; G), Q_i \rangle \quad \forall G \in \mathcal{F}_{N_i}^\varnothing, N_i > 0 \text{ fixed, } 1 \leq i \leq t \\
& Q_i \text{ is a } |\mathcal{F}_{n_i}^{\sigma_i}| \times |\mathcal{F}_{n_i}^{\sigma_i}| \text{ positive semidefinite matrix, } 1 \leq i \leq t.
\end{array}
\tag{D4'}
$$

In general, using more types will obtain a better bound for the optimal value of (P1), but will be computationally more difficult, as one needs to consider more constraints and optimise over the entries of more semidefinite matrices. Sometimes, though, it is necessary to use more types to achieve a tight bound, as we will see in the next section.

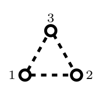
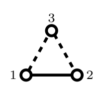
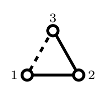
3.9 A Solution to the Pentagon Problem

In this final section, we will show how using the semidefinite method as described above generates a solution to the Erdős pentagon problem: First, we will use the semidefinite method to find an upper bound for the value of $\phi(C_5)$ in the space of triangle-free graphs. Then, we will show how this implies that the Erdős pentagon problem can be answered in the affirmative. We note that this solution was first found by Grzesik [11], and many notation choices here align with that used in his paper.

Theorem 45. Let $\mathcal{H} = \{$$\}$. For all positive homomorphisms $\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$,

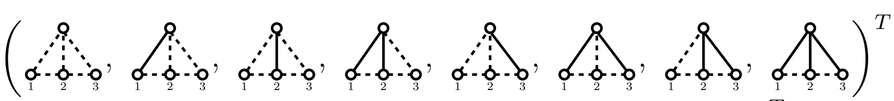
$$\phi(C_5) \leq \frac{24}{625}.$$

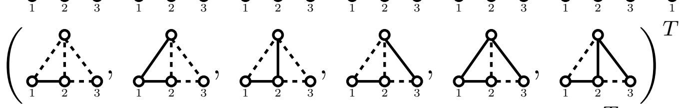
Proof. We will use the semidefinite method to prove this theorem. Consider three types $\sigma_0, \sigma_1, \sigma_2$ on the vertex set $\{1, 2, 3\}$, as follows:

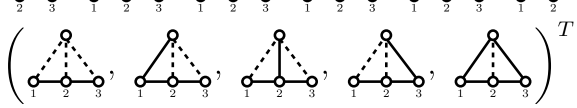
$\sigma_0 =$ , $\sigma_1 =$ , $\sigma_2 =$ .

That is, σ_0 is the type with no edges, σ_1 is the type with the edge $\{12\}$, and σ_2 is the type with two edges $\{12, 23\}$. We will choose $n_{\sigma_0} = n_{\sigma_1} = n_{\sigma_2} = 4$, noticing that $4 \geq |V(\sigma_i)|$ in each instance, and $N_0 = N_1 = N_2 = 5$.

Now, note that $|\mathcal{F}_4^{\sigma_0}| = 8$, $|\mathcal{F}_4^{\sigma_1}| = 6$, and $|\mathcal{F}_4^{\sigma_2}| = 5$ - namely, there are eight σ_0-flags of size four, six σ_1-flags of size four, and five σ_2-flags of size four. Thus, we can construct the following vectors:

$v_0 =$ 

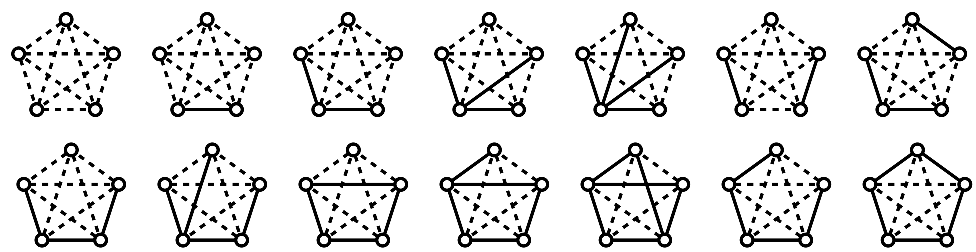
$v_1 =$ 

$v_2 =$ .

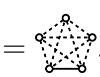
(We will not draw the non-edge 13 into any of these flags, since 13 is not an edge in any of $\sigma_0, \sigma_1, \sigma_2$.) Henceforth, we may refer to the ith entry of v_j as "the ith σ_j-flag". Now, following the semidefinite method above, we want to show that $\frac{24}{625}$ is a feasible value for the following semidefinite program:

$$\begin{array}{rl} \text{minimise} & \lambda \\ \text{subject to} & \lambda \geq p(C_5; G) \\ & \quad + \langle p([\![v_0 v_0^T]\!]_{\sigma_0}; G), P\rangle \\ & \quad + \langle p([\![v_1 v_1^T]\!]_{\sigma_1}; G), Q\rangle \\ & \quad + \langle p([\![v_2 v_2^T]\!]_{\sigma_2}; G), R\rangle \quad \forall G \in \mathcal{F}_5^{\varnothing} \\ & P, Q, R \succeq 0. \end{array}$$

Note that by our definition of v_0, v_1, v_2, we will have that P is an 8×8 matrix, Q is a 6×6 matrix, and R is a 5×5 matrix. Henceforth, we will use the notation P_{ij}, Q_{ij}, R_{ij} to denote entry ij of the matrices P, Q, R respectively as determined by our fixed ordering of $\mathcal{F}_4^{\sigma_i}$ in v_0, v_1, v_2. There will be 14 linear constraints in this program, each one corresponding to a different element of $\mathcal{F}_5^{\varnothing}$, the graphs in which can be enumerated as follows.



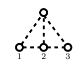
We will not go through the process of manually computing each of the constraints in this exposition via computing the matrices $[\![v_0 v_0^T]\!]_{\sigma_0}, [\![v_1 v_1^T]\!]_{\sigma_1}, [\![v_2 v_2^T]\!]_{\sigma_2}$, but we will demonstrate a more strategic computation of one of these constraints as an example.

