Substructure Densities in Extremal Combinatorics

by

Timothy F. N. Chan

A thesis submitted to the

School of Mathematics
Monash University
Australia

and

Mathematics Institute
University of Warwick
United Kingdom

for the joint degree of

Doctor of Philosophy

February 2021

Contents

Copyright notice

Declaration

This thesis is an original work of my research. It is submitted to both Monash University and the University of Warwick in partial fulfilment of the requirements of the Monash–Warwick Joint PhD, and contains no material which has been accepted for the award of any other degree. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis. In particular, it presents results from the following publications:

- T. F. N. Chan, A. Grzesik, D. Král' and J. A. Noel: *Cycles of length three and four in tournaments*, J. Combin. Theory Ser. A **175** (2020)
- T. F. N. Chan, D. Král', J. A. Noel, Y. Pehova, M. Sharifzadeh and J. Volec: *Characterization of quasirandom permutations by a pattern sum*, Random Structures Algorithms **57** (2020), 920–939
- T. F. N. Chan, D. Král', B. Mohar and D. R. Wood: *Inducibility and universality for trees*, preprint arXiv:2102.02010 (2020)

Timothy F. N. Chan
February 23, 2021

Acknowledgements

My deepest gratitude goes towards my supervisors David R. Wood and Daniel Král'. Their guidance has been invaluable, and I cannot emphasise enough how much they have done for me. To my collaborators, all of whom have been a joy to work with, *thank you*. To my parents, who have always supported me, and to my friends, who have made this whole journey a fun ride, *thank you*!

This research was supported by an Australian Government Research Training Program (RTP) Scholarship. Additionally, I would like to acknowledge Hays Whitlatch for bringing the results of Eric Zhang [102] to my attention, Yanitsa Pehova and Andrzej Grzesik for technical advice on producing this thesis, and Robert Hickingbotham for his helpful input.

Abstract

One of the primary goals of *extremal combinatorics* is to understand how an object's properties are influenced by the presence or multiplicity of a given substructure. Most classical theorems in the area, such as Mantel's Theorem, are phrased in terms of substructure *counts* such as the number of edges or the number of triangles in a graph. Gradually, however, it has become more popular to express results in terms of the *density* of substructures, where the substructure counts are normalised by some natural quantity. This approach has several benefits; results are often more succinctly stated using densities, and it becomes easier to focus on the asymptotic behaviour of objects.

In this thesis, we study three topics concerning density. We begin Chapter 1 by contextualising the study of combinatorial density and justifying its importance within extremal combinatorics. We also introduce the relevant combinatorial objects, results, and questions that are central to the later chapters. Particular attention is paid to developing the theory of graph limits and flag algebras, two modern fields that rely heavily on the notion of density.

In Chapter 2, we investigate the interplay between the densities of cycles of length 3 and 4 in large tournaments. In particular, we prove two cases of a conjecture of Linial and Morgenstern (2016) that the minimum density of 4-cycles in a graph with a fixed density of 3-cycles is attained by a particular random construction.

In Chapter 3, we explore quasirandom permutations. A permutation is said to be quasirandom if the density of every subpermutation matches the expected density in a random permutation. Our main result is that quasirandomness can be characterised by a property which, on the surface, appears significantly weaker.

Lastly, in Chapter 4, we resolve a problem posed by Bubeck and Linial (2016) on the inducibility of trees. The inducibility of a tree X is defined as the maximum possible density of X in a large tree. We show that there exist non-path, non-star trees with positive inducibility, but that all such trees have inducibility bounded away from 1. We also show that there exists a sequence of trees in which every possible subtree appears asymptotically with positive density.

Chapter 1

Introduction

Extremal combinatorics is an area of discrete mathematics that investigates how the global parameters of a combinatorial object influence its properties [7, 8, 39]. The study of extremal combinatorics is commonly understood to have originated with the work of Mantel, who proved in 1907 that every n-vertex graph with more than $n^2/4$ edges has the property that it contains a triangle as a subgraph [79]. Now named Mantel's Theorem, this result and its generalisations form the foundation of *extremal graph theory*, the most classical subfield of extremal combinatorics.

More specifically, Mantel's Theorem can be seen as the case $r = 3$ of the equally well-known Turan's Theorem [96], which determines the minimum number of edges an n-vertex graph must have to force the existence of an r-vertex clique K_r as a subgraph. Both of these theorems address the *forbidden subgraph problem*:

Given a graph H, what is the maximum number of edges in an n-vertex, H-free graph?

When H is not bipartite, the problem is essentially solved by the celebrated Erdős–Stone Theorem [43], which states that the threshold number of edges is

$$\left[\frac{\chi(H)-2}{\chi(H)-1} + o(1)\right]\binom{n}{2},$$

where $\chi(H)$ is the chromatic number of H and the little-o is interpreted as $n \to \infty$. Proved in 1946 and more recently described by Bollobás as "the fundamental theorem of extremal graph theory" [7], this landmark result illustrates some of the themes that have driven the field in subsequent decades. For example, the perhaps surprising appearance of the chromatic number in the above expression inspired the study of graph parameters beyond simply the number of vertices and edges. Graph invariants such as chromatic number, connectivity, minimum degree, and average degree have since become standard parameters that are investigated in extremal graph theory [8].

Secondly, the concession of a $o(1)$ term reflects a focus on asymptotic (as opposed to exact) behaviour. This increases the scope for powerful, general theorems that may otherwise be unapproachable or burdened by notation that detracts from the key behaviour. It is worth highlighting that the $o(1)$ term appears inside the factor of $\binom{n}{2}$, a very natural expression that corresponds to the maximum possible number of edges in an n-vertex graph. This suggests that, asymptotically at least, the most succinct way to describe the threshold is not as a relationship between the number of edges and vertices, but rather as a proportion of possible edges that are realised in the graph. In other words, the dependence of the threshold on n can be asymptotically removed if we normalise by $\binom{n}{2}$ to obtain an edge *density* between 0 and 1. More broadly, density can be viewed as a meaningful normalisation of substructure counts that draws attention to asymptotic behaviour. As we shall see, questions of an extremal nature can be asked for a wide range of combinatorial objects, and with an appropriate choice of normalisation factor for each object we can more easily discuss the commonalities in their behaviour.

Intriguingly, there is another generalisation of Mantel's Theorem that traces its roots back to the 1940s. Although complete balanced bipartite graphs demonstrate the existence of triangle-free graphs with $\lfloor n^2/4 \rfloor$ edges, Rademacher and later Erdős [44] surprisingly proved that exceeding this bound immediately forces the existence of not just one, but $\lfloor n/2 \rfloor$ triangles. Building on this interesting phenomenon, the Erdős–Rademacher Problem asks *how many* triangles are forced in an n-vertex graph with m edges, where $m > \lfloor n^2/4 \rfloor$. Note that whereas the property of interest in the forbidden subgraph problem is the *existence* of a given substructure, the Erdős–Rademacher Problem is interested in the property of having *multiple* copies of an object. It was conjectured that the graph forcing the fewest triangles was a complete multipartite graph with all parts having equal size except for one smaller part. This conjecture received substantial attention for several decades [6, 47, 53, 75], but remained open until its eventual solution in 2008 by Razborov [86], using his newly developed *flag algebra* method.

The flag algebra method and the complementary theory of *combinatorial limits* are a modern approach to extremal combinatorics that fully embraces the concept of density. The language of these fields suppresses $o(1)$ terms entirely, placing the focus on asymptotic results while avoiding the need for cumbersome notation. Equipped with appropriate definitions of density, these modern techniques can be applied to many classes of combinatorial objects; indeed, the foundational work of Razborov on flag algebras [85] was expressed in general, model-theoretic terms in anticipation of its extensive utility.

In this thesis we will explore three extremal questions, all centered around the concept of density, and apply them to three different classes of combinatorial objects. The first is a direct analogy of the Erdős–Rademacher Problem in the setting of *tournaments*, i.e., oriented complete graphs. A conjecture of Linial and Morgenstern [72] asserts that the minimum density of (directed) 4-cycles in a tournament forced by a given density of 3-cycles is achieved by a specific blow-up construction. If true, this would demonstrate behaviour similar to that proved by Razborov for K_2 and K_3 in graphs. Utilizing tools from spectral graph theory and optimisation, we confirm the conjecture when the density of 3-cycles exceeds $1/72$ and additionally describe the family of extremal constructions when the density exceeds $1/32$. Along the way, we show how the framework of combinatorial limits expresses these results.

The second question is about the density of subpermutations (called *patterns*) in permutations. Building on the foundational work of Chung, Graham, and Wilson [32] on *quasirandomness*, Cooper [34] defined *quasirandom permutations*, which have the same pattern densities as random permutations despite being deterministic. We formulate this concept in the convenient language of permutation limits, and apply the flag algebra method to prove that quasirandomness is equivalent to a property that appears, at first glace, to be weaker. In particular, we show that if the densities of a specific set of eight patterns match those expected in a random permutation, then the densities of all other patterns match as well. This is an analogue of a theorem about quasirandom graphs, extending a line of research on permutations initiated by Graham (see [34], page 141).

Lastly, we turn our attention to the concept of density in *sparse* graphs, which has been a topic of recent interest [9, 11, 12, 60, 67, 80]. In the sparse setting, to ask any meaningful questions about density we must normalise our substructure counts in an alternative way. In our case, we will resolve two open problems posed by Bubeck and Linial [22] about the density of subtrees in trees, and compare them to the behaviour observed in general graphs. The first of these questions concerns the *inducibility* of trees. The inducibility of a tree X is defined as the maximum possible density of X in a large tree. We show that there exist non-path, non-star trees with positive inducibility, but that all such trees have inducibility bounded away from 1. We also affirmatively answer a question about *universality* by showing that there exists a sequence of trees in which every possible subtree appears asymptotically with positive density.

The remainder of this chapter is devoted to establishing notation and defining the objects and questions that will be studied in this thesis. The tools that we use will at first be presented in the unified context of graphs before we progress to the settings of tournaments, permutations, and trees. Note that some general knowledge of measure theory will be assumed; the relevant measure-theoretic concepts can be found, for example, in [48].

1.1 Graphs

A graph is a mathematical structure that models a collection of objects and a symmetric, binary relation between them. More formally, a (simple) graph G consists of a finite set V of *vertices*, together with a set $E \subseteq \binom{V}{2}$ of *edges*. By convention we write $G = (V, E)$, and refer to the vertex and edge sets of G with $V(G)$ and $E(G)$ respectively. The number of vertices of G, denoted by $|V(G)|$ or simply $|G|$, is known as the *order* of G, and the number of edges of G, denoted by $|E(G)|$, is known as the *size* of G. The *degree* of a vertex v in G, written $\deg_G(v)$, is the number of edges of G that contain v.

Subgraphs. Given two graphs H and G, we say that H is a *subgraph* of G and write $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. For a set of vertices $V' \subseteq V(G)$, the *induced subgraph* on V', denoted by $G[V']$, has vertex set V' and edge-set $E(G) \cap \binom{V'}{2}$.

Morphisms. A *homomorphism* from a graph H to a graph G is a map $\varphi : V(H) \to V(G)$ such that $\varphi(u)\varphi(v)$ is an edge of G whenever uv is an edge of H. An *isomorphism* from H to G is a bijective homomorphism such that $\varphi(u)\varphi(v)$ is an edge of G if and only if uv is an edge of H. If there exists an isomorphism from H to G, then we say that H is *isomorphic* to G.

Random graphs. For $n \in \mathbb{N}$ and $p \in [0, 1]$, the *Erdős–Rényi random graph* $G(n, p)$ is the graph with n vertices and each edge included independently with probability p. Thus, $G(n, p)$ is expected to have $p\binom{n}{2}$ edges. Whenever we work with randomly-constructed graphs on n vertices, we say that a statement holds *with high probability* if it holds with probability converging to 1 as $n \to \infty$.

Asymptotics. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is $O(1)$ if there exists constants $c, N \in \mathbb{R}$ with $|f(x)| \leqslant c$ for all $x > N$. The function f is said to be $o(1)$ if for *every* positive $c \in \mathbb{R}$ there exists a number N_c with $|f(x)| \leqslant c$ for all $x > N_c$.

1.1.1 Graph limits

In this section, we introduce the key definitions, intuitions, and results from the theory of *graph limits*. As the theoretical underpinnings lean heavily on tools from probability theory, this subsection of the thesis is free of proofs; for a detailed treatment of the topic, see the monograph of Lovász [74].

The field of graph limits emerged in a series of papers by Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi [14,16,17,76,77] in the period 2006–2012, essentially setting out to answer the following question:

What does it mean for a sequence of graphs to converge?

For graphs with quadratically-many edges, this question has been satisfactorily resolved, and we now detail the answer. If G and H are graphs with $|H| \leqslant |G|$, the *density* of H in G, denoted $d(H,G)$, is the probability that a randomly chosen subset of $|H|$ vertices of G induces a subgraph isomorphic to H. In other words,

$$d(H,G) = \frac{\#(H,G)}{\binom{|G|}{|H|}},$$

where $\#(H,G)$ is the number of subgraphs of G isomorphic to H. (If $|G| > |H|$, we set $d(H,G)$ to be 0.) We note here that an equivalent theory of graph limits can be built on an alternative definition of density: that of *homomorphism density*, the probability that a random map $\varphi : V(H) \to V(G)$ defines a graph homomorphism. While arguably less intuitive, the homomorphism approach can sometimes lead to simpler analytic expressions, as well shall see in Chapter 2.

A sequence $(G_n)_{n\in\mathbb{N}}$ of graphs is *convergent* (or *locally convergent*) if the sequence of densities $(d(H,G_n))_{n\in\mathbb{N}}$ converges for every graph H. In general, we will only be interested in graph sequences $(G_n)_{n\in\mathbb{N}}$ where $|G_n| \to \infty$. Examples of convergent graph sequences include the sequence of complete graphs $(K_n)_{n\in\mathbb{N}}$, the sequence of balanced complete bipartite graphs $(K_{n,n})_{n\in\mathbb{N}}$, and the sequence of complete bipartite graphs $(K_{n,\lfloor \alpha n\rfloor})_{n\in\mathbb{N}}$ with part sizes converging to some ratio $\alpha \in (0,1)$. Another key example is the sequence of Erdős–Rényi random graphs $(G(n,p))_{n\in\mathbb{N}}$, which can be shown to converge with probability 1 using the Borel-Cantelli Lemma and Azuma-Hoeffding Inequality, relatively simple tools from probability theory.

Now that we have established a notion of convergence and have given examples of convergent sequences, the natural question to ask is:

Do convergent sequences of graphs have a limit object?

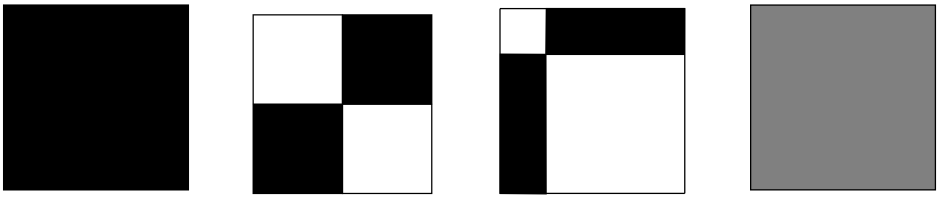


Figure 1.1: The graphons corresponding to the graph sequences $(K_n)_{n\in\mathbb{N}}$, $(K_{n,n})_{n\in\mathbb{N}}$, $(K_{n,3n})_{n\in\mathbb{N}}$, and $(G(n,1/2))_{n\in\mathbb{N}}$.

It turns out that a convergent sequence of graphs is best represented not by an infinite graph, but rather an analytic object called a *graphon*. Furthermore, in the space of graphons, many graph-theoretic problems and constructions have a simpler description (see [15] for a survey). A graphon is a measurable function $W : [0,1]^2 \to [0,1]$ that is symmetric in the sense that $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. One way to think of a graphon is as a continuous limit of the adjacency matrix of a large graph; we will see shortly that this analogy is misleading, but it nevertheless provides a good first intuition.

Typically, graphons are depicted as a shaded unit square, with the value of $W(x,y)$ represented by a shade ranging from white (denoting a value of 0) to black (denoting a value of 1). Because of the analogy to adjacency matrices, it is customary to draw the point $(0,0)$ in the top left corner. Several examples of graphons are given in Figure 1.1.

If W is a graphon, then a *W-random graph of order n* is the n-vertex graph obtained by sampling n points $x_1, \dots, x_n$ uniformly and independently in the interval $[0,1]$ and adding an edge between vertex i and vertex j with probability $W(x_i, x_j)$. As this definition is probabilistic, graphons can be seen as a type of random graph model. In particular, if W is the graphon that is uniformly equal to p for some $p \in [0,1]$, then a W-random graph of order n is exactly the Erdős–Rényi random graph $G(n,p)$.