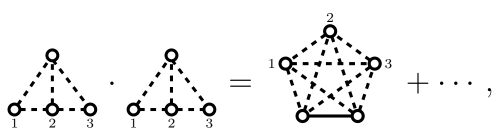
More specifically, we will compute the constraint for when $G =$ , the five-vertex graph with one edge. Clearly, $p(C_5; G) = 0$ in this instance. Then, we can ask about which entries of the matrices $[\![v_0 v_0^T]\!]_{\sigma_0}, [\![v_1 v_1^T]\!]_{\sigma_1}, [\![v_2 v_2^T]\!]_{\sigma_2}$ will contain a term with the flag G, and thus which entries of the matrices P, Q, R will appear with nonzero coefficients in our constraint. By definition, an entry of a matrix $[\![v_i v_i^T]\!]_{\sigma_i}$ will contain a term with the flag G if and only if we can place the two corresponding σ_i-flags inside G in such a way that the fourth vertices of the two flags are distinct.

Since G has only one edge, we cannot place two σ_2-flags inside G, since σ_2 as a type has two edges. It is possible to place two σ_1-flags inside G, but since σ_1 as a type has one edge already, the two flags cannot contain any edges that are not between vertices in the type, so we can only place two copies of the first σ_1-flag inside G. Finally, we can also place two

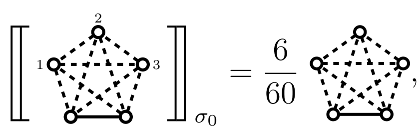
σ_0-flags inside G, but since G has only one edge and σ_0 as a type has no edges, only one of the flags we place can contain an edge, and the flag that contains an edge must contain exactly one edge. Thus, we can place two copies of the first σ_0-flag inside G (by letting the edge in G go between the fourth vertices of the two copies of the flag), or a copy of the first σ_0-flag and a copy of the second, third, or fifth σ_0-flags inside G.

Thus, the only entries of the matrices which will appear in our constraint are p_{11}, p_{12}, p_{13}, p_{15}, and q_{11}. (The entries p_{21}, p_{31}, and p_{51} will appear as well, but positive semidefinite matrices are symmetric by definition, so that $p_{12} = p_{21}$, $p_{13} = p_{31}$, and $p_{15} = p_{51}$.) This essentially leaves us with three cases to deal with:

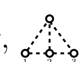
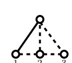
- First, we have the case where we try to place two copies of the first σ_0-flag, , inside G. The only way we can do this is to map the three vertices of σ_0 to the three vertices of degree zero. Thus, by the definition of the product, the corresponding entry of $v_0 v_0^T$ will be

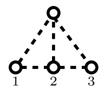
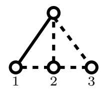
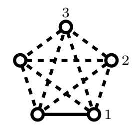
  

  and applying the downward operator, we get that the corresponding entry of $[\![v_0 v_0^T]\!]_{\sigma_0}$ will be

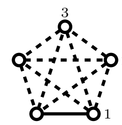
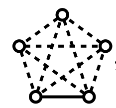
  

  as of the $5 \cdot 4 \cdot 3 = 60$ injective maps from $\{1, 2, 3\}$ to $V(G)$, there will be 3! maps which map all of $\{1, 2, 3\}$ to vertices of degree zero. Thus, the corresponding term of $\langle p([\![v_0 v_0^T]\!]_{\sigma_0}; G), P\rangle$ that will appear in our constraint is $\frac{6}{60} p_{11}$.

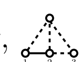
- Next, we have the case where we try and place one copy of the first σ_0-flag, , and one copy of the second σ_0-flag, , inside G. (The cases where we try and place one copy of the first σ_0 flag and one copy of the third or fifth σ_0-flag are identical by symmetry.) The only way we can do this is to map vertex 1 in σ_0 to a vertex of degree 1 in G, and map vertices 2 and 3 in σ_0 to a vertex of degree 0 in G. Thus, by the definition of the product, the corresponding entry of $v_0 v_0^T$ will be

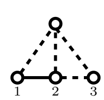
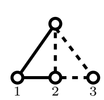
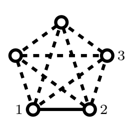
 $\cdot$  $= \dfrac{1}{2}$  $+ \cdots ,$

and applying the downward operator, we get that the corresponding entry of $[\![v_0 v_0^T]\!]_{\sigma_0}$ will be

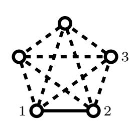
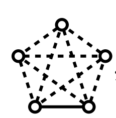
$\dfrac{1}{2} \Big[\!\!\Big[$  $\Big]\!\!\Big]_{\sigma_0} = \dfrac{1}{2} \cdot \dfrac{12}{60}$ ,

as of the 60 injective maps from $\{1, 2, 3\}$ to $V(G)$ we can choose a vertex of degree 1 to map the vertex 1 to in 2 ways, and vertices of degree 0 to map the vertices 2 and 3 to in $3 \cdot 2 = 6$ ways. Thus, the corresponding term of $\langle p([\![v_0 v_0^T]\!]_{\sigma_0}; G), P\rangle$ that will appear in our constraint is $\frac{12}{120} p_{12}$. Noting that we could place two flags in G in two different orders (corresponding to two different products), though, we will also get a term of $\frac{12}{120} p_{21}$ appearing in our constraint. Thus, by the symmetry of P and considering that we could use the third or fifth σ_0-flag instead of the second σ_0-flag, the terms $\frac{24}{120} p_{13}$ and $\frac{24}{120} p_{15}$ will also appear in $\langle p([\![v_0 v_0^T]\!]_{\sigma_0}; G), P\rangle$.

- Finally, we have the case where we try and place two copies of the first σ_1-flag, , inside G. The only way we can do this is to map vertices 1 and 2 in σ_1 to the vertices of degree 1 in G, and map vertex 3 in σ_1 to a vertex of degree 0 in G. Thus, by the definition of the product, the corresponding entry of $v_1 v_1^T$ will be

 $\cdot$  $=$  $+ \cdots ,$

and applying the downward operator, we get that the corresponding entry of $[\![v_1 v_1^T]\!]_{\sigma_1}$ will be

$\Big[\!\!\Big[$  $\Big]\!\!\Big]_{\sigma_1} = \dfrac{6}{60}$ ,

as of the 60 injective maps from $\{1, 2, 3\}$ to $V(G)$, there are 2! ways to map the vertices 1 and 2 to vertices of degree 1 in G, and 3 ways to map vertex 3 to a vertex of degree 0 in G. Thus, the corresponding term of $\langle p([\![v_1 v_1^T]\!]_{\sigma_1}; G), Q\rangle$ that will appear in our constraint is $\frac{6}{60} q_{11}$.