Having established a method of sampling a graph H from a graphon W, we can define the *density* $d(H,W)$ of H in W to be the probability that a W-random graph of order $|H|$ is isomorphic to H. In particular, if $v_1, \dots, v_{|H|}$ are the vertices of H and $\mathrm{Aut}(H)$ is the automorphism group of H, then

$$d(H,W) = \frac{|H|!}{|\operatorname{Aut}(H)|} \int_{[0,1]^{|H|}} \prod_{v_i v_j \in E(H)} W(x_i, x_j) \prod_{v_i v_j \notin E(H)} (1 - W(x_i, x_j))\, dx_1 \dots x_{|H|}.$$

Essentially, the integral fixes a labelling of the vertices of H and computes the probability that it is isomorphic (as a labelled graph) to a vertex-ordered W-random graph of order $|H|$; the factor $|H|!/|\operatorname{Aut}(H)|$ de-orders the vertices. Finally, we say that a graphon W is the *limit* of a convergent sequence $(G_n)_{n\in\mathbb{N}}$ of graphs if for every graph H,

$$d(H,W) = \lim_{n\to\infty} d(H,G_n).$$

Theorem 1.1. *(Lovász and Szegedy [76]) Let W be a graphon, and for each $n \in \mathbb{N}$ let G_n be a W-random graph of order n. The sequence $(G_n)_{n\in\mathbb{N}}$ is convergent and W is its limit with probability* 1*.*

Based on the examples in Figure 1.1, one might start to wonder where the intuition of graphons as a continuous limit of adjacency matrices begins to fall flat. After all, the depicted graphons are more-or-less macroscopic versions of the adjacency matrices of the associated graph sequences. However, it must be remembered that the adjacency matrix of a graph depends on the ordering of its vertices. While the most common way of representing the adjacency matrix of $K_{n,n}$ is by grouping vertices from the same side of the bipartition together, it is also possible to order the vertex set by interweaving vertices from each side one-at-a-time. The resulting adjacency matrices consist of an alternating pattern of 0s and 1s. If such a sequence of matrices could be said to have a continuous limit, it certainly could not be the second graphon depicted in Figure 1.1. If anything, the limit ought to be the graphon that takes the value $1/2$ everywhere, i.e., the graphon for $G(n,p)$. However, if H is a non-bipartite graph, then $d(H,K_{n,n}) = 0$ for every n but the density of H in the random graphon is non-zero. Therefore, the random graphon cannot be a limit for the sequence $(K_{n,n})_{n\in\mathbb{N}}$.

Clearly then, it is faulty to view graphons as the limit of adjacency matrices. Repairing this intuition requires the well-known concept of graph regularity, introduced by Szemerédi in 1978 [94]. For graph limits, a weaker notion of regularity due to Frieze and Kannan [49] is used. Given a large graph G, the Regularity Lemma equipartitions the vertex set of G into many parts. It then builds an auxiliary real-valued matrix whose rows and columns are indexed by these parts, and whose (i,j)-th entry describes the number and approximate distribution of edges between the i-th and j-th parts of the equipartition. By repeatedly refining these regular partitions, it is possible to better and better approximate the structure of G, and at the same time ensure some conformity between the resulting auxiliary matrices, akin to increasing the resolution of a blurry image. While the specifics are outside the scope of this thesis, the graphon associated with a convergent sequence $(G_n)_{n\in\mathbb{N}}$ can be more accurately viewed as the continuous limit of the auxiliary matrices constructed in this way. Indeed, the Regularity Lemma is the main tool used to prove the following fundamental theorem about the existence of graph limits.

Theorem 1.2. *(Lovász and Szegedy [76]) Every convergent sequence* $(G_n)_{n\in\mathbb{N}}$ *of graphs has a limit graphon* W*.*

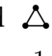
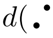
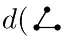
Having addressed the existence of graph limits, it remains to consider uniqueness. In other words, do there exist distinct graphons W_1 and W_2 such that $d(H, W_1) = d(H, W_2)$ for every graph H? The answer is yes, but in a highly controlled way. Call W_1 and W_2 *weakly isomorphic* if $d(H, W_1) = d(H, W_2)$ for every graph H. A general method of constructing weakly isomorphic graphons is as follows: Let $\varphi : [0, 1] \to [0, 1]$ be a *measuring-preserving map*, that is, a map for which the preimage of any measurable set $S \subseteq [0, 1]$ is measurable and has the same measure as S. Then the graphons W and W^{φ} are weakly isomorphic because φ preserves integrals. The final theorem of this subsection states that measure-preserving maps are essentially the only way to generate weakly isomorphic graphons.

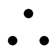
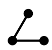
Theorem 1.3. *(Borgs, Chayes, and Lovász [13]) If* W_1 *and* W_2 *are weakly isomorphic graphons, then there exist measure-preserving maps* $\varphi_1, \varphi_2 : [0, 1] \to [0, 1]$ *such that the graphons* $W_1^{\varphi_1}$ *and* $W_2^{\varphi_2}$ *are equal almost everywhere.*

This concludes the overview of the theoretical background behind graph limits. Although graphons will not be directly featured in this thesis, the analogous notions for tournaments and permutations will be central to Chapters 2 and 3.

1.1.2 Flag algebras

The celebrated method of *flag algebras* was developed by Alexander Razborov in 2007 [85], and one of its first applications was to solve [86] the Erdős–Rademacher Problem, which at the time had been open for decades. Like graph limits, flag algebras study the density of graphs inside graph sequences whose orders tend to infinity, thereby ignoring non-asymptotic behaviour. However, whereas graph limits typically study analytic properties such as convergence, existence, and uniqueness, one way to think of flag algebras is as an attempt to elucidate the syntactic structure underlying many standard arguments in extremal combinatorics.

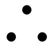
In this section, we often represent a graph by drawing it. For example, is the empty graph on two vertices and  is the complete graph on three vertices. In this notation, consider the equation $1 = d($$, G) + d($$, G) + d($$, G) + d($$, G)$, which holds for every graph G. Even more succinctly, if we allow ourselves to replace the expression $d(H, G)$ with simply H, we can write

$1 =$  $+$  $+$  $+$  .

What other general equalities hold about subgraph densities? One example is that whenever H is a graph with at most k vertices,

$$H = \sum_{F \in \mathcal{F}_k} d(H, F) \cdot F, \tag{1.1}$$

where $\mathcal{F}_k$ is the set of graphs on k vertices. Here, we implicitly assume that we have some background graph G, with H and F being written as shorthand for $d(H, G)$ and $d(F, G)$. The equation is a reflection of the fact that choosing $|H|$ vertices at random is no different from choosing $k \geqslant |H|$ vertices at random and then choosing a subset of $|H|$ of those vertices. For example, when $H = K_2$ and $k = 3$,

 $= 0 \cdot$  $+ \frac{1}{3} \cdot$  $+ \frac{2}{3} \cdot$  $+ 1 \cdot$  (1.2)

In fact, we can say something even more general. Let H_1, H_2, and F be graphs with $|H_1| = k_1$, $|H_2| = k_2$, and $|F| = k_1 + k_2$. Define $d(H_1, H_2; F)$ to be the probability that a randomly chosen set of k_1 vertices of F induces a copy of H_1 and the remaining vertices induce a copy of H_2. Then

$$d(H_1, G) \cdot d(H_2, G) = (1 + o(1)) \sum_{F \in \mathcal{F}_{k_1+k_2}} d(H_1, H_2; F) \cdot d(F, G) \tag{1.3}$$

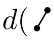
for any graph G with $|G| \to \infty$. Although we omit a formal proof, the intuition is similar to the reasoning behind (1.1); to choose k_1 and k_2 vertices from G, one may first choose a set of $k_1 + k_2$ vertices and then partition this set. The $o(1)$ error term arises because on the left side of the equation, our random choice of k_1 vertices might overlap with our random choice of k_2 vertices, while on the right side, our sets are guaranteed to be disjoint. Of course, when $|G|$ is large, it makes essentially no difference whether vertices are chosen with or without replacement.

To codify the general equalities above into an algebraic object, first let $\mathcal{F}$ be the set of graphs, and let $\mathbb{R}\mathcal{F}$ be the space of formal linear combinations of elements of $\mathcal{F}$, i.e., the set of expressions of the form $a_1 F_1 + \cdots + a_n F_n$, where $a_i \in \mathbb{R}$ and $F_i \in \mathcal{F}$. Define $\mathcal{A}$ to be the algebraic quotient of $\mathbb{R}\mathcal{F}$ by the expression $H - \sum_{F \in \mathcal{F}_k} d(H, F) \cdot F$ for each $H \in \mathcal{F}$ and each $k \geqslant |H|$. Essentially, the quotient enforces equations of the form (1.1) in $\mathcal{A}$. Next, given graphs H_1, H_2 (considered as elements of $\mathcal{A}$), we define the following multiplication operation that encodes equations of the form (1.3):

$$H_1 \times H_2 = \sum_{F \in \mathcal{F}_{k_1+k_2}} d(H_1, H_2; F) \cdot F.$$

Extending linearly, we obtain a multiplication operation on $\mathcal{A}$. Note that the $o(1)$ term that appears in (1.3) is suppressed, so expressions in $\mathcal{A}$, when interpreted as statements about subgraph densities, are accurate only when $|G| \to \infty$.

So far, what we have is an algebra $\mathcal{A}$. There is one more device that we can introduce to increase the expressiveness of this algebra. Let σ be a fixed (typically small) graph whose vertices are labelled. This labelled graph is called a *type*. Let $\mathcal{F}^\sigma$ be the set of graphs H that are equipped with a labelled copy of σ. More formally, $\mathcal{F}^\sigma$ is the set of graphs H that are equipped with an injective map θ from $V(\sigma)$ to $V(H)$, such that θ is an isomorphism (of labelled graphs) from σ to $H[\mathrm{Im}(\theta)]$. Let $\mathcal{F}^\sigma_k$ be the subset of $\mathcal{F}^\sigma$ for which $|H| = k$. The elements of $\mathcal{F}^\sigma$ are called *σ-flags*, the vertices of $\mathrm{Im}(\theta)$ are said to be *flagged*, and the subgraph of H induced by $\mathrm{Im}(\theta)$ is called the *root* of the σ-flag.

For two σ-flags H and G with $|H| \leqslant |G|$ whose maps are given by θ and θ', respectively, define the density $d(H, G)$ of H in G to be the probability that a randomly chosen subset of $|H| - |\sigma|$ vertices in $V(G) \setminus \theta'(V(\sigma))$, together with $\theta'(V(\sigma))$, induces a subgraph that is isomorphic to H through an isomorphism f satisfying $f(\theta') = \theta$. Informally speaking, we fix the root of G to be the root of H, and ask for the probability that by randomly choosing an additional $|H| - |\sigma|$ vertices of G we induce a graph isomorphic (as a partially labelled graph) to H. For example, if σ is the one-vertex flag and $\circ$ is drawn to represent flagged vertices, then $d($, $) = 0$, $d($, $) = 1/2$, and $d($, $) = 1$ (the labelling of the vertices of σ is not depicted as there is only one possible labelling of K_1). More generally, if σ is the one-vertex flag, G is a σ-flag, and v is the flagged vertex of G, then $d($$, G)$ is equal to $\deg_G(v)/(|G| - 1)$, a quantity known in classical terms as the *relative degree* of v. Thus, the point of introducing σ-flags is that it becomes possible to write expressions referencing *specific* vertices of a graph, unlike in the algebra $\mathcal{A}$, where we can only talk about properties such as the overall edge density $d($$, G)$.

We now define the product of two elements from $\mathcal{F}^\sigma$. Fix σ-flags $H, H_1, H_2 \in \mathcal{F}^\sigma$ with $|H| \leqslant k$ for the rest of this section, and let $\ell = |H_1| + |H_2| - |\sigma|$. Suppose $F \in \mathcal{F}^\sigma_\ell$ and let $\theta : V(\sigma) \to V(F)$ be the map associated with F. Define $d(H_1, H_2; F)$ to be the probability that a randomly chosen subset of $V(F) \setminus \theta(V(\sigma))$ of size $|H_1| - |\sigma|$ together with $\theta(V(\sigma))$ induces a subgraph isomorphic (as a partially labelled graph) to H_1, and that the remaining $|H_2| - |\sigma|$ vertices together with $\theta(V(\sigma))$ induce a subgraph isomorphic (as a partially labelled graph) to H_2. Then the natural analogues of (1.1) and (1.3) hold for graphs G with $|G| \to \infty$:

$$H = \sum_{F \in \mathcal{F}_k^\sigma} d(H, F) \cdot F, \text{ and} \tag{1.4}$$
$$d(H_1, G) \cdot d(H_2, G) = (1 + o(1)) \sum_{F \in \mathcal{F}_\ell^\sigma} d(H_1, H_2; F) \cdot d(F, G).$$

Thus, we can construct an algebra $\mathcal{A}^\sigma$ from the space $\mathbb{R}\mathcal{F}^\sigma$ of formal linear combinations of elements of $\mathcal{F}^\sigma$ in the same way that we constructed $\mathcal{A}$ from $\mathbb{R}\mathcal{F}$, namely by taking the quotient of $\mathbb{R}\mathcal{F}^\sigma$ by relations of the form (1.4) and defining a multiplication operation

$$H_1 \times H_2 = \sum_{F \in \mathcal{F}_\ell^\sigma} d(H_1, H_2; F) \cdot F.$$

Two examples of this multiplication are

where we have omitted terms with coefficient 0. Note that σ is the identity element of this operation as $H_1 \times \sigma = H_1$. We call $\mathcal{A}^\sigma$ the *flag algebra* of type σ.

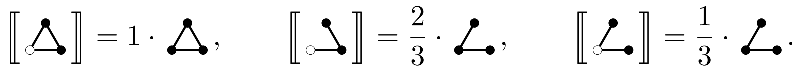
Lastly, we define a way to translate expressions in $\mathcal{A}^\sigma$ to expressions in $\mathcal{A}$. As mentioned above, σ-flags allow us to count subgraphs containing specific vertices of a large graph G. Averaging these counts over all possible choices of σ in G, we can recover standard subgraph counts. In particular, for any σ-flag H and graph G,

$$\mathbb{E}_\sigma(d(H, G^\sigma)) = p(\sigma, H^\varnothing) \cdot d(H^\varnothing, G),$$

where the expected value is taken over all σ-flags G^σ whose underlying graph is G, $H^\varnothing$ is the underlying graph corresponding to H, and $p(\sigma, H^\varnothing))$ is the probability that a random injective map from $V(\sigma)$ to $V(H^\varnothing)$ yields a σ-flag isomorphic to H. For this reason, we define the *averaging* (or *unlabelling*) operator $[\![\cdot]\!] : \mathcal{A}^\sigma \to \mathcal{A}$ as the linear operator defined on the elements of $H \in \mathcal{F}^\sigma$ by

$$[\![H]\!] = p(\sigma, H^\varnothing) \cdot H^\varnothing.$$

For example, when σ is the one-vertex flag, we have



Equipped with the advanced language of flag algebras, we can very efficiently prove an asymptotic version of Mantel's Theorem:

Theorem 1.4. *If* $K_3 = 0$*, then* $K_2 \leqslant 1/2$.

Proof. We begin by averaging the expression $(1 - 2K_2^{\sigma})^2$, which we know is non-negative because it is the square of a real number. Using $K_3 = 0$, we obtain

$$\begin{aligned}
0 &\leqslant \left[\!\left[\left(1 - 2K_2^{\sigma}\right)^2 \right]\!\right] \\
&= \left[\!\left[1 - 4K_2^{\sigma} + 4{K_2^{\sigma}}^2 \right]\!\right] \\
&= \left[\!\left[1 - 4K_2^{\sigma} + 4K_3^{\sigma} + 4P_3^{\sigma} \right]\!\right] \\
&= 1 - 4K_2 + 4K_3 + \frac{4}{3}P_3 \\
&= 1 - 4K_2 + \frac{4}{3}P_3.
\end{aligned}$$

By (1.2) with $K_3 = 0$, we have $2K_2 = \frac{2}{3}\overline{P_3} + \frac{4}{3}P_3$, so the last expression above is equal to $1 - 2K_2 - \frac{2}{3}\overline{P_3}$. Therefore, $2K_2 + \frac{2}{3}\overline{P_3} \leqslant 1$, and in particular, $K_2 \leqslant 1/2$. $\square$

In fact, the proof above even tells us something about the structure of extremal examples—since we derive the inequality $2K_2 + \frac{2}{3}\overline{P_3} \leqslant 1$, any graph G with $|G| \to \infty$ that satisfies $K_2 = 1/2$ and $K_3 = 0$ must have $\overline{P_3} = 0$. This property is exhibited by the extremal graph $K_{n,n}$, the complete balanced bipartite graph. Proving Mantel's Theorem (which concerns finite graphs) from Theorem 1.4 requires a simple *blow-up* argument that we briefly detail. Suppose that there exists a graph G with $d(K_2, G) > 1/2$ and $d(K_3, G) = 0$. Let $V(G) = \{v_1, \ldots, v_{|G|}\}$. For every integer $n \geqslant 1$, let G_n be the graph with $|G|n$ vertices organised into $|G|$ equal parts, and whenever v_iv_j is an edge of G, add all edges between vertices of the i-th and j-th parts of G_n. Then the sequence $(G_n)_{n\in\mathbb{N}}$ has $K_2 = d(K_2, G) > 1/2$ and $K_3 = d(K_3, G) = 0$, contradicting Theorem 1.4.

In principle, the above proof of Theorem 1.4 could have been expressed without the need for the flag algebra framework. Nevertheless, the formalism provides a succinct way to represent an argument that no doubt is used frequently in extremal combinatorics. Later, in Section 1.3 and Chapter 3, we will see that it also lends itself to the search for computerised proofs of combinatorial inequalities. For an overview of results that have been obtained in this way, see the survey of Razborov [87], and for a more detailed and formal introduction to flag algebras, we refer the reader to the PhD theses of Volec [98] and Grzesik [57].

Despite the many results that have been shown using flag algebras, we conclude this subsection with a brief discussion about its theoretical limitations. Observe that the proof of Theorem 1.4 began with an inequality of the form $0 \leqslant [\![A^2]\!]$ for some $A \in \mathcal{A}^\sigma$. More generally, the main source of inequalities on flag algebras is the following analogue of the Cauchy–Schwarz inequality:

Theorem 1.5 (Razborov [85])**.** *Let σ be a type. If $f, g \in \mathcal{A}^\sigma$ then*

$$[\![f^2]\!]\,[\![g^2]\!] \geqslant [\![fg]\!]^2 .$$

In particular, $[\![f^2]\!]\,[\![\sigma]\!] \geqslant [\![f]\!]^2$, which also implies

$$[\![f^2]\!] \geqslant 0.$$

This inequality forms the basis of computational methods in flag algebras. Indeed, one might hope that *every* linear inequality that holds between graph densities can be derived from Theorem 1.5. Along these lines, Lovász and Szegedy [78] proved that all such inequalities can be approximated arbitrarily well by a sum of Cauchy–Schwarz inequalities. However, an undecidability result of Hatami and Norine [61] (see also [5]) later showed that there exist linear graph density inequalities whose truth cannot be computationally verified.

1.2 Tournaments

As previously stated, graphs model symmetric, binary relations between objects. However, many real-world phenomena, such as in social choice theory, are more accurately modelled by anti-symmetric relations [18]. This motivates the study of *directed graphs*, i.e., graphs in which edges (called *arcs*) have a direction. One of the most well-known classes of directed graphs are *tournaments*, which are complete graphs where every edge is given an orientation.

Tournaments were studied as early as the 1930s by Rédei [88], and are so-called because they can be used to model a round-robin competition between players (represented by vertices), with the edge between two players directed towards the winner of their head-to-head match. An example of a tournament on n vertices $\{v_1, \ldots, v_n\}$ is the *transitive tournament*, where an ordering of the vertices is fixed and an arc is directed from v_i to v_j whenever v_i occurs before v_j in the ordering. Another important example is the *random tournament*, where every edge is given a random orientation (with probability $1/2$ in each direction).

In Chapter 2, we investigate a problem about tournaments posed by Linial and Morgenstern [72], who set out to determine the possible densities of cycles of length

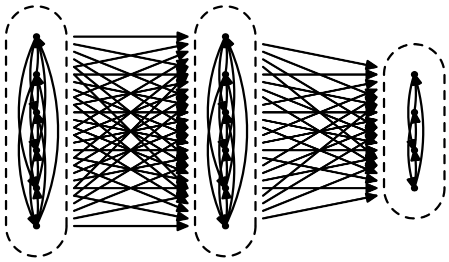


Figure 1.2: An illustration of the random blow up construction with three parts.

4 in a large tournament with a fixed density of cycles of length 3. The *(homomorphism) density* of the (directed) cycle C_ℓ of length ℓ in a tournament T, denoted by $t(C_\ell, T)$, is the probability that a random mapping from $V(C_\ell)$ to $V(T)$ is a homomorphism (i.e. arcs of C_ℓ map to arcs of T). Note that, for fixed ℓ, a tournament T on n vertices contains $t(C_\ell, T)n^\ell/\ell + O(n^{\ell-1})$ cycles of length ℓ. In fact, for $\ell \in \{3, 4, 5\}$, the error term is zero as every homomorphism of C_ℓ to T is injective.

It was shown in [30, Fact 1] that $t(C_3, T) \leqslant 1/8$ for every tournament T. Linial and Morgenstern proved that, for $d \in [0, 1/8]$, the asymptotically feasible densities of 4-cycles in tournaments with $t(C_3, T) = d + o(1)$ form an interval [72, Proof of Lemma 1.3]. They also showed that $t(C_4, T) \leqslant 2t(C_3, T)/3$, and that this bound is tight, thus determining the *maximum* 4-cycle density in a large tournament when $t(C_3, T)$ is fixed. Attempting to find the correct *minimum* 4-cycle density, they proved $t(C_4, T) \geqslant \frac{12t(C_3,T)^2}{1+16t(C_3,T)}$, but suspected that the true behaviour was more complex. They conjectured that the tournament that asymptotically minimises the density of 4-cycles is a blow-up of a transitive tournament with all but one part of equal size and one smaller part, with arcs inside each part oriented randomly. (They call this construction a *random blow-up*, see Figure 1.2.) If true, the structure of the extremal examples would parallel the solution of the Erdős–Rademacher problem. Figure 1.3 visualises the conjectured feasible region of 3-cycle and 4-cycle densities.

The bound $t(C_4, T) \geqslant \frac{12t(C_3,T)^2}{1+16t(C_3,T)}$ supports the conjecture of Linial and Morgenstern when $t(C_3, T) = 1/32 + o(1)$. We further confirm the conjecture in the case where the proposed extremal examples have two or three parts and provide a full description of extremal tournaments in the two-part case. In contrast to many of the recent proofs in this area that use the flag algebra method, our approach is based on the analysis of the spectrum of adjacency matrices of tournaments. We remark that descriptions of the extremal graphs for the Erdős–Rademacher problem have been given by Pikhurko and Razborov [82], and by Liu, Pikhurko, and Staden [73].

Now, we formally define the random blow-up construction from [72], parametrised

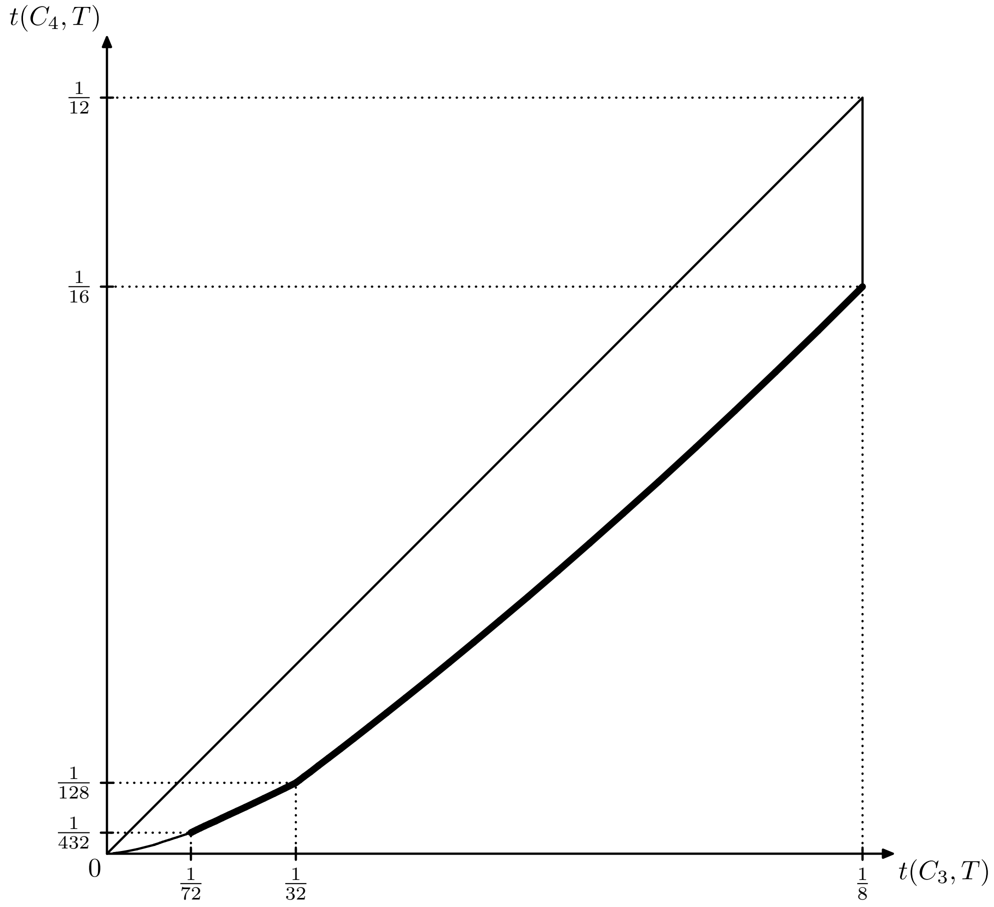


Figure 1.3: The conjectured region of asymptotically feasible densities of C_3 and C_4 in tournaments. The lower bound for $t(C_3, T) \in \{1/8, 1/32\}$ and the upper bound were proved in [72]. The rest of the lower bound is conjectured except for the part depicted in bold, which we prove.

by $z \in [0, 1]$, which will motivate the definition of the function g below. If $z = 0$, then the construction is simply a large transitive tournament. Otherwise, we let n be an integer chosen large with respect to $1/z$ and define an n-vertex tournament T as follows. The vertices of T are split into $\lfloor z^{-1} \rfloor + 1$ disjoint parts $V_1, \ldots, V_{\lfloor z^{-1} \rfloor + 1}$ such that $\lfloor z^{-1} \rfloor$ parts contain exactly $\lfloor zn \rfloor$ vertices and the remaining part contains the rest of the vertices (note that if z^{-1} and zn are integers, then the last part is empty). If two vertices v and v' respectively belong to distinct parts V_i and V_j with $i < j$, then the tournament T contains an edge from v to v'. If v and v' instead belong to the *same* part, then the edge between them is oriented from v to v' with probability $1/2$, i.e., the vertices of each part induce a randomly oriented tournament (see Figure 1.2 for an illustration with $z = 3/8$).

It is easy to see that $t(C_3, T) = t(C_4, T) = 0$ if $z = 0$, and that for any other z the expected value of $t(C_3, T)$ and $t(C_4, T)$ is a constant depending on z only. Furthermore, if $z \in (0, 1]$, it can be shown by writing the number of homomorphisms from C_3 and C_4 to T as a sum of indicator variables that $t(C_3, T)$ and $t(C_4, T)$ have $O(n^{-3})$ and $O(n^{-4})$ variance, respectively. Therefore, it follows from Chebyshev's inequality that with high probability,

$$t(C_3, T) = \frac{1}{8}\left(\lfloor z^{-1}\rfloor z^3 + \left(1 - \lfloor z^{-1}\rfloor z\right)^3\right) + o(1) \text{ and}$$

$$t(C_4, T) = \frac{1}{16}\left(\lfloor z^{-1}\rfloor z^4 + \left(1 - \lfloor z^{-1}\rfloor z\right)^4\right) + o(1).$$

The conjecture of Linial and Morgenstern [72] asserts that the above construction is asymptotically optimal. In light of this, we write *the regime of k parts* to denote the set of values of $t(C_3, T)$ between $1/(8k^2)$ and $1/(8(k-1)^2)$, corresponding to the range of values for which the above construction has its vertices split into k parts. In particular, our results focus on the regimes of two and three parts, which refer to values of $t(C_3, T)$ in the ranges $[1/32, 1/8]$ and $[1/72, 1/32]$, respectively.

To formally state the conjecture, define $g : [0, 1/8] \to [0, 1]$ as follows: $g(0) = 0$ and

$$g\left(\frac{1}{8}\left(\lfloor z^{-1}\rfloor z^3 + \left(1 - \lfloor z^{-1}\rfloor z\right)^3\right)\right) = \frac{1}{16}\left(\lfloor z^{-1}\rfloor z^4 + \left(1 - \lfloor z^{-1}\rfloor z\right)^4\right)$$

for $z \in (0, 1]$.

Conjecture 1.6 (Linial and Morgenstern [72, Conjecture 2.2][1])**.** *For every tournament T,*

$$t\left(C_4, T\right) \geqslant g\left(t(C_3, T)\right) + o(1).$$

Although $0 \leqslant t(C_3, T) \leqslant 1/8$ for every tournament T, the conjecture is currently only known to hold for tournaments with 3-cycle density asymptotically equal to 0, 1/8, or 1/32 [72].

In Chapter 2, we confirm the conjecture for all 3-cycle densities in the range $[1/72, 1/8]$ (Theorem 2.1). Using a different method, we characterise the asymptotic structure of extremal tournaments for densities in the range $[1/32, 1/8]$ (Theorem 2.8).

[1] Linial and Morgenstern phrased their conjecture equivalently in terms of subgraph counts as opposed to homomorphism densities; the quantities c_3 and c_4 in their paper are asymptotically equal to $2t(C_3, T)$ and $6t(C_4, T)$, respectively.

1.3 Permutations

Permutations, which describe the ways in which a sequence of items can be arranged, are an important object of study in combinatorics, cryptography, and algebra. Formally, a permutation of *order* (or *length*) k, also known as a k-permutation, is a bijection π from $[k]$ to $[k]$. The order of a permutation π is denoted by $|\pi|$.

Although there are many ways to represent a k-permutation π, the convention used in this thesis is to write $\pi = \pi(1)\pi(2)\dots\pi(k)$. For example, 2143 is the permutation $\pi : [4] \to [4]$ with $\pi(1) = 2$, $\pi(2) = 1$, $\pi(3) = 4$, and $\pi(4) = 3$. Permutations can also be represented pictorially, as shown in Figure 1.4.

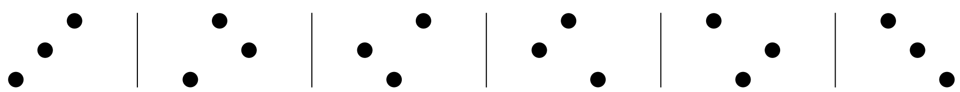


Figure 1.4: The six permutations of order 3. From left to right: 123, 132, 213, 231, 312, 321. A dot i-th from the left and j-th from the bottom indicates that $\pi(i) = j$.

We now define the natural notion of density in permutations. If $A = \{a_1, \dots, a_\ell\}$ is a subset of $[k]$ with $a_1 < \cdots < a_\ell$, the subpermutation of π *induced* by A is the unique permutation π' of order $|A| = \ell$ such that $\pi'(i) < \pi'(j)$ if and only if $\pi(a_i) < \pi(a_j)$ for every $i, j \in [\ell]$. For example, if $\pi = 521364$ and $A = \{1, 4, 5\}$, then A picks out the subsequence 536 from 521364, so the corresponding induced subpermutation is 213. More briefly, we can say the elements 536 of the permutation 521364 induce the permutation 213. Subpermutations are also often referred to as *patterns*. If π and Π are two permutations with $|\pi| \leqslant |\Pi| = n$, then the *pattern density* of π in Π, which is denoted by $d(\pi, \Pi)$, is the probability that the subpermutation of Π induced by a random $|\pi|$-element subset of $[n]$ is π. If $|\Pi| < |\pi|$, then we set $d(\pi, \Pi)$ to be 0. We often refer to pattern density simply as *density* in this thesis.

A well-known concept in the study of combinatorial densities is that of *quasirandomness*. Informally speaking, a combinatorial object is said to be quasirandom if it looks like a truly random object of the same kind. The theory of quasirandom graphs can be traced back to the work of Rödl [89], Thomason [95], and Chung, Graham, and Wilson [32] from the 1980s. It turned out that if any of a multitude of properties of random graphs involving subgraph density, edge distribution, and adjacency matrix eigenvalues are satisfied by a large graph, then all of the properties are satisfied. Results of a similar kind have been obtained for many other types of combinatorial objects, such as groups [56], hypergraphs [28, 54, 55, 62, 69], set systems [29], subsets of integers [31] and tournaments [23, 30, 35, 58].

One example of such a result is by Chung, Graham, and Wilson [32], who showed that if the edge density of a large graph G is p and the density of cycles of length 4 is $p^4 + o(1)$, then the density of all subgraphs is close to their expected density in a random graph. This can be seen as the case $H = C_4$ of the *forcing conjecture*, which strengthens a famous conjecture of Sidorenko [90]:

Conjecture 1.7 (Skokan and Thoma [91]). *Let G be a graph with edge density p, and let H be a bipartite graph that is not a tree. If $d(H, G) = (1 + o(1))p^{|E(H)|}$, then G is quasirandom in the sense that $d(F, G) = (1 + o(1))p^{|E(F)|}$ for all graphs F.*

In this section we are concerned with the quasirandomness of permutations as studied by Cooper [34]. Observe that the expected density of any permutation π in a large random permutation is $1/|\pi|!$. Based on this key property, we say that a sequence $\{\Pi_i\}_{i\in\mathbb{N}}$ of permutations is *quasirandom* if

$$\lim_{i\to\infty} d(\pi, \Pi_i) = \frac{1}{|\pi|!}$$

for every permutation π.

Our research is motivated by the following question of Graham (see [34, page 141]): *Is there an integer k such that a sequence $\{\Pi_i\}_{i\in\mathbb{N}}$ of permutations is quasirandom if and only if*

$$\lim_{i\to\infty} d(\pi, \Pi_i) = \frac{1}{k!}$$

for every k-permutation π? This question was answered affirmatively by Král' and Pikhurko [71], who established that $k = 4$ has this property. It was later noticed that this statement is equivalent to a result in statistics on non-parametric independence tests by Yanagimoto [100] that improved an older result of Hoeffding [65].

Theorem 1.8 (Král' and Pikhurko [71]). *A sequence $\{\Pi_i\}_{i\in\mathbb{N}}$ of permutations is quasirandom if and only if*

$$\lim_{i\to\infty} d(\pi, \Pi_i) = \frac{1}{4!}$$

for every 4-permutation π.

The statement of Theorem 1.8 does not hold for 3-permutations [33, 71]—in other words, there exists a non-quasirandom sequence of permutations in which the density of every 3-permutation converges to $1/3!$.

Theorem 1.8 says that if the limit densities of *all* 4-permutations in a sequence are equal to $1/4!$, then the sequence is quasirandom. Hence, it is natural to ask whether it is possible to replace the set of all 4-permutations in the statement of Theorem 1.8 with a smaller set. Inspecting the proof given in [71], Zhang [102] observed that there exists a 16-element set of 4-permutations with this property.