Thus, it follows that the constraint corresponding to $G \in \mathcal{F}_5^{\varnothing}$ reads

$$\lambda \geq \frac{1}{120}(12p_{11} + 24p_{12} + 24p_{13} + 24p_{15} + 12q_{11}).$$

By computing the remaining constraints, either in a similar way to the above or through manually computing the matrices $[\![v_0v_0^T]\!]_{\sigma_0}, [\![v_1v_1^T]\!]_{\sigma_1}, [\![v_2v_2^T]\!]_{\sigma_2}$, we can rewrite our semidefinite program as follows, where the constraints correspond to the flags in $\mathcal{F}_5^{\varnothing}$ in the order given above:

$$\begin{aligned}
\text{minimise} \qquad & \lambda \\
\text{subject to} \qquad & \lambda \geq \frac{1}{120}(120p_{11}) \\
& \lambda \geq \frac{1}{120}(12p_{11} + 24p_{12} + 24p_{13} + 24p_{15} + 12q_{11}) \\
& \lambda \geq \frac{1}{120}(8p_{12} + 8p_{13} + 8p_{14} + 8p_{15} + 8p_{16} + 8p_{17} + 4p_{22} + 4p_{33} + 4p_{55} + \\
& \quad + 8q_{12} + 8q_{13} + 4r_{11}) \\
& \lambda \geq \frac{1}{120}(12p_{14} + 12p_{16} + 12p_{17} + 12p_{18} + 6q_{22} + 6q_{33} + 12r_{13}) \\
& \lambda \geq \frac{1}{120}(48p_{18} + 24r_{33}) \\
& \lambda \geq \frac{1}{120}(16p_{23} + 16p_{25} + 16p_{35} + 8q_{11} + 16q_{14}) \\
& \lambda \geq \frac{1}{120}(8p_{27} + 8p_{36} + 8p_{45} + 8q_{14} + 8q_{24} + 8q_{34} + 4q_{44} + 4r_{11}) \\
& \lambda \geq \frac{1}{120}(4p_{23} + 4p_{24} + 4p_{25} + 4p_{26} + 4p_{34} + 4p_{35} + 4p_{37} + 4p_{56} + 4p_{57} + \\
& \quad + 4q_{12} + 4q_{13} + 4q_{15} + 4q_{16} + 4q_{23} + 4r_{12} + 4r_{14}) \\
& \lambda \geq \frac{1}{120}(4p_{27} + 4p_{28} + 4p_{36} + 4p_{38} + 4p_{45} + 4p_{58} + 4q_{15} + 4q_{16} + 4q_{25} + \\
& \quad + 4q_{36} + 4r_{13} + 2r_{22} + 4r_{23} + 4r_{34} + 2r_{44}) \\
& \lambda \geq \frac{1}{120}(8p_{44} + 8p_{66} + 8p_{77} + 16q_{23} + 16r_{15}) \\
& \lambda \geq \frac{1}{120}(4p_{48} + 4p_{68} + 4p_{78} + 4q_{26} + 4q_{35} + 2q_{55} + 2q_{66} + 4r_{15} + 4r_{23} + \\
& \quad + 4r_{25} + 4r_{34} + 4r_{35} + 4r_{45})
\end{aligned}$$

$$
\begin{aligned}
&\lambda \geq \frac{1}{120}(12p_{88} + 24r_{35} + 12r_{55}) \\
&\lambda \geq \frac{1}{120}(4p_{46} + 4p_{47} + 4p_{67} + 4q_{24} + 4q_{26} + 4q_{34} + 4q_{35} + 4q_{45} + 4q_{46} + \\
&\qquad + 4r_{12} + 4r_{14} + 4r_{24}) \\
&\lambda \geq 1 + \frac{1}{120}(20q_{56} + 20r_{24}) \\
&P, Q, R \succeq 0
\end{aligned}
$$

This semidefinite program can be solved to yield the positive semidefinite matrices

$$
P = \frac{1}{625}\begin{pmatrix}
24 & -36 & -36 & 24 & -36 & 24 & 24 & -36 \\
-36 & 277 & 97 & -79 & 97 & -79 & -259 & 54 \\
-36 & 97 & 277 & -79 & 97 & -259 & -79 & 54 \\
24 & -79 & -79 & 247 & -259 & 67 & 67 & -36 \\
-36 & 97 & 97 & -259 & 277 & -79 & -79 & 54 \\
24 & -79 & -259 & 67 & -79 & 247 & 67 & -36 \\
24 & -259 & -79 & 67 & -79 & 67 & 247 & -36 \\
-36 & 54 & 54 & -36 & 54 & -36 & -36 & 54
\end{pmatrix},
$$

$$
Q = \frac{1}{2500}\begin{pmatrix}
1728 & -1551 & -1551 & -1308 & 687 & 687 \\
-1551 & 2336 & 742 & 908 & 2557 & -4084 \\
-1551 & 742 & 2336 & 908 & -4084 & 2557 \\
-1308 & 908 & 908 & 1728 & -254 & -254 \\
687 & 2557 & -4084 & -254 & 15264 & -14424 \\
687 & -4084 & 2557 & -254 & -14424 & 15264
\end{pmatrix},
$$

$$
R = \frac{1}{625}\begin{pmatrix}
1512 & 568 & -380 & 568 & -376 \\
568 & 475 & -191 & 0 & -93 \\
-380 & -191 & 192 & -191 & -2 \\
568 & 0 & -191 & 475 & -93 \\
-376 & -93 & -2 & -93 & 190
\end{pmatrix},
$$

and $\lambda = \frac{24}{625}$, as desired. □

Finally, we can use the previous theorem to yield a solution to the Erdős pentagon problem:

Theorem 46. Let G be a triangle-free graph on n vertices. Then $c(C_5; G) \leq \left(\frac{n}{5}\right)^5$.