We identify several 8-element sets that have this property. In fact, the sets S that we identify have the stronger property that fixing the *sum* of densities of elements of S is enough to force quasirandomness; it is not necessary to fix the density of each individual element of S. In reference to the forcing conjecture of Skokan and Thoma, we say that a set S of k-permutations is Σ-*forcing* ("sum-forcing") if for every sequence $\{\Pi_i\}_{i\in\mathbb{N}}$ of permutations,

$$\lim_{i\to\infty}\sum_{\pi\in S} d(\pi,\Pi_i) = \frac{|S|}{k!} \text{ if and only if } \{\Pi_i\}_{i\in\mathbb{N}} \text{ is quasirandom.}$$

Our main result is the characterisation of all Σ-forcing sets of 4-permutations (Theorem 3.1 in Chapter 3). The first of the sets listed in Theorem 3.1 was previously identified by Bergsma and Dassios [4], who studied the Σ-forcing property in statistics.

1.3.1 Permutation flags

In Section 1.1 we introduced flag algebras for graphs. However, the flag algebra framework is applicable to a wide range of combinatorial structures, having originally been expressed by Razborov [85] (primarily a logician and theoretical computer scientist) in model-theoretic terms. As the results appearing in Chapter 3 rely heavily on flag algebras, in this subsection we will develop the flag algebra of permutations with an emphasis on practical applications. The exposition will be accompanied by an example of the semidefinite method.

Let $\mathcal{F}$ be the set of permutations, and let $\mathcal{F}_k$ be the set of permutations of order k. We first need to define the algebra $\mathcal{A}$. As the construction has essentially already been given in Section 1.1, and indeed will be given again when we construct the flag algebra $\mathcal{A}^\tau$ below, we simply describe $\mathcal{A}$ as the quotient of the space $\mathbb{R}\mathcal{F}$ by relations of the form

$$\pi = \sum_{\varphi\in\mathcal{F}_k} d(\pi,\varphi)\cdot\varphi, \tag{1.5}$$

with multiplication given by

$$\pi_1 \times \pi_2 = \sum_{\varphi\in F_\ell} d(\pi_1,\pi_2;\varphi)\cdot\varphi,$$

where $\ell = |\pi_1| + |\pi_2|$ and $d(\pi_1,\pi_2;\varphi)$ is the probability that $|\pi_1|$ randomly chosen elements of φ induce π_1 and the remaining $|\pi_2|$ elements induce π_2.

Fix a permutation τ for the rest of this section. A τ-*rooted permutation* is a permutation with $|\tau|$ distinguished elements such that the subpermutation induced by these elements is τ; the distinguished elements are referred to as *roots.* When presenting rooted permutations, the roots will be underlined. For example, $\underline{12}3$, $1\underline{23}$, and $\underline{1}2\underline{3}$ are distinct τ-rooted permutations for $\tau = 12$. Define $\mathcal{F}^\tau$ to be the set of τ-rooted permutations, $\mathcal{F}^\tau_k$ to be the set of τ-rooted permutations of order k, and $\mathbb{R}\mathcal{F}^\tau$ to be the set of formal linear combinations of elements of $\mathcal{F}^\tau$ with real coefficients.

For two τ-rooted permutations π and φ with $|\pi| \leqslant |\varphi|$, define the *density* $d(\pi, \varphi)$ of π in φ to be the probability that a randomly chosen subset of $|\pi| - |\tau|$ non-root elements of φ, together with the root of φ, induce π. For example $d(\underline{12}3, 1\underline{2}6\underline{4}35) = 1/4$ because the only τ-rooted subpermutation of $1\underline{2}6\underline{4}35$ that induces $\underline{12}3$ is the subpermutation corresponding to $\underline{24}5$. Furthermore, given τ-rooted permutations π_1, π_2, and φ with $|\varphi| = |\pi_1| + |\pi_2| - |\tau|$, we define $d(\pi_1, \pi_2; \varphi)$ to be the probability that a randomly chosen subset of $|\pi_1| - |\tau|$ unrooted elements of φ together with the root of φ induce π_1, and that the remaining $|\pi_2| - |\tau|$ unrooted elements of φ together with the root of φ induce π_2.

Closely paralleling the exposition for graphs, we observe two important relationships between pattern densities. Namely, given a large τ-rooted background permutation Π and τ-rooted permutations π, π_1, and π_2 with $|\pi| \leqslant k$, let $\ell = |\pi_1| + |\pi_2| - |\tau|$. Then

$$\pi = \sum_{\varphi \in F^\tau_k} d(\pi, \varphi) \cdot \varphi, \text{ and} \tag{1.6}$$

$$d(\pi_1, \Pi) \cdot d(\pi_2, \Pi) = (1 + o(1)) \sum_{\varphi \in F^\tau_\ell} d(\pi_1, \pi_2; \varphi) \cdot d(\varphi, \Pi),$$

for any permutation Π with $|\Pi| \to \infty$. Thus, we can construct a flag algebra $\mathcal{A}^\tau$ from the space $\mathbb{R}\mathcal{F}^\tau$ of formal linear combinations of elements of $\mathcal{F}^\tau$ by taking the quotient of $\mathbb{R}\mathcal{F}^\tau$ by relations of the form (1.6) and defining a multiplication operation

$$\pi_1 \times \pi_2 = \sum_{\varphi \in F^\tau_\ell} d(\pi_1, \pi_2; \varphi) \cdot \varphi. \tag{1.7}$$

An example of this multiplication is

$$\underline{1}2 \times \underline{1}23 = 1 \cdot \underline{1}234 + \frac{2}{3} \cdot \underline{1}243 + \frac{2}{3} \cdot \underline{1}324 + \frac{1}{3} \cdot \underline{1}342 + \frac{1}{3} \cdot \underline{1}423 + 0 \cdot \underline{1}432.$$

As in the case of graphs, there is also a linear *averaging* (or *unlabelling*) operator $[\![\cdot]\!] : \mathcal{A}^\tau \to \mathcal{A}$, defined on the elements of $\pi \in \mathcal{F}^\tau$ by $[\![\pi]\!] = \binom{|\pi|}{|\tau|}^{-1} \cdot \pi^\varnothing$, where $\pi^\varnothing \in \mathcal{F}$ is the underlying (unrooted) permutation corresponding to π. This is because only one of the $\binom{|\pi|}{|\tau|}$ choices of $|\tau|$ roots in $\pi^{\varnothing}$ leads to a permutation isomorphic to π, unlike in the case of graphs.

We now turn towards the practical aspects of the flag algebra method. An $n \times n$ symmetric real matrix M is said to be *positive semidefinite* if $x^T M x \geqslant 0$ for all $x \in \mathbb{R}^n$, where x^T denotes the vector transpose of x. In particular, if Π^τ is a τ-rooted permutation, we can take each entry of x to be a flag density of the form $d(\psi, \Pi^\tau)$ for some τ-rooted permutation ψ. Moreover, averaging the expression $x^T M x$ over τ-flags Π^τ with underlying (unrooted) permutation Π preserves non-negativity, so

$$[\![x^T M x]\!] \geqslant 0. \tag{1.8}$$

Finding positive semidefinite matrices M that generate useful inequalities is the core of the so-called *semidefinite method* in flag algebras.

Let us consider a question along the lines of Mantel's Theorem (Theorem 1.4). Suppose we wish to find an upper bound on the density of the permutation 12 in a large permutation Π satisfying $d(123, \Pi) = 0 + o(1)$. Using our shorthand of drawing a permutation (see Figure 1.4) to represent its density in Π, our problem is to maximise 12 given that $123 = 0$. Applying (1.5) with $\pi = 12$ and $k = 3$, we have

$$12 = \sum_{\varphi \in \mathcal{F}_3} d\left(12, \varphi\right) \cdot \varphi = 123 + \frac{2}{3} 132 + \frac{2}{3} 213 + \frac{1}{3} 231 + \frac{1}{3} 312 + 0\, 321. \tag{1.9}$$

As $123 = 0$ and $123 + 132 + 213 + 231 + 312 + 321 = 1$, we immediately obtain the non-trivial bound

$$12 \leqslant \max\left\{\frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1}{3}, 0\right\} = \frac{2}{3},$$

with equality holding if and only if $132 + 213 = 1$. However, it was proved by Albert et al. [1] that there is no large permutation Π with $132 + 213 = 1$, so the bound $12 \leqslant 2/3$ is not tight. Indeed, if $|\Pi| \geqslant 5$, the non-asymptotic inequality $d(132, \Pi) + d(213, \Pi) < 1$ is implied by the Erdős–Szekeres Theorem [45], which states that every permutation of length at least 5 contains 123 or 321 as a subpermutation.

To improve the bound on 12, the idea is to use (1.8) to prove an inequality

$$0 \leqslant \sum_{\varphi \in \mathcal{F}_3} \alpha_\varphi \cdot \varphi,$$

where α_φ is a (possibly negative) constant depending on φ. For instance, let M be a 4×4 positive semidefinite matrix and let $x^T = (\underline{1}2, 2\underline{1}, \underline{2}1, 1\underline{2})$, where the unfilled circles represent rooted elements (so $\underline{1}2$ depicts the permutation $\underline{1}2$). Then by (1.7), the product $x^T M x$ is a linear combination of 1-rooted permutations of length 3. Applying (1.8) to unlabel $x^T M x$ and adding the resulting inequality to (1.9) gives us the bound

 $$\leqslant \max_{\varphi \in \mathcal{F}_3} \left[d(\pi, \varphi) + \alpha(\varphi) \right].$$

Therefore, if $\alpha(132)$ and $\alpha(213)$ are both negative, and the other values of α are not too large, we have a chance to improve the bound on .

Following the template above, a typical application of flag algebras to maximise the density of a small permutation π (or a linear combination of permutations) in a large permutation Π under some conditions is as follows:

- Choose some integer $k > |\pi|$. Using (1.5), write $d(\pi, \Pi)$ as a linear combination of densities of k-permutations in Π.

- Choose a small permutation τ and an integer $\ell < |\tau|$ such that $2\ell - |\tau| = k$. Let x be the vector whose entries are the elements of $\mathcal{F}_\ell^\tau$ (the τ-rooted permutations of order ℓ).

- Find a positive semidefinite $|\mathcal{F}_\ell^\tau| \times |\mathcal{F}_\ell^\tau|$ matrix M that minimises the quantity $\max_{\varphi \in \mathcal{F}_k} [d(\pi, \varphi) + \alpha(\varphi)]$, where $\alpha(\varphi)$ is the coefficient of φ in $[\![x^T M x]\!]$.

Sliačan and Stromquist [93] provide an example of such a proof (with a 4×4 matrix M) that can be verified by hand. More generally, however, it is not clear how to guess the matrix M. The problem of finding the optimal matrix is an example of a *semidefinite program* (SDP). Thankfully, the literature on the field of semidefinite programming is extensive (see [52] for an overview). Several solvers such as CSDP [10] and SDPA [50] pre-date the development of flag algebras, and there are also software packages customised for use in flag algebras, including Flagmatic 2.0 [97] and Permpack [92].

1.4 Trees

An important class of graphs is the set of *trees*. A tree is a connected graph that contains no cycles, or equivalently, a graph in which there is a unique path between every pair of vertices. Trees have been studied since at least 1857 [24] and are of fundamental importance as a data structure in computer science and algorithm design [68].

Trees are examples of *sparse* graphs. While the distinction between dense and sparse graphs is, to some degree, a matter of perspective, trees fall firmly into the latter class. It is easily shown by induction that every tree on n vertices has exactly

$n-1$ edges. In particular, trees have $o(n^2)$ edges and the density of edges in T converges to 0. Therefore, when we randomly sample k vertices from a tree T for some fixed k, we expect to see $o(1)$ edges in the subgraph of T induced by these k vertices. This implies that the limit density of the empty graph $\overline{K_n}$ in T is 1, and the limit density of every other graph in T is 0. In particular, every sequence of trees converges to the graphon that is zero everywhere.

Clearly, the standard definition of subgraph density, which normalises subgraph counts by $\binom{n}{k}$, does not capture any meaningful information about trees. There have been several attempts to address this problem. On the side of developing a general limit theory for sparse graphs, multiple competing definitions of convergence exist, although each has its drawbacks. Perhaps the simplest of these notions is that of *Benjamini–Schramm convergence* [3], which samples a random vertex v from a large, sparse graph G and looks at the distribution of subgraphs induced by the set of vertices at distance at most d from v. Research in the area of sparse limits primarily focusses on understanding the different notions of convergence and their relationships to each other, see [9, 11, 12, 42, 60, 80, 81].

Regarding the study of graph densities in the class of trees specifically, Bubeck and Linial [22] defined an alternative normalisation of subgraph counts as follows. Let T be a tree. We denote by $Z_k(T)$ the number of k-vertex subtrees in T. An *embedding* of a tree S in T is a subtree of T isomorphic to S. Note that in our usage, an embedding can be associated with (possibly multiple) injective homomorphisms from S to T, and all injective homomorphisms from S to T with the same image are associated with a single embedding. The *density* of a k-vertex tree S in T, denoted by $d(S,T)$, is the number of embeddings of S in T divided by $Z_k(T)$; if the number of vertices of T is less than k, we set $d(S,T) = 0$.

From here, the theory of *tree profiles* can be developed. The *k-profile* of a tree T, denoted by $p^{(k)}(T)$, is the vector whose entries are indexed by the set of non-isomorphic k-vertex trees, where the entry of $p^{(k)}(T)$ indexed by a tree S is equal to $d(S,T)$. Note that $p^{(k)}(T)$ is the zero vector when the number of vertices of T is less than k, and the entries of $p^{(k)}(T)$ sum to 1 when the number of vertices of T is at least k. We say that a sequence $(T_n)_{n\in\mathbb{N}}$ of trees is *convergent* if the k-profiles $p^{(k)}(T_n)$ converge entrywise for every $k \in \mathbb{N}$. As $p^{(k)}(T_n)$ is a $[0,1]$-valued vector and any countable product of sequentially compact spaces is itself sequentially compact, every sequence of trees has a convergent subsequence.

Let N_k be the number of non-isomorphic k-vertex trees. The main result of [22] is that for every integer k, the limit set of k-profiles of trees (seen as a subset of $[0,1]^{N_k}$) is convex. This is in stark contrast to the case of graph profiles, where not only is the limit set of k-profiles non-convex, describing its convex hull is not computationally feasible [61].

It is natural to check whether existing questions about graph densities have an analogue in the case of trees. For example, perhaps the most simple question one can ask about limit densities of graphs is:

What is the maximum possible density of a fixed graph H in a large graph G?

This maximum is called the *inducibility* of H, and was introduced by Pippenger and Golumbic [83]; also see [20, 46, 59, 63, 64, 70, 99, 101] for graph inducibility and [1, 2, 84, 93] for the equivalent notion in permutations. We remark that the dual problem of finding the *minimum* possible density of H in a large graph G is closely related to the influential conjecture of Sidorenko [90] mentioned in Section 1.3. Mimicking the literature on graphs, the inducibility of a tree S is defined as the maximum limit density of S in a convergent sequence of trees. In other words, the inducibility of S is equal to

$$\limsup_{n\to\infty} \max\{d(S,T) : T \text{ is an } n\text{-vertex tree}\}.$$

We note that the definition of tree inducibility used here is that of Bubeck and Linial [22], which differs slightly from the definition used in [36, 37, 40, 41].

Clearly, paths have inducibility 1 since every subtree of a path is a path. Similarly, stars have inducibility 1 since every subtree of a star is a star. The second result proved by Bubeck and Linial in [22] is that paths and stars are the *only* trees with inducibility 1. Motivated by this result, they asked [22, Problem 4] whether there are additional trees with inducibility arbitrarily close to 1, or if not, whether there are infinitely many trees with inducibility bounded away from 0 by a fixed constant:

- Does there exist $\varepsilon > 0$ such that the inducibility of every tree that is neither a star nor a path is at most $1 - \varepsilon$?
- Does there exist $\varepsilon > 0$ such that there are infinitely many trees with inducibility at least ε?

We answer both these questions affirmatively (Theorems 4.1 and 4.2 in Chapter 4). Both theorems give explicit values for ε, although we make no attempt to optimise these values.

Another property related to profiles that holds for general graphs is the existence of *universal graphs*, that is, large graphs in which the limit density of every graph is positive. It is easily shown that $G_{n,p}$ is an example of a universal graph. Bubeck and Linial [22, Problem 5] asked whether there exist universal trees:

- Does there exist a convergent sequence $(T_n)_{n\in\mathbb{N}}$ of trees in which the limit density $\lim_{n\to\infty} d(S, T_n)$ of every tree S is positive?

Our final result is an explicit construction of such a sequence of trees (Theorem 4.3).

Regarding the state of the other problems appearing in [22], Bubeck, Edwards, Mania and Supko [21] and Czabarka, Székely and Wagner [38] independently resolved [22, Problem 3] by showing that if the limit density of a k-vertex path P_k in a (convergent) sequence of trees equals 0, then the limit density of the k-vertex star S_k equals 1. Further, results on 5-profiles of trees can be found in [21], where additional questions raised in [22, Problems 1 and 7] have been answered.