Proof. By way of contradiction, suppose there exists a triangle-free graph G on n vertices with $c(C_5;G) = \left(\frac{n}{5}\right)^5 + \alpha > \left(\frac{n}{5}\right)^5$, for some $\alpha > 0$. Consider the sequence of graphs $(G_k)_{k\geq 1}$ given by $G_m = G^{(m)}$ for all $m \in \mathbb{N}$ - that is, let G_n be the balanced blow-up of G where every vertex of G is replaced by an independent set of m vertices and every edge of G is replaced by a complete bipartite graph $K_{m,m}$. Observe that by construction, every graph in the sequence (G_k) is triangle-free, and the graph G_m will have nm vertices and $c(C_5;G_m) \geq \left(\left(\frac{n}{5}\right)^5 + \alpha\right) m^5$, as one way to choose a 5-cycle in G_m is to first choose a 5-cycle in G, then choose a 5-cycle in G_m by selecting one vertex from each of the corresponding independent sets in G_m.

It is well-known that such a sequence of blow-up graphs is convergent (see, for example, [4]). For some brief intuition as to why this is true, consider that for a fixed graph H, $p(H;G_k)$ converges to the probability that the vertices of H can be "coloured" with numbers corresponding to vertices in $V(G)$ such that vertices of a given colour v are only adjacent to vertices with colours corresponding to vertices in $N_G(v)$. Thus, there is a limit functional ϕ associated with the sequence (G_k).

But now, consider that $\phi(C_5)$ can be bounded as follows:

$$\begin{aligned}
\phi(C_5) &= \lim_{k\to\infty} p(C_5;G_k) \\
&\geq \lim_{k\to\infty} \frac{\left(\left(\frac{n}{5}\right)^5 + \alpha\right) k^5}{\binom{nk}{5}} \\
&= \lim_{k\to\infty} \frac{120\left(\left(\frac{nk}{5}\right)^5 + \alpha k^5\right)}{(nk)^5 + O(nk)^4} \\
&= \frac{24}{625} + \frac{120\alpha}{n^5} \\
&> \frac{24}{625},
\end{aligned}$$

contradicting Theorem 45. $\square$

Chapter 4

Further Work

4.1 A Survey of Results on the Pentagon Problem

While Grzesik's 2012 paper [11] resolved Erdős's original problem by showing that every triangle-free graph on n vertices contains at most $\left(\frac{n}{5}\right)^5$ pentagons, it is worth noting that several others have obtained further results on the pentagon problem. In this section, we will briefly detail these results.

In [15], Hatami et al. also used flag algebras to give a different, independent proof of Theorem 46. However, they made two further contributions to our existing knowledge of the problem: Firstly, we proved in Theorem 45 that $\phi(C_5) \leq \frac{24}{625}$ for all $\phi \in \mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$ when $\mathcal{H} = \{\triangle\}$, but one might wonder precisely which positive homomorphisms achieve this upper bound, and how the corresponding convergent sequences of graphs might be related (if at all) to blow-ups of the pentagon. Indeed, we noted in our proof of Theorem 46 that for a fixed graph G, the sequence of blow-up graphs $(G_k)_{k \geq 1}$ given by $G_m = G^{(m)}$ for all $m \in \mathbb{N}$ is convergent. Theorem 34 then tells us that there is a corresponding limit functional to this convergent sequence of graphs, which we will denote ϕ_G. Hatami et al. then proved the following ([15], Theorem 3.2):

Theorem 47. Let $\mathcal{H} = \{\triangle\}$. The homomorphism $\phi = \phi_{C_5}$ is the unique element in $\mathrm{Hom}^+(\mathcal{A}^\varnothing, \mathbb{R})$ satisfying $\phi(C_5) = \frac{24}{625}$.

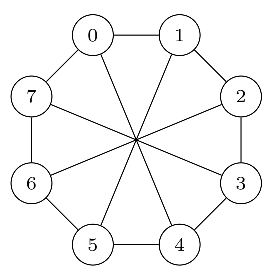
Returning to the original graph-theoretic formulation of the problem, one might ask which graphs G meet the upper bound $c(C_5; G) = \left(\frac{n}{5}\right)^5$. We certainly know that the balanced blow-up of the pentagon, $C_5^{(n/5)}$, meets this upper bound, but are there any other graphs that do? Hatami et al. also proved that there are no other such graphs ([15], Corollary 3.3):

Theorem 48. Let G be a triangle-free graph on n vertices. Then $c(C_5; G) = \left(\frac{n}{5}\right)^5$ only when n is divisible by 5 and $G = C_5^{(n/5)}$.

While this provides a tight upper bound to the parameter $c(C_5; G)$ for the instances where $|V(G)|$ is a multiple of 5, this does not resolve the question of what the tightest possible upper bound is when $|V(G)|$ is not a multiple of 5. Because the upper bound $c(C_5; G) = \left(\frac{n}{5}\right)^5$ is only achieved by the balanced blow-up of the pentagon, a natural construction to consider would be trying to make a blow-up of the pentagon "as balanced as possible", by having each of the five independent sets in the graph consist of either $\lfloor \frac{n}{5} \rfloor$ or $\lceil \frac{n}{5} \rceil$ vertices. In 2017, Lidický and Pfender showed that this construction resulted in a tight upper bound for every $n \geq 5$, with one sporadic exception ([21], Theorem 2):

Theorem 49. Let G be a triangle-free graph. Then $c(C_5; G) \leq \prod_{i=0}^{4} \left\lfloor \frac{n+i}{5} \right\rfloor$. Moreover, the only triangle-free graphs on $n \geq 5$ vertices for which $c(C_5; G) = \prod_{i=0}^{4} \left\lfloor \frac{n+i}{5} \right\rfloor$ are the blow-ups of a 5-cycle $C_5^{(\lfloor \frac{n}{5} \rfloor, \lfloor \frac{n+1}{5} \rfloor, \lfloor \frac{n+2}{5} \rfloor, \lfloor \frac{n+3}{5} \rfloor, \lfloor \frac{n+4}{5} \rfloor)}$, and the Möbius ladder ML_8 for $n = 8$.

The graph ML_8 is as follows:



We can easily confirm that we do have $c(C_5; ML_8) = 8$: Indeed, there are eight pentagons in this graph, the vertex sets of which can be found by taking any five consecutively-numbered vertices in the cycle (considering the numbers mod 8).

4.2 A Modification of the Pentagon Problem

In this section, we will discuss an elementary proof of a modification of the pentagon problem, recently found by Grzesik and Kielak [13].

While the pentagon problem asks about the maximum number of pentagons in triangle-free graphs, one might wonder about analogous questions for longer cycles. For example, what's the maximum number of C_7's in graphs without C_3's or C_5's, the maximum number of C_9's in graphs without C_3's, C_5's, or C_7's, and more generally, the maximum number of odd cycles C_k in graphs without any smaller odd cycles? Based on Erdős's original statement of the pentagon problem, one might in particular wonder whether this number can be maximised by balanced blow-ups of the odd cycle C_k. In their paper, Grzesik and Kielak proved that this is indeed the case, by showing the following theorem:

Theorem 50. For each odd integer $k \geq 7$, any graph on n vertices without odd cycles of length less than k contains at most $\left(\frac{n}{k}\right)^k$ cycles of length k.

It is worth mentioning that the method Grzesik and Kielak used to prove Theorem 50 was based on that used in a paper by Kráľ, Norin, and Volec [19], which showed that *every* n-vertex graph has at most $\frac{2n^k}{k^k}$ induced cycles of length k. That result was the farthest progress towards resolving a 1975 conjecture of Pippenger and Golumbic [25], which states that for every $k \geq 5$, an n-vertex graph has at most $\frac{n^k}{k^k - k}$ induced cycles of length k.

We will now follow the proof of Theorem 50 as given in [13], and our notation choices here will align with that of their paper.

Proof. Let $k \geq 7$ be a fixed odd integer, and let G be a graph on n vertices without odd cycles of length less than k. Observe that every C_k is an induced cycle - if there is a chord, there will be a smaller odd cycle as k is odd. Recall that $d(v, w)$ denotes the minimal distance between $v, w \in V(G)$.

Let $v_0 v_1 \cdots v_{k-1}$ be a fixed cyclic ordering of the vertices of a k-cycle. Now, define a *good sequence* to be a sequence of vertices $D = (z_i)_{i=0}^{k-1}$, where $z_i = v_i$ for $i = 0, 1$, and $i \geq 4$, and $z_2 = v_3, z_3 = v_2$. Note that there are $2k$ different cyclic orderings of a given k-cycle and one good sequence for every fixed ordering, so there are $2k$ different good sequences corresponding to each k-cycle in G.

Furthermore, define a *partial good sequence* to be any sequence of $l \leq k$ vertices which could extend to a good sequence: If $l = k - 1$, a partial good sequence must be a good sequence. If $l = 0, 1$ or $3 \leq l \leq k - 1$, a partial good sequence $D = (z_i)_{i=0}^{l-1}$ must satisfy that $z_0z_1z_3z_2...z_{l-1}$ is an induced path in G. Finally, if $l = 2$, a partial good sequence $D = (z_i)_{i=0}^{2}$ must satisfy that $z_0z_1 \in E(G)$, $d(z_1, z_2) = 2$, and $z_2 \notin N(z_0)$.

Now, for any good or partial good sequence of vertices $D = (z_i)_{i=0}^{l-1}$, where $l \leq k$, let the special sets $A_0, \cdots, A_{l-1}$ be defined on D as follows:

$$
\begin{aligned}
A_0(D) &= V(G),\\
A_1(D) &= N(z_0),\\
A_2(D) &= \{w \notin N(z_0) \mid w \in V(G), d(z_1, w) = 2\},\\
A_3(D) &= N(z_1) \cap N(z_2),\\
A_4(D) &= \{w \mid w \in V(G), z_0z_1z_3z_2w \text{ is an induced path}\},\\
A_i(D) &= \{w \mid w \in V(G), z_0z_1z_3z_2z_4 \cdots z_{i-1}w \text{ is an induced path}\} \text{ for } 5 \leq i \leq l-2,\\
A_{l-1}(D) &= \{w \mid w \in V(G), z_0z_1z_3z_2z_4 \cdots z_{l-2}w \text{ is an induced path}\} \text{ when } l < k,\\
A_{l-1}(D) &= \{w \mid w \in V(G), z_0z_1z_3z_2z_4 \cdots z_{l-2}w \text{ is an induced cycle}\} \text{ when } l = k.
\end{aligned}
$$

We make two observations: Firstly, for all $l \leq k$ and $1 \leq i \leq l-1$, $A_i(D)$ depends only on the vertices $z_0, \cdots, z_{i-1}$. Secondly, the set of vertices $A_i(D)$ is the maximal set of vertices which could extend a partial good sequence $z_0, \cdots, z_{i-1}$ with i vertices to be a (partial) good sequence with $i+1$ vertices. In terms of G, most of these sets of vertices are those which would extend the corresponding previous vertices in the good sequence to be a (induced) path, save for $A_{k-1}(D)$ when $l = k$, which instead is the set of vertices which extend the previous vertices in the good sequence to be an induced cycle, and $A_2(D)$, which is the set of second neighbours of z_1 which are not neighbours of z_0.

Now, for a good sequence $D = (z_i)_{i=0}^{k-1}$, define the *weight* $w(D)$ of D as

$$w(D) := \prod_{i=0}^{k-1} \frac{1}{|A_i(D)|}.$$

It is worth noting that this is well-defined for every good sequence, since $z_i \in A_i(D)$ for all $0 \leq i \leq k-1$ and hence $|A_i(D)|$ is positive for every i. In fact, for the same reason, the product $\prod_{i=0}^{l-1} \frac{1}{|A_i(D)|}$ is well-defined even if D is a partial good sequence with $l < k$ vertices. We will now prove a lemma concerning the sum of weights over all good sequences in G:

Lemma 51. The sum of weights of all good sequences in G is at most 1.

Proof. We will show that for all $0 \leq l \leq k-1$ and for every partial good sequence $D = (z_0, \cdots, z_l)$, the sum of the weights of all good sequences starting with D is at most $\displaystyle\prod_{i=0}^{l} \frac{1}{|A_i(D)|}$. In particular, letting $l = 0$, the sum of weights of all good sequences starting with a given vertex is at most $\frac{1}{n}$ and thus the sum of weights of all good sequences is 1.

We will demonstrate this by backward induction on l, using the base case $l = k-1$ and showing that the result for a given $l \leq k-1$ implies the result for $l-1$. The base case $l = k-1$ follows from the definition of the weight, since $z_0, \cdots, z_{k-1}$ only corresponds to one good sequence. Now suppose that the sum of weights of all good sequences starting with any partial good sequence D with $l+1$ vertices is at most $\displaystyle\prod_{i=0}^{l} \frac{1}{|A_i(D)|}$. For any partial good sequence D' with l vertices, we want to bound the sum of weights of all good sequences starting with D'.