Chapter 2

Cycles in Tournaments

2.1 Overview

As discussed in Section 1.3, Linial and Morgenstern [72] conjectured that a particular construction minimises the density of 4-cycles among all tournaments with a fixed density of 3-cycles. In this chapter, we confirm the conjecture when the homomorphism density of 3-cycles is at least $1/72$:

Theorem 2.1. *Let T be a tournament with $t(C_3, T) \geqslant 1/72$. Then $t(C_4, T) \geqslant g(t(C_3, T)) + o(1)$.*

This result is equivalent to Theorem 2.14. Additionally, when $t(C_3, T) \geqslant 1/32$, Theorem 2.8 fully characterises the extremal tournaments. The proofs of both theorems rely on spectral analysis of adjacency matrices of tournaments.

2.2 Preliminaries

In this section, we introduce the notation used throughout the chapter, as well as the notions of tournament matrices and tournament limits. The set of integers $1, \ldots, n$ is denoted by $[n]$. Some of the matrices that we consider have complex eigenvalues and the complex unit will be denoted by ι. If A is a matrix (or a vector), then we write A^T for its transpose and A^* for its conjugate transpose; in particular, if A is real, then $A^T = A^*$. The *trace* of a square matrix A is the sum of the entries in its diagonal and is denoted by $\operatorname{Tr} A$. We let $\langle \cdot \mid \cdot \rangle$ denote the standard inner (dot) product on $\mathbb{R}^n$. We use $\mathbb{J}_n$ to denote the square matrix of order n such that each entry of $\mathbb{J}_n$ is equal to one; if n is clear from the context, we will omit the subscript. Note that $\mathbb{J}_n$ has one eigenvalue equal to n and the remaining $n-1$ eigenvalues are equal to zero. The n-dimensional column vector with all entries equal to one is denoted by $\vec{j}_n$ and we again omit the subscript when n is clear from the context. Note that $\mathbb{J}_n = \vec{j}_n \vec{j}_n^T$.

2.2.1 Tournament matrices

We say that a square matrix A of order n is a *tournament matrix* if A is non-negative and $A + A^T = \mathbb{J}$; in particular, if A is a tournament matrix, then each diagonal entry of A is equal to $1/2$. Every n-vertex tournament T can be associated with a tournament matrix A of order n, which we refer to as the *adjacency matrix* of T, in the following way. Each diagonal entry A is equal to $1/2$ and, for $i \neq j$, the entry of A in the i-th row and the j-th column (denoted $A_{i,j}$) is equal to 1 if T contains an arc from the i-th vertex to the j-th vertex, and it is equal to 0 otherwise. The following proposition readily follows.

Proposition 2.2. *Let T be a tournament on n vertices, A be the adjacency matrix of T and $\ell \geqslant 3$. The number of homomorphisms of C_ℓ to T is $\mathrm{Tr}\, A^\ell + O(n^{\ell-1})$.*

Recall that the trace of a matrix is equal to the sum of its eigenvalues and that the eigenvalues of the ℓ-th power of a matrix are the ℓ-th powers of its eigenvalues. In view of Proposition 2.2, for $\ell \geqslant 1$, we define $\sigma_\ell(A)$ for a square matrix A of order n to be

$$\sigma_\ell(A) = \frac{1}{n^\ell} \sum_{i=1}^{n} \lambda_i^\ell = \frac{1}{n^\ell} \operatorname{Tr} A^\ell$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are the eigenvalues of A. Note that the normalisation of the sum is chosen in such a way that $\sigma_1(A) = 1/2$ for every tournament matrix A.

Next, we argue that Conjecture 1.6 is equivalent to the following.

Conjecture 2.3. *If A is a tournament matrix, then $\sigma_4(A) \geqslant g(\sigma_3(A))$.*

Indeed, Conjecture 2.3 implies Conjecture 1.6 by Proposition 2.2. In the other direction, suppose that there exists a tournament matrix A of order n such that $\sigma_4(A) < g(\sigma_3(A))$. We consider the following (random) tournament T with $k \cdot n$ vertices, $k \in \mathbb{N}$: the vertices of T are split into n sets $V_1, \ldots, V_n$, each containing k vertices, and a vertex $v \in V_i$ is joined by an arc to a vertex $v' \in V_j$ with probability $A_{i,j}$; note that v' is joined by an arc to v with probability $A_{j,i} = 1 - A_{i,j}$, i.e., the tournament T is well defined. Since n is fixed, for $\ell \in \{3, 4\}$ and large k, the number of homomorphisms from C_ℓ to T is $\sigma_\ell(A)(nk)^\ell + O(k^{\ell-1})$ with high probability and so Conjecture 1.6 fails for $t(C_3, T) \approx \sigma_3(A)$.

We conclude this subsection by establishing the following lemma. A similar result has been proved by Brauer and Gentry [19], but for a slightly different definition of a tournament matrix.

Lemma 2.4. *If A is a tournament matrix, then every eigenvalue of A has non-negative real part.*

Proof. Let λ be any eigenvalue of A and let v be a corresponding eigenvector. Observe that

$$\begin{aligned}0 \leqslant \overline{(\vec{j}^T v)}(\vec{j}^T v) = v^* \mathbb{J} v = v^*(A + A^T)v = v^*(Av) + (v^* A^T)v \\ = v^*(\lambda + \overline{\lambda})v = (\lambda + \overline{\lambda})v^* v\,.\end{aligned}$$

Since v^*v is a non-negative real number, it follows that $\lambda + \overline{\lambda}$ is a non-negative real. In particular, the real part of λ is non-negative. □

2.2.2 Tournament limits

As we have seen in Section 1.1, the theory of graph limits provides analytic tools to represent and analyse large graphs. In an analogous way, one can develop a limit theory for tournaments, in which the foundational results for graphons translate to similar statements for tournament limits with essentially the same proofs. Below, we define tournament limits and outline some of the basic results that we will use.

A *tournament limit* is a measurable function $W : [0,1]^2 \to [0,1]$ such that $W(x,y) + W(y,x) = 1$ for all $(x,y) \in [0,1]^2$. One can define the density of the cycle C_ℓ in W as follows:

$$t(C_\ell, W) = \int_{x_1,\dots,x_\ell \in [0,1]} W(x_1,x_2)W(x_2,x_3)\cdots W(x_{\ell-1},x_\ell)W(x_\ell,x_1)\,\mathrm{d}x_1 \cdots x_\ell\,.$$

Note that any n-dimensional tournament matrix A can be represented by a tournament limit W_A by dividing $[0,1]$ into sets $I_1,\dots,I_n$ of measure $1/n$ and setting W equal to $A_{i,j}$ on the set $I_i \times I_j$. It is easily observed that $t(C_\ell, W_A)$ is precisely $\sigma_\ell(A)$. The following proposition links densities of cycles in tournament limits and in tournaments.

Proposition 2.5. *The following two statements are equivalent for every sequence $(s_\ell)_{\ell \geqslant 3}$ of non-negative reals:*

- *There exists a tournament limit W such that $t(C_\ell, W) = s_\ell$ for every $\ell \geqslant 3$.*
- *There exists a sequence $(T_i)_{i \in \mathbb{N}}$ of tournaments with increasing orders such that*

  $$\lim_{i\to\infty} t(C_\ell, T_i) = s_\ell$$

  for every $\ell \geqslant 3$.

The first statement easily implies the second by letting T_i be a *W-random tournament* of order i; that is, we let $x_1, \ldots, x_i$ be independent uniformly random points of $[0,1]$ and join the ith vertex to the jth with probability $W(x_i, x_j)$. For the other direction, the tournament limit W can be constructed by adapting one of the existing arguments in the graph case, e.g., the argument of Lovász and Szegedy [76] based on weak regularity that was briefly discussed in Subsection 1.1.1. In light of Proposition 2.5, Conjecture 1.6 is equivalent to the following.

Conjecture 2.6. *For every tournament limit W,*

$$t(C_4, W) \geqslant g(t(C_3, W)).$$

The notion of regularity decompositions of graphs readily extends to tournaments. We present here the notion of weak regular partitions introduced by Frieze and Kannan in [49] adapted to the setting of tournament limits. We use $|X|$ to denote the measure of a measurable subset X of $[0,1]$. Given a tournament limit W and $\varepsilon \in (0,1)$, a partition $Z_1, \ldots, Z_n$ of $[0,1]$ into sets of measure $1/n$ is *weak ε-regular* for W if

$$\left| \int_{(x,y)\in X\times Y} W(x,y)\,\mathrm{d}x\,\mathrm{d}y - \sum_{i,j=1}^{n} A_{i,j}\cdot |Z_i \cap X| \cdot |Z_j \cap Y| \right| \leqslant \varepsilon$$

for all measurable subsets X and Y of $[0,1]$, where A is the tournament matrix defined by

$$A_{i,j} = \frac{\int_{(x,y)\in Z_i\times Z_j} W(x,y)\,\mathrm{d}x\,\mathrm{d}y}{|Z_i|\cdot|Z_j|}.$$

We say that a tournament limit W' is a *weak ε-regular approximation* of W if there exists a weak ε-regular partition $\{Z_1, \ldots, Z_n\}$ such that $W'(x,y) = A_{i,j}$ for $(x,y) \in Z_i \times Z_j$, $i,j \in [n]$, where A is the tournament matrix associated with the partition.

The results of Frieze and Kannan [49] adapted to the setting of tournament limits and the corresponding arguments for graph limits [76] yield the following: for every tournament limit W and $k \geqslant 2$, there exists a weak $1/k$-regular partition $\{Z_{k,1}, \ldots, Z_{k,n_k}\}$ with the following properties: (a) n_k is bounded by a function of k, and (b) the partitions are *refining* in the sense that, for every $k < k'$ and $i' \in [n_{k'}]$, the set $Z_{k',i'}$ is contained in $Z_{k,i}$ for some $i \in [n_k]$. It can be shown analogously to the graph case that the corresponding weak $1/k$-regular approximations converge to W in L_1. In particular,

$$\lim_{k\to\infty} \sigma_\ell(A_k) = t(C_\ell, W)$$

for every $\ell \geqslant 3$, where, for $k \in \mathbb{N}$, A_k is the tournament matrix associated with the partition $Z_{k,1}, \ldots, Z_{k,n_k}$.

We conclude with a proposition on the density of C_3 in a weak regular approximation of a tournament limit. The proof of the proposition is also valid in a more general setting of step approximations of tournament limits, which need not be weak regular, however, we prefer stating the proposition in the restricted setting of weak regular approximations to avoid introducing additional notation not needed for our exposition.

Proposition 2.7. *Let W be a tournament limit and W' a weak ε-regular approximation of W. Then $t(C_3, W) \leqslant t(C_3, W')$.*

Proof. We begin by showing that any tournament limit U satisfies

$$t(C_3, U) = \frac{1}{2} - \frac{3}{2} \int_{x \in [0,1]} \left(\int_{y \in [0,1]} U(x, y) \, \mathrm{d}y \right)^2 \mathrm{d}x \,. \tag{2.1}$$

To do this, we derive two identities based on the symmetry of variables x, y and z. Firstly,

$$\begin{aligned}
1 &= \int_{x,y,z \in [0,1]} (U(x,y) + U(y,x))(U(x,z) + U(z,x))(U(y,z) + U(z,y)) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&= 2 \int_{x,y,z \in [0,1]} U(x,y)U(y,z)U(z,x) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&\quad + 6 \int_{x,y,z \in [0,1]} U(x,y)U(y,z)U(x,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&= 2t(C_3, U) + 6 \int_{x,y,z \in [0,1]} U(x,y)U(y,z)U(x,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \,,
\end{aligned} \tag{2.2}$$

and similarly,

$$\begin{aligned}
&\int_{x \in [0,1]} \left(\int_{y \in [0,1]} U(x,y) \, \mathrm{d}y \right)^2 \mathrm{d}x \\
&\qquad = \int_{x,y,z \in [0,1]} U(x,y)U(x,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&\qquad = \int_{x,y,z \in [0,1]} U(x,y)U(x,z)(U(y,z) + U(z,y)) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&\qquad = 2 \int_{x,y,z \in [0,1]} U(x,y)U(x,z)U(y,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \,.
\end{aligned} \tag{2.3}$$

Noticing that the integrals on the last lines in (2.2) and (2.3) are the same, the equality (2.1) is obtained. Hence, the inequality from the statement of the proposition is equivalent to

$$\int_{x\in[0,1]} f'(x)^2 \,\mathrm{d}x \leqslant \int_{x\in[0,1]} f(x)^2 \,\mathrm{d}x\,, \tag{2.4}$$

where for brevity we have set

$$f'(x) = \int_{y\in[0,1]} W'(x,y)\,\mathrm{d}y \qquad \text{and} \qquad f(x) = \int_{y\in[0,1]} W(x,y)\,\mathrm{d}y\,.$$

Since

$$f'(x) = \frac{1}{|Z_i|}\int_{x'\in Z_i} f(x')\,\mathrm{d}x'$$

for every x in a part Z_i of the weak ε-regular partition defining the tournament limit W',

$$\begin{aligned}\int_{x\in Z_i} f'(x)^2\,\mathrm{d}x &= |Z_i|\left(\frac{1}{|Z_i|}\int_{x'\in Z_i} f(x')\,\mathrm{d}x'\right)^2\\ &\leqslant \int_{x\in Z_i} f(x)^2\,\mathrm{d}x\,,\end{aligned} \tag{2.5}$$

where the last line follows from the Cauchy–Schwarz inequality. Summing the inequalities obtained from applying (2.5) to each Z_i yields (2.4). □

2.3 Regime of two parts

Our goal in this section is to prove Conjecture 2.3 in the case that $\sigma_3(A) \geqslant 1/32$, as well as describe the tournament matrices which achieve equality. We then apply this result to characterise the extremal tournament limits for Conjecture 2.6 for $t(C_3, W) \geqslant 1/32$. Throughout the proof of the next theorem, we will frequently use the property that the trace of a product of matrices is invariant under "cyclic permutations", i.e., $\mathrm{Tr}\,(M_1M_2\cdots M_k) = \mathrm{Tr}\,(M_2\cdots M_kM_1)$.

Theorem 2.8. *Let A be a tournament matrix of order n. If $\sigma_3(A) \geqslant 1/32$, then $\sigma_4(A) \geqslant g(\sigma_3(A))$ and equality holds if and only if there exists a vector $z \in \mathbb{R}^n$ such that $A_{i,j} = 1/2 + z_i - z_j$ for $i, j \in [n]$.*

Proof. Fix a tournament matrix A of order n. Let $B = \mathbb{J} - 2A$. Note that B is a skew-symmetric matrix, i.e., $B = -B^T$. It follows (see, e.g., [51, p. 293]) that B can be written as $B = ULU^T$ where the columns $v_1, v_2, \ldots, v_n$ of U form an orthonormal basis of $\mathbb{R}^n$ and L has the form

$$L = \begin{bmatrix} 0 & \lambda_1 n & 0 & 0 & \cdots & 0 & 0 \\ -\lambda_1 n & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 n & \cdots & 0 & 0 \\ 0 & 0 & -\lambda_2 n & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_k n \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k n & 0 \end{bmatrix}$$

if n is even, and

$$L = \begin{bmatrix} 0 & \lambda_1 n & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\lambda_1 n & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 n & \cdots & 0 & 0 & 0 \\ 0 & 0 & -\lambda_2 n & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_k n & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k n & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}$$

otherwise, where $k = \lfloor n/2 \rfloor$ and $\lambda_1, \ldots, \lambda_k$ are real numbers. (Note that they are not the eigenvalues of B.) Since replacing v_{2i-1} and v_{2i} with $v_{2i-1}\cos\beta + v_{2i}\sin\beta$ and $v_{2i}\cos\beta - v_{2i-1}\sin\beta$, respectively, does not change the matrix B (this corresponds to rotating the basis inside the plane spanned by v_{2i-1} and v_{2i}), we can assume that the vectors $v_2, v_4, \ldots, v_{2k}$ are orthogonal to the vector $\vec{j}$. Set $\alpha_i = \cos^{-1}\left\langle v_{2i-1} \mid n^{-1/2}\vec{j}\right\rangle$ for $i \in [k]$, and additionally set $\alpha_{k+1} = \cos^{-1}\left\langle v_{2k+1} \mid n^{-1/2}\vec{j}\right\rangle$ if n is odd.

We next examine $\operatorname{Tr} A^3$ and $\operatorname{Tr} A^4$ in terms of $\mathbb{J}$ and B. We start with the trace of A^3:

$$8\operatorname{Tr} A^3 = \operatorname{Tr}(\mathbb{J} - B)^3 = \operatorname{Tr}\mathbb{J}^3 - 3\operatorname{Tr}\mathbb{J}^2 B + 3\operatorname{Tr}\mathbb{J}B^2 - \operatorname{Tr} B^3 .$$

Since both B and B^3 are skew-symmetric, it follows that $\operatorname{Tr}\mathbb{J}^2 B = 0$ and $\operatorname{Tr} B^3 = 0$. We next analyse the term $\operatorname{Tr}\mathbb{J}B^2$. Since $v_1, \ldots, v_n$ are mutually orthogonal and $v_2, v_4, \ldots, v_{2k}$ are orthogonal to $\vec{j}$, we have

$$\operatorname{Tr}\mathbb{J}B^2 = \frac{1}{n}\operatorname{Tr}\mathbb{J}^2 B^2 = \frac{1}{n}\operatorname{Tr}\mathbb{J}B^2\mathbb{J} = \frac{1}{n}\operatorname{Tr}\mathbb{J}(ULU^T)^2\mathbb{J} = \frac{1}{n}\operatorname{Tr}\mathbb{J}UL^2U^T\mathbb{J}$$

$$= -n^2\sum_{i=1}^{k}\lambda_i^2\left\langle v_{2i-1} \mid \vec{j}\right\rangle^2 = -n^3\sum_{i=1}^{k}\lambda_i^2\cos^2\alpha_i .$$

Hence, we obtain

$$8\sigma_3(A) = 1 - 3\sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i. \tag{2.6}$$

Similarly, we can express the trace of A^4 as follows:

$$16 \operatorname{Tr} A^4 = \operatorname{Tr} \mathbb{J}^4 - 4 \operatorname{Tr} \mathbb{J}^3 B + 4 \operatorname{Tr} \mathbb{J}^2 B^2 + 2 \operatorname{Tr} \mathbb{J} B \mathbb{J} B - 4 \operatorname{Tr} \mathbb{J} B^3 + \operatorname{Tr} B^4 .$$

Since B and B^3 are skew-symmetric, it follows that $\operatorname{Tr} \mathbb{J}^3 B = 0$, $\operatorname{Tr} \mathbb{J} B \mathbb{J} B = 0$ and $\operatorname{Tr} \mathbb{J} B^3 = 0$. Also, $\operatorname{Tr} B^4 = 2n^4 \sum_{i=1}^{k} \lambda_i^4$ by the cyclic property. Consequently,

$$16\sigma_4(A) = 1 - 4\sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i + 2\sum_{i=1}^{k} \lambda_i^4. \tag{2.7}$$

Recall that, if $\sigma_3(A) \in [1/32, 1/8]$, then $g(\sigma_3(A)) = \frac{1}{16}(z^4 + (1-z)^4)$ where $z \in [1/2, 1]$ such that $\sigma_3(A) = \frac{1}{8}(z^3 + (1-z)^3)$. So, for $\sigma_3(A)$ in the considered range, we have

$$8\sigma_3(A) = z^3 + (1-z)^3 = 1 - 3(z - z^2) .$$

By comparing this equation to (2.6), it must be the case that

$$z - z^2 = \sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i . \tag{2.8}$$

It follows that

$$\begin{aligned} 16g(\sigma_3(A)) &= z^4 + (1-z)^4 = 1 - 4z + 6z^2 - 4z^3 + 2z^4 \\ &= 1 - 4(z - z^2) + 2(z - z^2)^2 \\ &= 1 - 4\sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i + 2\left(\sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i\right)^2 . \end{aligned}$$

Combining this with (2.7), we see that the inequality $\sigma_4(A) \geqslant g(\sigma_3(A))$ holds if and only if

$$\sum_{i=1}^{k} \lambda_i^4 \geqslant \left(\sum_{i=1}^{k} \lambda_i^2 \cos^2 \alpha_i\right)^2 , \tag{2.9}$$

and $\sigma_4(A) = g(\sigma_3(A))$ if and only if (2.9) holds with equality.

Since $v_1, \ldots, v_n$ form an orthonormal basis of $\mathbb{R}^n$,

$$\sum_{i=1}^{n} \left\langle v_i \mid n^{-1/2}\vec{j}\right\rangle^2 = \left\langle n^{-1/2}\vec{j} \mid n^{-1/2}\vec{j}\right\rangle^2 = 1.$$

Thus, $\sum_{i=1}^{k}\cos^2\alpha_i = 1$ if n is even and $\sum_{i=1}^{k+1}\cos^2\alpha_i = 1$ otherwise. In either case, $\sum_{i=1}^{k}\cos^4\alpha_i \leqslant 1$ and the equality holds if and only if exactly one of the values of $\alpha_1,\ldots,\alpha_k$ is equal to zero and the remainder are equal to $\pi/2$. Since the Cauchy–Schwarz inequality implies that

$$\left(\sum_{i=1}^{k}\lambda_i^2\cos^2\alpha_i\right)^2 \leqslant \left(\sum_{i=1}^{k}\lambda_i^4\right)\cdot\left(\sum_{i=1}^{k}\cos^4\alpha_i\right), \tag{2.10}$$

the inequality (2.9) indeed holds.

Now, assume that the inequality (2.9) holds with equality. As we have seen, this can only occur if exactly one of the α_i are zero and the rest are $\pi/2$. By symmetry, we can assume that $\alpha_1 = 0$ and $\alpha_i = \pi/2$ for $i > 1$. It follows that $\lambda_2 = \cdots = \lambda_k = 0$, and $v_1 = n^{-1/2}\vec{j}$. Hence, as $B = ULU^T$, the entry $B_{i,j}$ is equal to $\lambda_1 n^{1/2}(v_{2,j} - v_{2,i})$ for all $i,j \in [n]$. It follows that, for $\sigma_3(A) \in [1/32, 1/8]$, if $\sigma_4(A) = g(\sigma_3(A))$, then $A_{i,j} = 1/2 + z_i - z_j$ where $z_i = \lambda_1 n^{1/2} v_{2,i}/2$. Conversely, any matrix of this form satisfies (2.9) with equality and therefore satisfies $\sigma_4(A) = g(\sigma_3(A))$. □

Reinterpreting Theorem 2.8 in the language of tournament limits, we obtain the following corollary.

Corollary 2.9. *Let W be a tournament limit. If $t(C_3, W) \geqslant 1/32$, then $t(C_4, W) \geqslant g(t(C_3, W))$ and the equality holds if and only if there exists a measurable function $f : [0,1] \to [0,1/2]$ such that $W(x,y) = 1/2 + f(x) - f(y)$ for almost every $(x,y) \in [0,1]^2$.*

Proof. Let $(W_k)_{k\in\mathbb{N}}$ be a sequence of refining weak $1/k$-regular approximations of W and let A_k, $k \in \mathbb{N}$, be the corresponding tournament matrices. Since $t(C_3, W) \geqslant 1/32$, we obtain $\sigma_3(A_k) = t(C_3, W_k) \geqslant 1/32$ by Proposition 2.7. Thus, by Theorem 2.8, we have

$$t(C_4, W_k) = \sigma_4(A_k) \geqslant g\left(\sigma_3(A_k)\right) = g(t(C_3, W_k)).$$

and so $t(C_4, W) \geqslant g(t(C_3, W))$ by the fact that $(W_k)_{k\in\mathbb{N}}$ converges to W in L_1.

To prove the structure of W in the case of equality, assume that $t(C_4, W) = g(t(C_3, W))$. Let n_k be the order of A_k for $k \in \mathbb{N}$, and let $\lambda_{k,1},\ldots,\lambda_{k,\lfloor n_k/2\rfloor}$, $\alpha_{k,1},\ldots,\alpha_{k,\lfloor n_k/2\rfloor}$ and B_k be defined as in the proof of Theorem 2.8 (we may assume that n_k is even and so $\alpha_{k,\lfloor n_k/2\rfloor+1}$ is not defined). The analysis of the case of equality in the proof of Theorem 2.8 implies that

$$\lim_{k\to\infty}\lambda_{k,1} = \sqrt{\frac{1 - 8t(C_3, W)}{3}} \qquad \text{and} \qquad \lim_{k\to\infty}\alpha_{k,1} = 0. \tag{2.11}$$

Let $\{Z_{k,1}, \ldots, Z_{k,n_k}\}$ be the partition of the interval $[0,1]$ corresponding to W_k, and let $w_k = (w_{k,1}, \ldots, w_{k,n_k})$ be the vector v_2 as defined in the proof of Theorem 2.8. Define a function $f_k : [0,1] \to \mathbb{R}$ by setting

$$f_k(x) = \frac{\lambda_{k,1} n_k^{1/2}}{2} \left(w_{k,i} - \frac{\sum_{i'=1}^{n_k} w_{k,i'}}{n_k} \right)$$

where $i \in [n_k]$ such that $x \in Z_{k,i}$. It follows from (2.11) that

$$\lim_{k\to\infty} \int_{x,y\in[0,1]} |W_k(x,y) - (1/2 + f_k(x) - f_k(y))| \, \mathrm{d}x \, \mathrm{d}y = 0\,. \tag{2.12}$$

Also observe that the definition of f_k implies that

$$\int_{x\in[0,1]} f_k(x) = 0\,. \tag{2.13}$$

We next define functions $\tilde{f}_k : [0,1] \to \mathbb{R}$ by setting

$$\tilde{f}_k(x) = \frac{1}{|Z_{k,i}|} \int_{(x',y)\in Z_{k,i}\times[0,1]} W(x',y) \, \mathrm{d}x' \, \mathrm{d}y - \frac{1}{2}$$

for $x \in Z_{k,i}$, $i \in [n_k]$. Note that

$$\tilde{f}_k(x) = \int_{y\in[0,1]} W_k(x,y) \, \mathrm{d}y - \frac{1}{2}\,. \tag{2.14}$$

In particular, $\tilde{f}_k(x) \in [-1/2, 1/2]$ for all $x \in [0,1]$. Observe that the just defined functions satisfy that

$$\int_{x\in Z_{k,i}} \tilde{f}_{k'}(x) \, \mathrm{d}x = \int_{x\in Z_{k,i}} \tilde{f}_k(x) \, \mathrm{d}x$$

for every $k \in \mathbb{N}$, $i \in [n_k]$ and $k' \geqslant k$. In particular, the sequence $(\tilde{f}_k)_{k\in\mathbb{N}}$ forms a martingale when viewed as a sequence of random variables on $[0,1]$. So, Doob's Martingale Convergence Theorem yields that the sequence $(\tilde{f}_k)_{k\in\mathbb{N}}$ L_1-converges to a function $\tilde{f} : [0,1] \to [-1/2, 1/2]$.

We derive by applying the L_1-convergence of $(\tilde{f}_k)_{k\in\mathbb{N}}$, (2.14), (2.13), the triangle inequality and (2.12) (in this order) that

$$\begin{aligned}
&\lim_{k\to\infty}\int_{x\in[0,1]}\left|\tilde f(x)-f_k(x)\right|\,\mathrm{d}x\\
&\quad=\lim_{k\to\infty}\int_{x\in[0,1]}\left|\tilde f_k(x)-f_k(x)\right|\,\mathrm{d}x\\
&\quad=\lim_{k\to\infty}\int_{x\in[0,1]}\left|\int_{y\in[0,1]}W_k(x,y)\,\mathrm{d}y-1/2-f_k(x)\right|\,\mathrm{d}x\\
&\quad=\lim_{k\to\infty}\int_{x\in[0,1]}\left|\int_{y\in[0,1]}W_k(x,y)-(1/2+f_k(x)-f_k(y))\,\mathrm{d}y\right|\,\mathrm{d}x\\
&\quad\leqslant\lim_{k\to\infty}\int_{x,y\in[0,1]}\left|W_k(x,y)-(1/2+f_k(x)-f_k(y))\right|\,\mathrm{d}x\,\mathrm{d}y=0\,.
\end{aligned}$$

This implies that the sequence $(f_k)_{k\in\mathbb{N}}$ also L_1-converges to the function $\tilde f$. It follows from (2.12) and the L_1-convergence of W_k to W that $W(x,y)$ is equal to $1/2+\tilde f(x)-\tilde f(y)$ for almost every $(x,y)\in[0,1]^2$.

It remains to shift $\tilde f$ so that its range lies in $[0,1/2]$. Let z_0 be the infimum of those values z such that the measure of $\tilde f^{-1}((-\infty,z])$ is positive, and define a function $f:[0,1]\to[0,1/2]$ as follows:

$$f(x)=\begin{cases}\tilde f(x)-z_0 & \text{if } \tilde f(x)\in[z_0,z_0+1/2], \text{ and}\\ 0 & \text{otherwise.}\end{cases}$$

Since $W(x,y)=1/2+\tilde f(x)-\tilde f(y)$ for almost every $(x,y)\in[0,1]^2$ and $W(x,y)\in[0,1]$ for all $(x,y)\in[0,1]^2$, the set of $x\in[0,1]$ such that the second case in the definition of $f(x)$ applies has measure zero. It follows that $W(x,y)=1/2+f(x)-f(y)$ for almost every $(x,y)\in[0,1]^2$ as desired. □

2.4 Regime of three parts

Having confirmed Conjecture 1.6 in the regime of two parts, we now turn towards the next case, namely $1/72\leqslant\sigma_3(A)\leqslant 1/32$ (Theorem 2.14). Indeed, the proof of Theorem 2.14 will apply to both regimes, although it does not characterise the extremal tournaments.

We start with analysing the following optimisation problem, which we refer to as the problem Spectrum. This optimisation problem is obtained from constraints that (normalised) eigenvalues of a non-negative matrix of order n with trace $n/2$ must satisfy. We state this formally in Lemma 2.10, which follows the statement of the problem.

	Optimisation problem Spectrum
Parameters:	reals $s_3 \in [0, 1/8]$ and $\rho \in [0, 1/2]$
	non-negative integers k and ℓ such that $k + \ell \geqslant 1$
Variables:	real numbers $r_1, \dots, r_k$, $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$
Constraints:	$0 \leqslant r_1, \dots, r_k \leqslant \rho$
	$0 \leqslant a_1, \dots, a_\ell$
	$\rho + \sum_{i=1}^{k} r_i + 2\sum_{i=1}^{\ell} a_i = 1/2$
	$\rho^3 + \sum_{i=1}^{k} r_i^3 + 2\sum_{i=1}^{\ell} (a_i^3 - 3a_ib_i^2) = s_3$
Objective:	minimise $\rho^4 + \sum_{i=1}^{k} r_i^4 + 2\sum_{i=1}^{\ell} (a_i^4 - 6a_i^2b_i^2 + b_i^4)$

Lemma 2.10. *Let A be a tournament matrix of order n with spectral radius equal to $\rho \cdot n$. Let k be one less than the number of real eigenvalues of A (counting multiplicities) and ℓ the number of conjugate pairs of complex eigenvalues (again counting multiplicities). Further, let $\rho \cdot n, r_1 \cdot n, \dots, r_k \cdot n$ be the $k+1$ real eigenvalues and $(a_1 \pm \imath b_1)n, \dots, (a_\ell \pm \imath b_\ell)n$ be the ℓ pairs of complex eigenvalues. Then the numbers $r_1, \dots, r_k$, $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$ satisfy all constraints in the optimisation problem* Spectrum *for the parameters $s_3 = \sigma_3(A)$, ρ, k and ℓ.*

Proof. Since $\rho \cdot n$ is the spectral radius of A, we have $r_i \leqslant \rho$ for every $i \in [k]$ and $\rho \cdot n$ is an eigenvalue of A by the Perron–Frobenius theorem. Since the real part of every eigenvalue of A is non-negative by Lemma 2.4, all $r_1, \dots, r_k$ and $a_1, \dots, a_\ell$ are non-negative. Since the diagonal entries of A are all $1/2$ and trace of A is equal to the sum of its eigenvalues, we have

$$\rho n + \sum_{i=1}^{k} r_i n + 2\sum_{i=1}^{\ell} a_i n = n/2.$$

Similarly, the trace of A^3 gives us

$$s_3 n^3 = \sigma_3(A) n^3 = \rho^3 n^3 + \sum_{i=1}^{k} r_i^3 n^3 + 2\sum_{i=1}^{\ell} \left(a_i^3 - 3a_ib_i^2\right) n^3.$$

Thus, we conclude that the numbers $r_1, \dots, r_k$, $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$ satisfy all constraints in the optimisation problem Spectrum. □

Note that the objective function of Spectrum is precisely $\sigma_4(A)$. Next, we analyse the structure of optimal solutions of the optimisation problem Spectrum.

Lemma 2.11. *Let $r_1, \dots, r_k$, $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$ be an optimal solution of the optimisation problem* Spectrum *with the parameters s_3, ρ, k and ℓ. Then, at least one of the following two cases must hold:*

- *There exist positive reals r' and r'' such that $r_1, \dots, r_k \in \{0, r', r'', \rho\}$ and $(a_1, b_1), \dots, (a_\ell, b_\ell) \in \{(0,0), (r',0), (r'',0)\}$.*
- *There exist reals a' and $b' \neq 0$ such that $r_1, \dots, r_k \in \{0, \rho\}$ and $(a_1, b_1), \dots, (a_\ell, b_\ell) \in \{(0,0), (a',b'), (a',-b')\}$.*

Proof. The method of Lagrange multipliers implies that the gradient of the objective function is a linear combination of the gradient of the two equality constraints when restricted to the entries indexed by $r_i \notin \{0, \rho\}$, by $a_i \neq 0$ and $b_i \neq 0$, i.e., when we are not on the boundary of the feasible set. In particular, the rank of the matrix M with rows being the described restrictions of the three gradient vectors is at most two.

We first analyse the case that one of the numbers $b_1, \dots, b_\ell$ is non-zero. Our aim is to show that the second case described in the statement of the lemma applies. By symmetry, we can assume that $b_1 \neq 0$. Also note the following holds for every $i \in [\ell]$: if $b_i \neq 0$, then $a_i \neq 0$. Indeed, if $a_i = 0$ and $b_i \neq 0$, then setting $b_i = 0$ does not affect the constraints and decreases the objective function, which contradicts that the solution is optimal. It follows that a_1 is positive.