Let $D' = (z_0, \cdots, z_{l-1})$ be fixed. By the induction hypothesis, for any $w \in V(G)$, if $D_w = (z_0, \cdots, z_{l-1}, w)$ is a partial good sequence, then the sum of weights of all good sequences starting with D_w is at most $\displaystyle\prod_{i=0}^{l} \frac{1}{|A_i(D_w)|} = \prod_{i=0}^{l} \frac{1}{|A_i(D)|}$, since $|A_i(D_w)| = |A_i(D)|$ for $0 \leq i \leq l$ and the number $|A_l(D_w)|$ only depends on $z_0, \cdots, z_{l-1}$, which are fixed, so that $|A_l(D_w)| = |A_l(D)|$ for every w. However, there are only $|A_l(D)|$ choices of w that could feasibly extend $z_0, \cdots, z_{l-1}$ to be a partial good sequence, so it follows that the sum of weights of all good sequences starting with $z_0, \cdots, z_{l-1}$ is at most $|A_l(D)| \cdot \displaystyle\prod_{i=0}^{l} \frac{1}{|A_i(D)|} = \prod_{i=0}^{l-1} \frac{1}{|A_i(D)|}$, as desired. □

Let $v_0 v_1 \cdots v_{k-1}$ be a fixed cyclic ordering of a k-cycle in G, and let $C = \{v_0, v_1, \cdots, v_{k-1}\}$ be the set of its vertices. Half the good sequences corresponding to C can be generated by shifting the indices of our fixed ordering and then swapping the third and fourth vertices in the resulting vertex sequence: Namely, letting $D_j = (v_j, v_{j+1}, v_{j+3}, v_{j+2}, v_{j+4}, \cdots, v_{j+k-1})$ where the indices are considered mod k, D_j will be a good sequence for all $0 \leq j \leq k-1$. In general, when considering indexed vertices in C in the remainder of this proof, we will consider the indices mod k. We will now prove several lemmas:

Lemma 52. Any vertex $w \in V(G)$ has at most two neighbours in C.

Proof. Suppose there exists a $w \in V(G)$ with three neighbours in C. If $w = v_m \in C$, then w will be adjacent to three vertices v_{m-1}, v_{m+1}, v_i for some i and $v_m v_i$ will be a chord, which must create a smaller odd cycle (as C itself is an odd cycle and hence one of the two paths from v_m to v_i will have even length), a contradiction.

If $w \notin C$, then w will be adjacent to some three vertices $v_i, v_j, v_l \in C$, where $i < j < l$. If two of these three vertices are adjacent in C, we will have a triangle, a contradiction. Considering the three paths $v_i v_{i+1} \cdots v_j$, $v_j v_{j+1} \cdots v_l$, $v_l v_{l+1} \cdots v_i$, one of these paths must have odd length as C is an odd cycle. Since no two of these three vertices were adjacent in C, by adding the two edges from the endpoint of this path to w, we will get a smaller odd cycle, a contradiction. □

Lemma 53. Let $w \in V(G)$. There are at most three vertices in C at distance exactly 2 from w, and any two such vertices are not adjacent.

Proof. If there are two vertices $v_i, v_{i+1} \in C$ at distance exactly 2 from w, then v_i and v_{i+1} either share a common neighbour outside of C - a contradiction, since this creates a triangle - or have two disjoint paths to w, in which case this creates a pentagon, a contradiction as $k \geq 7$. In the event that $k = 7$, this condition alone means that there are at most three vertices in C at distance exactly two from w.

Now let $k \geq 9$, and suppose there are four vertices in C, v_i, v_j, v_l, v_m at distance exactly 2 from w, no two of which are adjacent, and where $i < j < l < m$. Considering the four paths $v_i v_{i+1} \cdots v_j$, $v_j v_{j+1} \cdots v_l$, $v_l v_{l+1} \cdots v_m$, $v_m v_{m+1} \cdots v_i$, one of these paths must have odd length as C is an odd cycle. Consider the paths from the two endpoints of this odd-length path to w. These paths either meet at a common vertex that is not w or meet at w; in both cases, since none of v_i, v_j, v_l, v_m are adjacent, it follows that by adjoining the odd-length path (which has length at most $k - 6$) to the symmetric difference of the two paths from the endpoints to w (which has length at most 4), we will get an odd cycle in G of length at most $k - 2$, a contradiction. □

Now, let $n_{i,j} = |A_i(D_j)|$. We will use the previous two lemmas to prove a lemma which bounds a certain sum of $n_{i,j}$'s over most i and j, which will leave the final result to be obtainable from the AM-GM Inequality:

Lemma 54. $\displaystyle\sum_{j=0}^{k-1} \left(\frac{n_{1,j}}{2} + \sum_{i=2}^{k-1} n_{i,j} \right) \leq n(k-1).$

Proof. If $w \in V(G)$ is in a given $A_i(D_j)$, it will contribute 1 to $n_{i,j}$. Thus, recalling that $|V(G)| = n$, to show that this bound holds, it suffices to show that every $w \in V(G)$

contributes at most $k-1$ to the sum on the left-hand side of the above inequality. Lemma 52 allows us to break this up into three cases, where w has zero, one, and two neighbours in C respectively.

Case 1: Suppose that w has no neighbours in C. For every D_j, it follows that $w \notin A_1(D_j), A_3(D_j), A_4(D_j), \cdots, A_{k-1}(D_j)$, since all of these sets only contain vertices that are adjacent to some vertex in C, so w can only contribute to $n_{2,j}$. If we happen to have $d(w, v_j) = 2$ for some j, then by Lemma 53, it follows that $d(w, v_{j-1}) > 2$ and $d(w, v_{j+1}) > 2$, so w would not contribute to $n_{2,j-2}$ or $n_{2,j}$. Thus w contributes at most $k-2$ to the left-hand side of the inequality.