Suppose that there exist r_i such that $0 < r_i < \rho$. The restriction of the matrix M to the columns corresponding to a_1, b_1 and r_i is the following.

$$\begin{bmatrix} 2 & 0 & 1 \\ 6a_1^2 - 6b_1^2 & -12a_1b_1 & 3r_i^2 \\ 8a_1^3 - 24a_1b_1^2 & 8b_1^3 - 24a_1^2b_1 & 4r_i^3 \end{bmatrix} \tag{2.15}$$

Dividing the first column by 2 and the second by $4b_1$, dividing the second row by 3 and the third by 2, and subtracting the last column from the first one yields the following matrix, which has the same rank.

$$\begin{bmatrix} 0 & 0 & 1 \\ a_1^2 - b_1^2 - r_i^2 & -a_1 & r_i^2 \\ 2(a_1^3 - 3a_1b_1^2 - r_i^3) & b_1^2 - 3a_1^2 & 2r_i^3 \end{bmatrix}$$

This matrix is not full rank if and only if the determinant of its submatrix formed by the intersection of the second and third rows with the first and second columns, which is equal to

$$-(a_1^2 + b_1^2)^2 + (3a_1^2 - b_1^2)r_i^2 - 2a_1r_i^3\,,$$

is zero. However, this determinant can be rewritten as

$$-(2a_1r_i + a_1^2 + b_1^2)\left((r_i - a_1)^2 + b_1^2\right) ,$$

which is negative since $a_1 > 0$ and $b_1 \neq 0$. It follows that $r_i \in \{0, \rho\}$ for all $i \in [k]$.

Further suppose that there exists an index $i \in [\ell]$ for which $(a_i, b_i) \notin \{(0,0), (a_1, b_1), (a_1, -b_1)\}$. If $b_i = 0$, then the restriction of the matrix M to the columns corresponding to a_1, b_1 and a_i is the same as the restriction of the matrix M considered in the previous paragraph with r_i replaced by a_i and the corresponding column multiplied by two. In particular, the restriction cannot have rank two in this case. Hence, we can assume that $b_i \neq 0$ and so $a_i > 0$ (the argument is the same as when we argued that $a_1 > 0$). The restriction of the matrix M to the columns corresponding to a_1, b_1, a_i and b_i is the following.

$$\begin{bmatrix} 2 & 0 & 2 & 0 \\ 6a_1^2 - 6b_1^2 & -12a_1b_1 & 6a_i^2 - 6b_i^2 & -12a_ib_i \\ 8a_1^3 - 24a_1b_1^2 & 8b_1^3 - 24a_1^2b_1 & 8a_i^3 - 24a_ib_i^2 & 8b_i^3 - 24a_i^2b_i \end{bmatrix}$$

The rank of this matrix is the same as the rank of the following matrix (the rows are multiplied by $1/2$, $1/6$ and $1/4$, the columns by 1, $-a_1/2b_1$, 1 and $-a_i/2b_i$, respectively).

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ a_1^2 - b_1^2 & a_1^2 & a_i^2 - b_i^2 & a_i^2 \\ 2a_1^3 - 6a_1b_1^2 & 3a_1^3 - a_1b_1^2 & 2a_i^3 - 6a_ib_i^2 & 3a_i^3 - a_ib_i^2 \end{bmatrix}$$

By subtracting twice the second column from the first column and twice the fourth column from the third column, we obtain the following matrix.

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ -(a_1^2 + b_1^2) & a_1^2 & -(a_i^2 + b_i^2) & a_i^2 \\ -4a_1(a_1^2 + b_1^2) & 3a_1^3 - a_1b_1^2 & -4a_i(a_i^2 + b_i^2) & 3a_i^3 - a_ib_i^2 \end{bmatrix}$$

Since the last row of the matrix is a linear combination of the previous two rows (the operation that we have performed so far has preserved this property of the matrix M), it follows that

$$\frac{3a_1^2 - b_1^2}{a_1} = \frac{3a_i^2 - b_i^2}{a_i} \tag{2.16}$$

We now subtract the second row multiplied by the value given in (2.16) from the third row and obtain the following matrix, which has the rank two.

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ -(a_1^2+b_1^2) & a_1^2 & -(a_i^2+b_i^2) & a_i^2 \\ -(a_1^2+b_1^2)^2/a_1 & 0 & -(a_i^2+b_i^2)^2/a_i & 0 \end{bmatrix}$$

It follows that

$$\frac{(a_1^2+b_1^2)^2}{a_1} = \frac{(a_i^2+b_i^2)^2}{a_i} ,$$

which yields that

$$b_i^2 = -a_i^2 + \sqrt{\frac{a_i}{a_1}}(a_1^2+b_1^2) . \tag{2.17}$$

On the other hand, we derive from (2.16) that

$$b_i^2 = 3a_i^2 - \frac{a_i}{a_1}\left(3a_1^2 - b_1^2\right) . \tag{2.18}$$

We obtain by comparing (2.17) and (2.18) the following.

$$\begin{aligned} 0 &= 4a_i^2 - \frac{a_i}{a_1}\left(3a_1^2 - b_1^2\right) - \sqrt{\frac{a_i}{a_1}}(a_1^2+b_1^2) \\ &= \sqrt{a_i}(\sqrt{a_i}-\sqrt{a_1})\left(4a_i + 4\sqrt{a_1 a_i} + \frac{a_1^2+b_1^2}{a_1}\right) \end{aligned}$$

Since both a_1 and a_i are positive, this expression can be equal to zero only if $a_i = a_1$. Consequently, the equality (2.17) implies that $b_i = b_1$ or $b_i = -b_1$, which contradicts the choice of (a_i, b_i). Hence, we have established that if at least one of $b_1, \ldots, b_\ell$ is non-zero, then the second case indeed applies.

We now consider the case that $b_1 = \cdots = b_\ell = 0$. Suppose that the first case in the statement of the lemma does not apply. This implies that there exist three distinct positive reals α, β and γ such that at least one of the values $r_1, \ldots, r_k, a_1, \ldots, a_\ell$ is α, at least one is β and at least one is γ. Consequently, the matrix M contains the following submatrix possibly after dividing some columns by two (the columns correspond to those of the variables $r_1, \ldots, r_k, a_1, \ldots, a_\ell$ that are equal to α, β and γ, respectively; the columns corresponding to the variables $a_1, \ldots, a_\ell$ are divided by two).

$$\begin{bmatrix} 1 & 1 & 1 \\ 3\alpha^2 & 3\beta^2 & 3\gamma^2 \\ 4\alpha^3 & 4\beta^3 & 4\gamma^3 \end{bmatrix}$$

The determinant of this matrix is equal to

$$12(\alpha^2\beta^3 + \beta^2\gamma^3 + \gamma^2\alpha^3 - \alpha^2\gamma^3 - \beta^2\alpha^3 - \gamma^2\beta^3)\,,$$

which is equal to

$$12(\alpha - \beta)(\beta - \gamma)(\gamma - \alpha)(\alpha\beta + \alpha\gamma + \beta\gamma)\,.$$

Since this expression is non-zero for all distinct positive reals α, β and γ, we conclude that the rank of M is three, which contradicts our assumption that the first case as described in the statement of the lemma does not apply. □

Before we can prove the main result of this section, we need two additional auxiliary lemmas.

Lemma 2.12. *Let $z \in (0, 1]$, and let $x_1, \dots, x_n$ be non-negative reals such that $x_1 + \dots + x_n = 1/2$ and $x_i \leqslant z$ for every $i \in [n]$. Then*

$$\sum_{i=1}^{n} x_i^3 \leqslant \lfloor z^{-1}/2 \rfloor \cdot z^3 + (1/2 - \lfloor z^{-1}/2 \rfloor \cdot z)^3,$$

and equality holds if and only if all but at most one of $x_1, \dots, x_n$ are equal to 0 or z.

Proof. Consider any n-tuple $x_1, \dots, x_n$ that maximises the sum $x_1^3 + \dots + x_n^3$ among all n-tuples of non-negative reals $x_1, \dots, x_n$ such that $x_1 + \dots + x_n = 1/2$ and $x_i \leqslant z$, $i \in [n]$. If $x_i \in \{0, z\}$ for all but at most one $i \in [n]$, then the sum of $x_1^3 + \dots + x_n^3$ is equal to $\sum_{i=1}^{n} x_i^3 \leqslant \lfloor z^{-1}/2 \rfloor \cdot z^3 + (1/2 - \lfloor z^{-1}/2 \rfloor \cdot z)^3$ and the lemma holds. Otherwise, there exist x_i and x_j such that $0 < x_i \leqslant x_j < z$. Choose $\varepsilon > 0$ such that $\varepsilon < x_i$ and $\varepsilon < z - x_j$, and replace x_i with $x_i - \varepsilon$ and x_j with $x_j + \varepsilon$. This preserves the sum $x_1 + \dots + x_n$ and increases the sum $x_1^3 + \dots + x_n^3$, which contradicts the choice of the n-tuple $x_1, \dots, x_n$. □

Linial and Morgenstern [72] proved that, among the random blow-ups of transitive tournaments with the fixed density of C_3, the density of C_4 is minimised if all parts have the same size except possibly for a single smaller part. This statement is equivalent to the following.

Lemma 2.13 (Linial and Morgenstern [72, Lemma 2.7])**.** *Let $x_1, \dots, x_n$ be any non-negative reals such that their sum is $1/2$. Then*

$$x_1^4 + \dots + x_n^4 \geqslant g(x_1^3 + \dots + x_n^3)\,.$$

We are now ready to prove the main result of this section.

Theorem 2.14. *Let A be a tournament matrix of order n. If $\sigma_3(A) \geqslant 1/72$, then $\sigma_4(A) \geqslant g(\sigma_3(A))$.*

Proof. Let $s_3 = \sigma_3(A)$. We start with lower bounding the spectral radius of A. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A. By Lemma 2.4, the real parts of all the eigenvalues are non-negative, which implies that

$$s_3 = \sum_{i=1}^{n} \left(\frac{\lambda_i}{n}\right)^3 \leqslant \sum_{i=1}^{n} \left(\frac{\operatorname{Re}\lambda_i}{n}\right)^3 .$$

By Lemma 2.12, the last sum is at most

$$\lfloor \rho_A^{-1}/2 \rfloor \cdot \rho_A^3 + (1/2 - \lfloor \rho_A^{-1}/2 \rfloor \cdot \rho_A)^3$$

where ρ_A is the spectral radius of A divided by n. Consequently, ρ_A is at least z where z is the unique real between 0 and 1/2 satisfying that

$$s_3 = \lfloor z^{-1}/2 \rfloor \cdot z^3 + (1/2 - \lfloor z^{-1}/2 \rfloor \cdot z)^3 .$$

Note that $z \geqslant 1/6$ since $s_3 \geqslant 1/72$.

Lemma 2.10 now yields that the theorem will be proven if we show that the optimal solution of the problem Spectrum is at least $g(s_3)$ for s_3, any $\rho \geqslant z$ and all non-negative integers k and ℓ. By Lemma 2.11, this would be implied by the following two claims, which correspond to the two cases described in the statement of Lemma 2.11.

Claim 1. If $r_1, \dots, r_k$ are any positive real numbers that have at most three distinct values and that satisfy $r_1 + \cdots + r_k = 1/2$ and $r_1^3 + \cdots + r_k^3 = s_3$, then $r_1^4 + \cdots + r_k^4 \geqslant g(s_3)$.

Claim 2. If m and m' are positive integers, $\rho \geqslant z$, a is a non-negative real and b is a real such that $m\rho + 2m'a = 1/2$ and $m\rho^3 + 2m'(a^3 - 3ab^2) = s_3$, then $m\rho^4 + 2m'(a^4 - 6a^2b^2 + b^4) \geqslant g(s_3)$.

Claim 1 follows from Lemma 2.13 (even without the restriction on the number of the distinct values that $r_1, \dots, r_k$ may have). So, we focus on proving Claim 2 in the remainder of the proof. Note that $m \in \{1, 2\}$ in this case since $\rho \geqslant z \geqslant 1/6$.

To prove Claim 2, we fix m and m' and consider the following optimisation problem: minimise the sum $m\rho^4 + 2m'(a^4 - 6a^2b^2 + b^4)$ subject to $m\rho + 2m'a = 1/2$,

$m\rho^3 + 2m'(a^3 - 3ab^2) = s_3$, $\rho \geqslant z$ and $a \geqslant 0$. The method of Lagrange multipliers implies that the following matrix is not full rank

$$\begin{bmatrix} m & 2m' & 0 \\ 3m\rho^2 & 6m'a^2 - 6m'b^2 & -12m'ab \\ 4m\rho^3 & 8m'a^3 - 24m'ab^2 & 8m'b^3 - 24m'a^2b \end{bmatrix} \tag{2.19}$$

for the values of ρ, a and b that minimise the sum unless $\rho = z$ or $a = 0$. However, dividing the first column by m and the remaining two by m', permuting the columns and renaming the variables yields the same matrix as in (2.15), which we have analysed in the proof of Lemma 2.11. In particular, the matrix in (2.19) has full rank unless $a = 0$ or (possibly) $b = 0$. We conclude that the expression $m\rho^4 + 2(a^4 - 6a^2b^2 + b^4)$ is minimised when at least one of the following holds: $\rho = z$, $a = 0$ or $b = 0$. We next analyse these three cases.

The case $\rho = z$. In this case, Lemma 2.12 implies that $m\rho^3 + 2m'a^3 < s_3$ unless $a \geqslant \rho$, i.e., there is no such feasible solution unless $a \geqslant \rho$. If indeed $a \geqslant \rho$, then since $m\rho + 2m'\rho = 1/2$ and $\rho \geqslant 1/6$, it must be that $m = m' = 1$, $\rho = a = 1/6$, $s_3 = 1/72$ and $b = 0$, in which case $m\rho^4 + 2m'(a^4 - 6a^2b^2 + b^4) = 1/432 = g(s_3)$.

The case $a = 0$. If $a = 0$, then, as noted in the proof of Lemma 2.11, setting $b = 0$ does not affect the constraints, but does decrease the objective function. Hence this case reduces to the final case $b = 0$.

The case $b = 0$. If $b = 0$, then using Lemma 2.13 we obtain that $m\rho^4 + 2m'a^4 \geqslant g(m\rho^3 + 2m'a^3) = g(s_3)$.

Hence, we have shown that $m\rho^4 + 2m'(a^4 - 6a^2b^2 + b^4) \geqslant g(s_3)$ for all ρ, a and b such that $m\rho + 2m'a = 1/2$, $m\rho^3 + 2m'(a^3 - 3ab^2) = s_3$, $\rho \geqslant z$ and $a \geqslant 0$. The proof of Claim 2 is now finished and so is the proof of the theorem. □

Proposition 2.5 yields the following corollary in the tournament limit setting.

Corollary 2.15. *Let W be a tournament limit. If $t(C_3, W) \geqslant 1/72$, then $t(C_4, W) \geqslant g(t(C_3, W))$.*

2.5 Concluding remarks

Unfortunately, the proof of Theorem 2.14 does not immediately work in higher regimes because, in the case of four or more parts, the solution to the optimisation problem Spectrum beats the conjectured minimum. Of course, as solutions need not be realisable as the eigenvalues of a tournament matrix, this does not invalidate Conjecture 1.6. It is possible that the current method can be pushed further by introducing to the optimisation problem additional constraints that reflect the properties that eigenvalues of tournament matrices must satisfy.

Meanwhile, in the regime of two parts, we have been able to fully determine the asymptotic structure of extremal examples. These constructions can be extended to the remaining regimes as follows. Fix $k \in \mathbb{N}$, $z \in [1/(k+1), 1/k]$ and $i, i' \in [k+1]$ such that $|i - i'| = 1$. We construct a tournament T with n vertices as follows. The vertices of T are split into $k+1$ parts $V_1, \ldots, V_{k+1}$ such that $|V_i| = n - k\lfloor zn \rfloor$ and $|V_j| = \lfloor zn \rfloor$ if $j \neq i$. If two vertices v and v' respectively belong to distinct parts V_j and $V_{j'}$ with $j < j'$ and $\{j, j'\} \neq \{i, i'\}$, then the tournament T contains an arc from v to v'. If, instead, v and v' belong to the *same* part V_j, where $j \notin \{i, i'\}$, then the edge between v and v' is oriented from v to v' with probability $1/2$, i.e., the vertices of every such part induce a randomly oriented tournament. Finally, each vertex $v \in V_i \cup V_{i'}$ is assigned a real number $p_v \in [0, 1/2]$ and the edge between v and $v' \in V_i \cup V_{i'}$ is directed from v to v' with probability $1/2 + p_v - p_{v'}$. If the expected number of triangles in T is equal to $\frac{1}{8}\left(k\lfloor zn \rfloor^3 + (n - k\lfloor zn \rfloor)^3\right)$, then the expected value of the density $t(C_4, T)$ is

$$g\left(\frac{1}{8}\left(k\lfloor zn \rfloor^3 + (n - k\lfloor zn \rfloor)^3\right)\right) + o(1)$$

and both of these random variables are concentrated. In particular, unless z^{-1} is a positive integer, there are infinitely many different types of extremal tournaments.

It is also interesting to note that the problem of determining the set of feasible densities of cycles of length three and four is equivalent to the analogous problems for transitive tournaments of order three and four and for the cycle and transitive tournament of order four [72, Proposition 1.1]. To see this, let T_k be the transitive tournament of order k and let $t(T_k, T)$ be the probability that a random mapping from $V(T_k)$ to $V(T)$ is a homomorphism. The following holds for every n-vertex tournament T:

$$8t(C_3, T) + 24t(T_4, T) - 6t(C_4, T) = 1 - O(n^{-1}).$$

Thus, the problem of minimizing the density of C_4 when the density of C_3 is fixed is equivalent to minimizing the density of T_4 when the density of T_3 is fixed, and also equivalent to minimizing the density of C_4 when T_4 is fixed, in the sense that a complete solution to any of these three problems yield complete solutions to the remaining two.