Case 2: Suppose that w has one neighbour in C; without loss of generality, let it be v_0. For every D_j, it follows that $w \notin A_3(D_j), A_{k-1}(D_j)$, since these sets only contain vertices that are adjacent to two vertices in C. First, let us consider the case of w contributing to $n_{i,j}$ when $i \neq 2$: If w contributes to $n_{1,j}$, it needs to be adjacent to v_j. If w contributes to $n_{i,j}$ for $4 \leq i \leq k-2$, it needs to be adjacent to v_{i+j-1}. Thus, in this case, w can only contribute to $n_{1,0}, n_{4,k-3}, n_{5,k-4}, \cdots, n_{i,k-i+1}, \cdots, n_{k-2,3}$.

If w contributes to $n_{2,j}$, it must satisfy $w \notin N(v_j)$ and $d(v_{j+1}, w) = 2$. By Lemma 53, there are at most three vertices in C at distance exactly 2 from w, but one of them is v_1 and $v_1 \in N(v_0)$, so w contributes to at most two different $n_{2,j}$'s. Thus, counting both cases, w contributes at most $k - 3 + \frac{1}{2}$ to the left-hand side of the inequality.

Case 3: Suppose that w has two neighbours in C. If these neighbours are not at distance 2 in C, then by considering the odd-length path in C between these neighbours and appending the two edges to w, we will get an odd cycle of length shorter than k in G, a contradiction. Without loss of generality, let the neighbours of w in C be v_{k-1} and v_1, so that w contributes to $n_{1,k-1}, n_{1,1}$, and no other $n_{1,j}$. Since w has exactly two neighbours in C, this also means that w contributes to $n_{3,k-2}$, and no other $n_{3,j}$.

By Lemma 53, it follows that the only vertices in C which can be at distance exactly 2 away from w are v_{k-2}, v_0, and v_2. Thus, it follows that w contributes only to $n_{2,k-3}$ and no other $n_{2,j}$'s. Finally, in order for w to contribute to $n_{i,j}$ for $4 \leq i \leq k-1$, the vertex immediately before w in the induced path or cycle must be v_{k-1}, as if it is v_1, the path or cycle will also contain v_{k-1} and hence it would not be induced. Thus, for $4 \leq i \leq k-1$, w contributes to $n_{4,k-4}, n_{5,k-5}, \cdots, n_{i,k-i}, \cdots, n_{k-1,1}$ and no other $n_{i,j}$. Counting all cases, w contributes exactly $k-1$ to the left-hand side of the inequality, as desired. □

By the definition of the weight of a good sequence and since $n_{0,j} = n$ for every $0 \leq j \leq k-1$, it follows that

$$\begin{aligned}\left(\prod_{j=0}^{k-1} w(D_j)\right)^{-\frac{1}{k(k-1)}} &= \left(\prod_{i=0}^{k-1}\prod_{j=0}^{k-1} n_{i,j}\right)^{\frac{1}{k(k-1)}} \\ &= \left(n^k 2^k \prod_{j=0}^{k-1} \frac{n_{1,j}}{2} \prod_{i=2}^{k-1} n_{i,j}\right)^{\frac{1}{k(k-1)}} \\ &= (2n)^{\frac{1}{k-1}} \left(\prod_{j=0}^{k-1} \frac{n_{1,j}}{2} \prod_{i=2}^{k-1} n_{i,j}\right)^{\frac{1}{k(k-1)}},\end{aligned}$$

and applying the AM-GM inequality (Theorem 2) to the given product of $n_{i,j}$'s, we get that

$$\left(\prod_{j=0}^{k-1} w(D_j)\right)^{-\frac{1}{k(k-1)}} \leq \frac{(2n)^{\frac{1}{k-1}}}{k(k-1)} \sum_{j=0}^{k-1} \left(\frac{n_{1,j}}{2} + \sum_{i=2}^{k-1} n_{i,j}\right).$$

Now, by Lemma 54, it follows that

$$\begin{aligned}\left(\prod_{j=0}^{k-1} w(D_j)\right)^{-\frac{1}{k(k-1)}} &\leq \frac{(2n)^{\frac{1}{k-1}} n(k-1)}{k(k-1)} \\ &= \frac{(2n^k)^{\frac{1}{k-1}}}{k},\end{aligned}$$

and simplifying, we get that equivalently,

$$\left(\prod_{j=0}^{k-1} w(D_j)\right)^{\frac{1}{k}} \geq \frac{k^{k-1}}{2n^k}.$$

Finally, by applying the AM-GM inequality to the product on the left-hand side, we get that

$$\frac{1}{k} \sum_{j=0}^{k-1} w(D_j) \geq \frac{k^{k-1}}{2n^k},$$

so that

$$\sum_{j=0}^{k-1} w(D_j) \geq \frac{k^k}{2n^k}.$$

Since the D_j account for only half the good sequences corresponding to C, it follows that the sum of weights of all good sequences corresponding to a fixed copy of C_k is at least $\left(\frac{k}{n}\right)^k$. By Lemma 51, the sum of weights of all good sequences in G is at most 1, and therefore the number of copies of C_k in G is at most $\left(\frac{n}{k}\right)^k$, as desired. □

It should be noted that Grzesik and Kielak also showed that the bound in Theorem 50 is only tight for the balanced blow-ups of a k-cycle, thus extending Hatami et al.'s result [15] on the equality case for the pentagon problem:

Theorem 55. The balanced blow-ups of a k-cycle $C_k^{(\frac{n}{k})}$ are the only graphs for which $c(C_k; G) = \left(\frac{n}{k}\right)^k$.

Proof sketch. If a graph G contains the maximum number of copies of C_k, every inequality we considered must be an equality. In particular, the bound we achieved in Lemma 54 is only tight if, for every k-cycle C, all other vertices of G are connected with two vertices of C, which are at distance 2 apart from each other. This implies that G is the blow-up of a k-cycle, and the tightness of the AM-GM inequality shows that this blow-up must be balanced. □

One might wonder why a good sequence is defined to be the cyclic ordering of the vertices in a cycle with two vertices swapped. Observe that for Lemma 51 to work, our sets $A_i(D)$ need to be the maximal possible sets of vertices which could possibly extend a fragment of a good sequence to its next vertex. Thus, if we simply defined a good sequence as the cyclic ordering of the vertices in a cycle, in Case 3 of Lemma 54, every vertex w which is adjacent to v_1, v_{k-1} would get counted twice among the $n_{1,j}$'s (as it's adjacent to both v_1 and v_{k-1}), twice among the $n_{2,j}$'s (as $v_{k-2}v_{k-1}w$ and v_0v_1w are both induced paths), and once among each subsequent set of $n_{i,j}$'s, which would not enable us to reach the bound of $k-1$ necessary for the ensuing computations to work. We can avoid this problem by defining $A_2(D)$ differently (and thus necessarily defining a good sequence differently) to allow this counting to work.

Observe that the theorem applies only to odd integers $k \geq 7$, so this unfortunately does not amount to an elementary proof of the pentagon problem. That being said, one might wonder where exactly this proof breaks down for the pentagon case. When applied to the pentagon, there is no analogue to Lemma 53, as two adjacent vertices in a cycle C being at distance exactly two from another vertex $w \notin C$ would simply mean that the paths from w to C would need to be disjoint, and adding in the edge in C, would induce another pentagon. This would mean that the best possible version of Lemma 53 we could obtain would be that the number of vertices in C at distance exactly 2 from w is at most the number of vertices in C that w is not adjacent to, and cases 1 and 2 of Lemma 54 would no longer work accordingly. Changing the weighting of the terms in the left-hand side of 54 (through factoring out different numbers from $n_{i,j}$ instead of a 2 from $n_{1,j}$) would not work, and similar situations occur, albeit perhaps with case 3 of Lemma 54 breaking instead, even if one modifies the definition of a good sequence to be a different permutation of $(v_0, v_1, v_2, v_3, v_4)$.

4.3 Conclusion

The Erdős pentagon problem asks about the maximum number of copies of C_5 a triangle-free graph on n vertices can have. In this thesis, we have seen how this simply-stated problem is much more difficult to resolve: While we examined this problem from multiple perspectives, including a simple approach by Győri (section 2.1), an algebraic approach (section 2.2), and an approach investigating structure in extremal graphs (section 2.3), the only approach which managed to yield a solution was through the semidefinite method for flag algebras (section 3.9).

While we could certainly conjecture and prove modifications of the pentagon problem (as we did in section 4.2), the most tantalising open question is whether there exists an elementary solution to this problem. The proof by flag algebras resolves Erdős's pentagon problem asymptotically and has been accepted by the mathematical community, but it was aided by a computer (in solving the semidefinite program); in fact, all known proofs to date resolving the pentagon problem have been computer-assisted.

Some may see the resolution to this problem as unsatisfying: While many proofs of theorems in extremal graph theory such as our first proof of Mantel's theorem (Theorem 17) are able to tell us something about the elementary structure of a graph (often, an extremal graph) that causes the theorem to be true, this has continued to elude us here. If anything, the proof we presented signals to us that this task may be daunting, since we were only able to investigate the pentagon density in a triangle-free graph through looking at the densities of every other triangle-free graph on five vertices, which in turn required us to "factor through" the densities of all four-vertex flags corresponding to all possible three-vertex types.

The late Erdős often talked about proofs that he thought to be from “The Book”, where “the perfect proofs for mathematical theorems” are maintained [1]; these were often succinct and elegant solutions for simply asked questions. Perhaps the pentagon problem has a solution which Erdős might exclaim to be in The Book, but if it does, it still remains to be seen what exactly this might look like.

References

[1] M. Aigner and G.M. Ziegler. *Proofs from the Book*. Springer-Verlag, 6th edition, 2018.

[2] N. Biggs, K. Lloyd, and R. J. Wilson. *Graph Theory, 1736-1936*. Clarendon Press, 1976.

[3] G. Blekherman, P. A. Parrilo, and R. R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, 2013.

[4] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi. Graph limits and parameter testing. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing - STOC '06*, 2006.

[5] M. K. De Carli Silva, F. M. De Oliveira Filho, and C. M. Sato. Flag algebras: A first glance. *Nieuw Archief voor Wiskunde*, 17(3):193–199, 2016.

[6] R. Diestel. *Graph theory*. Springer, 5th edition, 2017.

[7] P. Erdős. On some problems in graph theory, combinatorial analysis and combinatorial number theory. In *Graph Theory and Combinatorics (Cambridge, 1983)*, pages 1–17. Academic Press, 1984.

[8] P. Erdős, L. Lovász, and J. Spencer. Strong independence of graphcopy functions. In *Graph Theory and Related Topics*, pages 165–172. Academic Press, 1978.

[9] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bulletin of the American Mathematical Society*, 52(12):1087–1092, 1946.

[10] P. Frankl and V. Rödl. Near perfect coverings in graphs and hypergraphs. *European Journal of Combinatorics*, 6(4):317–326, 1985.

[11] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *Journal of Combinatorial Theory, Series B*, 102(5):1061–1066, 2012.

[12] A. Grzesik. *Flag Algebras in Extremal Graph Theory*. PhD thesis, Jagiellonian University, 2014.

[13] A. Grzesik and B. Kielak. On the maximum number of odd cycles in graphs without smaller odd cycles, 2018. arXiv:1806.09953.

[14] E. Győri. On the number of C_5's in a triangle-free graph. *Combinatorica*, 9(1):101–102, 1989.

[15] H. Hatami, J. Hladký, D. Král', S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. *Journal of Combinatorial Theory, Series A*, 120(3):722–732, 2013.

[16] P. Haxell, E. Long, J. Skokan, and M. Stein. Personal communication.

[17] P. Haxell and J. Skokan. Personal communication.

[18] P. Hell and J. Nešetřil. *Graphs and Homomorphisms*. Oxford University Press, 2008.

[19] D. Král', S. Norin, and J. Volec. A bound on the inducibility of cycles, 2018. arXiv:1801.01556.

[20] B. Lidický. Flag algebras and some applications. 50th Czech-Slovak Conference *Graphs 2015*, 2015.

[21] B. Lidický and F. Pfender. Pentagons in triangle-free graphs, 2017. arXiv:1712.08869.

[22] L Lóvasz. *Large Networks and Graph Limits*. American Mathematical Society, 2013.

[23] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Technical Report TR-2004-79*, 2004.

[24] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1907.

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

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

[27] A. Razborov. Flag algebras: An interim report. *The Mathematics of Paul Erdős II*, pages 207–232, 2013.

[28] A. Razborov. What is... a flag algebra? *Notices of the American Mathematical Society*, 60(10):1324–1327, 2013.

[29] G. Sabidussi. Graph derivatives. *Mathematische Zeitschrift*, 76(1):385–401, 1961.