Chapter 3

Quasirandom Permutations

3.1 Overview

In Section 1.3, we defined Σ-forcing sets of permutations as the sets S for which

$$\lim_{i\to\infty} \sum_{\pi\in S} d(\pi, \Pi_i) = \frac{|S|}{k!} \text{ if and only if } \{\Pi_i\}_{i\in\mathbb{N}} \text{ is quasirandom,}$$

whenever $\{\Pi_i\}_{i\in\mathbb{N}}$ is a sequence of permutations. In this chapter, we characterise all Σ-forcing sets of 4-permutations:

Theorem 3.1. *Let S be a set of 4-permutations. The set S is Σ-forcing if and only if S is one of the following five sets*

- $\{1234, 1243, 2134, 2143, 3412, 3421, 4312, 4321\}$,
- $\{1234, 1432, 2143, 2341, 3214, 3412, 4123, 4321\}$,
- $\{1324, 1342, 2413, 2431, 3124, 3142, 4213, 4231\}$,
- $\{1324, 1423, 2314, 2413, 3142, 3241, 4132, 4231\}$,
- $\{1234, 1243, 1432, 2134, 2143, 2341, 3214, 3412, 3421, 4123, 4312, 4321\}$,

or their complements.

The five listed sets are shown to be Σ-forcing in Theorems 3.5, 3.6, 3.7, and 3.8 (one of the sets follows by symmetry). The proofs of all of these results rely on flag algebras. Theorem 3.16 completes the proof of Theorem 3.1 by showing that no other set of 4-permutations is Σ-forcing. Some of the arguments are supported by supplementary data that is available online as a series of five appendices totalling 44 pages. The appendices can be downloaded as an ancillary file on arXiv at `https://arxiv.org/src/1909.11027/anc/Appendices.pdf`.

3.2 Preliminaries

In this section, we fix the notation used throughout the chapter. Let τ_1 and τ_2 be the permutations 12 and 21, respectively. The set of all k-permutations is denoted by S_k. Two permutations π and σ of the same order, say k, are *symmetric* if the permutation matrix of π can be obtained from the permutation matrix of σ by a sequence of reflections and rotations when viewed as $k \times k$ tables, i.e., if either

- $\pi(i) = \sigma(i)$, or
- $\pi(i) = \sigma^{-1}(i)$, or
- $\pi(i) = \sigma(k+1-i)$, or
- $\pi(i) = \sigma^{-1}(k+1-i)$, or
- $\pi(i) = k+1-\sigma(i)$, or
- $\pi(i) = k+1-\sigma^{-1}(i)$, or
- $\pi(i) = k+1-\sigma(k+1-i)$, or
- $\pi(i) = k+1-\sigma^{-1}(k+1-i)$

holds for all $i \in [k]$. For example, exactly the following seven permutations are symmetric to 12534 in addition to the permutation 12534 itself: 12453, 23145, 31245, 35421, 43521, 54132 and 54213.

A *permuton* is a Borel probability measure μ on $[0,1]^2$ that has uniform marginals, i.e., $\mu([x,x'] \times [0,1]) = x'-x$ for every $0 \leqslant x < x' \leqslant 1$ and $\mu([0,1] \times [y,y']) = y'-y$ for every $0 \leqslant y < y' \leqslant 1$. In other contexts, permutons are known as doubly stochastic measures or two-dimensional copulas. Given a permuton μ, a *μ-random permutation of order k* is obtained in the way that we now describe. We first sample k points $(x_1,y_1),\ldots,(x_k,y_k)$ in $[0,1]^2$ according to the probability measure μ. Note that the probability that an x- or y-coordinate is shared by multiple points is zero because μ has uniform marginals. By renaming the points, we can assume that $x_1 < \cdots < x_k$. The μ-random permutation $\pi \in S_k$ is then the unique permutation such that $\pi(i) < \pi(j)$ if and only if $y_i < y_j$ for every $i,j \in [k]$. We define the *pattern density* of $\pi \in S_k$ in the permuton μ to be the probability that a μ-random permutation of order k is π. A sequence $(\Pi_i)_{i\in\mathbb{N}}$ of permutations is *convergent* if $|\Pi_i|$ grows to infinity and the limit

$$\lim_{i\to\infty} d(\pi,\Pi_i)$$

exists for every permutation π. It can be shown [66,71] that if $(\Pi_i)_{i\in\mathbb{N}}$ is a convergent sequence of permutations, then there exists a unique permuton μ such that

$$\lim_{i\to\infty} d(\pi,\Pi_i) = d(\pi,\mu)$$

for every permutation π; the permuton μ is called the *limit* of the sequence $(\Pi_i)_{i\in\mathbb{N}}$. In the other direction, if μ is a permuton, then, with probability 1, a sequence of μ-random permutations with increasing orders converges and its limit is μ. We note that a sequence $(\Pi_i)_{i\in\mathbb{N}}$ of permutations is quasirandom if and only if its limit is the uniform measure on $[0,1]^2$.

The *support* of a Borel measure μ, denoted $\operatorname{supp}(\mu)$, is the set of all points x such that every open neighbourhood of x has positive measure under μ. Fix a permutation $\tau \in S_\ell$. A *τ-rooted permuton* is an $(\ell+1)$-tuple $\mu^\tau = (\mu, (x_1, y_1), \dots, (x_\ell, y_\ell))$ such that

- μ is a permuton,
- $(x_1, y_1), \dots, (x_\ell, y_\ell) \in \operatorname{supp}(\mu)$, $x_1 < \dots < x_\ell$, and
- $\tau(i) < \tau(j)$ if and only if $y_i < y_j$ for all $i, j \in [\ell]$.

The points $(x_1, y_1), \dots, (x_\ell, y_\ell)$ are referred to as *roots*. If μ^τ is a τ-rooted permuton, then a *μ^τ-random permutation of order $k \geqslant \ell$* is a τ-rooted permutation obtained by sampling $k-\ell$ points in $[0,1]^2$ according to the measure μ, forming a permutation of order k using the ℓ roots and the $k-\ell$ sampled points, and distinguishing the ℓ points corresponding to the roots of μ^τ to be the roots of the permutation. If π^τ is a τ-rooted permutation, we write $d(\pi^\tau, \mu^\tau)$ for the probability that a μ^τ-random permutation of order $|\pi^\tau|$ is π^τ.

Fix a permuton μ for the rest of this section. We define a mapping $h_\mu : \mathcal{A} \to \mathbb{R}$ by setting $h_\mu(\pi)$ to be $d(\pi, \mu)$ for every permutation π and extending linearly. Clearly, h_μ is a homomorphism from $\mathcal{A}$ to $\mathbb{R}$ that respects addition and multiplication by a real number. The mapping h_μ also respects the multiplication operation on $\mathcal{A}$, i.e., $h_\mu(A \times B) = h_\mu(A)h_\mu(B)$ for all $A, B \in \mathcal{A}$. Following the shorthand introduced in Section 1.3, we write $A \geqslant \alpha$ for an element $A \in \mathcal{A}$ and a real $\alpha \in \mathbb{R}$ if $h_\mu(A) \geqslant \alpha$ for every permuton μ. Analogously to the unrooted case, for a τ-rooted permuton μ^τ, we can define a homomorphism $h_{\mu^\tau} : \mathcal{A}^\tau \to \mathbb{R}$.

Next, for a permutation τ with $d(\tau, \mu) > 0$, we wish to define a probability distribution on τ-rooted permutons arising from μ. Formally, we define μ^τ to be a τ-rooted permuton obtained from μ by choosing $|\tau|$ points randomly according to the probability measure μ to be the roots (and sorting them according to their first coordinates) conditioned on the event that the chosen roots yield the permutation τ, i.e., μ^τ is a random τ-rooted permuton where the randomness comes from the choice of $|\tau|$ roots. The probability distribution on τ-rooted permutons in turn defines a probability distribution on homomorphisms from $\mathcal{A}^\tau$ to $\mathbb{R}$, and we will write h_μ^τ for a random homomorphism from $\mathcal{A}^\tau$ to $\mathbb{R}$ sampled according to this distribution. As in Section 1.3,

$$h_\mu\left(\llbracket A \rrbracket_\tau\right) = d(\tau, \mu) \cdot \mathbb{E}h^\tau_\mu(A)$$

for every $A \in \mathcal{A}^\tau$, and if M is a $k \times k$ positive semidefinite matrix, then the following holds for every vector $w \in (\mathcal{A}^\tau)^k$:

$$h_\mu\left(\llbracket w^T M w \rrbracket_\tau\right) \geqslant 0.$$

3.3 Σ-forcing sets

In this section, we present a sequence of flag algebra calculations to prove that the sets listed in Theorem 3.1 are Σ-forcing. To do so, we first need the following lemma.

Lemma 3.2. *Let μ be a permuton. If*

$$\mu\left([\min\{x_1, x_2\}, \max\{x_1, x_2\}] \times [\min\{y_1, y_2\}, \max\{y_1, y_2\}]\right) = |x_2 - x_1| \cdot |y_2 - y_1|$$

for all points $(x_1, y_1), (x_2, y_2) \in supp\,(\mu)$, then μ is the uniform measure.

Proof. Our goal is to show that $\operatorname{supp}(\mu) = [0,1]^2$. We start with showing that all points on the boundary of $[0,1]^2$ are contained in $\operatorname{supp}(\mu)$. Suppose that $\operatorname{supp}(\mu)$ does not contain the whole boundary of $[0,1]^2$. Since $\operatorname{supp}(\mu)$ is closed, it is enough to consider the points distinct from the four corners. By symmetry, we need to consider the following two cases.

- **There exists $x \in (0,1)$ such that $(x, 0) \notin \operatorname{supp}(\mu)$ but $(x, 1) \in \operatorname{supp}(\mu)$.** By the definition of the support of a measure, there exists $\varepsilon \in (0, \min\{x, 1-x\})$ such that

  $$\mu\left([x - \varepsilon, x + \varepsilon] \times [0, \varepsilon]\right) = 0.$$

  Let $y' \in [0,1]$ be the infimum among all reals such that $(x', y') \in \operatorname{supp}(\mu)$ for some $x' \in (x - \varepsilon, x)$. If there was no such y', then the measure of the rectangle $[x - \varepsilon, x] \times [0, 1]$ would be zero, which is impossible because the measure μ has uniform marginals. Observe that $y' \in [\varepsilon, 1]$. Since $\operatorname{supp}(\mu)$ is a closed set, there exists $x' \in [x - \varepsilon, x]$ such that $(x', y') \in \operatorname{supp}(\mu)$; if possible, choose x' distinct from x.

  We first consider the case that $x' < x$. The assumption of the lemma implies that the measure of the rectangle $[x', x] \times [y', 1]$ is $(x - x')(1 - y')$. On the other hand, the choice of y' implies that the measure of the rectangle $[x', x] \times [0, y']$

is zero. Consequently, the measure of the rectangle $[x', x] \times [0, 1]$ is $(x - x')(1 - y') < x - x'$, which is impossible.

It remains to analyse the case $x' = x$. The choice of y' implies that there exist $y'' \in (y', 1]$ and $x'' \in (x - \varepsilon, x)$ such that $(x'', y'') \in \operatorname{supp}(\mu)$. Since the measure of the rectangle $[x'', x] \times [y'', 1]$ is $(x - x'')(1 - y'')$ and the measure of the rectangle $[x'', x] \times [y', y'']$ is $(x - x'')(y'' - y')$, the measure of the rectangle $[x'', x] \times [y', 1]$ is $(x - x'')(1 - y')$. On the other hand, the choice of y' implies that the measure of the rectangle $[x'', x] \times [0, y']$ is zero, which yields that the measure of the rectangle $[x'', x] \times [0, 1]$ is less than $x - x''$, which is impossible.

- **There exists $x \in (0, 1)$ such that $(x, 0) \notin$ supp (μ) and $(x, 1) \notin$ supp (μ).** By the definition of the support of a measure, there exists $\varepsilon \in (0, \min\{x, 1 - x\})$ such that

$$\mu([x - \varepsilon, x + \varepsilon] \times [0, \varepsilon]) = 0 \text{ and } \mu([x - \varepsilon, x + \varepsilon] \times [1 - \varepsilon, 1]) = 0.$$

Let $y_1 \in [0, 1]$ be the infimum among all reals such that $(x_1, y_1) \in \operatorname{supp}(\mu)$ for some $x_1 \in (x - \varepsilon, x + \varepsilon)$. If there was no such y_1, then the measure of the rectangle $[x - \varepsilon, x + \varepsilon] \times [0, 1]$ would be zero, which is impossible because the measure μ has uniform marginals. Since $\operatorname{supp}(\mu)$ is a closed set, there exists $x_1 \in [x - \varepsilon, x + \varepsilon]$ such that $(x_1, y_1) \in \operatorname{supp}(\mu)$. Note that $y_1 \in [\varepsilon, 1 - \varepsilon]$. Similarly, let $y_2 \in [0, 1]$ be the supremum among all reals such that $(x_2, y_2) \in \operatorname{supp}(\mu)$ for some $x_2 \in (x - \varepsilon, x + \varepsilon)$ (again note that $y_2 \in [\varepsilon, 1 - \varepsilon]$) and we fix $x_2 \in [x - \varepsilon, x + \varepsilon]$ such that $(x_2, y_2) \in \operatorname{supp}(\mu)$. If possible, we choose x_1 and x_2 above such that $x_1 \neq x_2$.

We first consider the case that $x_1 \neq x_2$; by symmetry, we can assume that $x_1 < x_2$. The assumption of the lemma implies that the measure of the rectangle $[x_1, x_2] \times [y_1, y_2]$ is $(x_2 - x_1)(y_2 - y_1)$, and the choices of y_1 and y_2 imply that the measure of each of the rectangles $[x_1, x_2] \times [0, y_1]$ and $[x_1, x_2] \times [y_2, 1]$ is zero. It follows that the measure of the rectangle $[x_1, x_2] \times [0, 1]$ is $(x_2 - x_1)(y_2 - y_1) < x_2 - x_1$, which is impossible.

It remains to consider the case that $x_1 = x_2$. Since the measure of the rectangle $[x - \varepsilon, x + \varepsilon] \times [0, 1]$ is not zero, there exists $x_3 \in [x - \varepsilon, x + \varepsilon]$, $x_3 \neq x_1$, and $y_3 \in (y_1, y_2)$ such that $(x_3, y_3) \in \operatorname{supp}(\mu)$. By symmetry, we can assume that $x_1 < x_3$. The measures of the rectangles $[x_1, x_3] \times [y_1, y_3]$ and $[x_1, x_3] \times [y_3, y_2]$ are $(x_3 - x_1)(y_3 - y_1)$ and $(x_3 - x_1)(y_2 - y_3)$, respectively. Since the measure of each of the rectangles $[x_1, x_3] \times [0, y_1]$ and $[x_1, x_3] \times [y_2, 1]$ is zero, we conclude that the measure of the rectangle $[x_1, x_3] \times [0, 1]$ is $(x_3 - x_1)(y_2 - y_1) < x_3 - x_1$, which is impossible.

We have shown that all points on the boundary of $[0, 1]^2$ are contained in $\operatorname{supp}(\mu)$. Suppose that there exists a point $(x, y) \in (0, 1)^2$ that is not contained in $\operatorname{supp}(\mu)$, and let $\varepsilon \in (0, \min\{x, y, 1 - x, 1 - y\})$ be such that the whole set $[x - \varepsilon, x + \varepsilon] \times [y - \varepsilon, y + \varepsilon]$ is not contained in $\operatorname{supp}(\mu)$. Let y_1 be the supremum among all reals in $[0, y - \varepsilon]$ such that $(x_1, y_1) \in \operatorname{supp}(\mu)$ for some $x_1 \in (x - \varepsilon, x + \varepsilon)$, and let y_2 be the infimum among all reals in $[y + \varepsilon, 1]$ such that $(x_2, y_2) \in \operatorname{supp}(\mu)$ for some $x_2 \in (x - \varepsilon, x + \varepsilon)$. Further, let $x_1, x_2 \in [x - \varepsilon, x + \varepsilon]$ be such that $(x_1, y_1) \in \operatorname{supp}(\mu)$ and $(x_2, y_2) \in \operatorname{supp}(\mu)$. Note that y_1 can be 0 and y_2 can be 1, and $y_2 - y_1 \geqslant 2\varepsilon$.

We first consider the case that $x_1 \neq x_2$. By symmetry, we can assume that $x_1 < x_2$. Since the boundary of the square $[0, 1]^2$ is contained in $\operatorname{supp}(\mu)$, the measures of the rectangles $[x_1, x_2] \times [0, y_1]$ and $[x_1, x_2] \times [y_2, 1]$ are $(x_2 - x_1)y_1$ and $(x_2 - x_1)(1 - y_2)$, respectively. On the other hand, the choice of y_1 and y_2 implies that the measure of the rectangle $[x_1, x_2] \times [y_1, y_2]$ is zero. Consequently, the measure of the rectangle $[x_1, x_2] \times [0, 1]$ is $(x_2 - x_1)(1 - y_2 + y_1) < x_2 - x_1$, which is impossible.

To conclude the proof, we need to analyse the case $x_1 = x_2$. Let x_3 be any point in the interval $[x - \varepsilon, x + \varepsilon]$ distinct from $x_1 = x_2$. By symmetry, we can assume that $x_1 < x_3$. Again, since the boundary of the square $[0, 1]^2$ is contained in $\operatorname{supp}(\mu)$, it follows that the measures of the rectangles $[x_1, x_3] \times [0, y_1]$ and $[x_1, x_3] \times [y_2, 1]$ are $(x_3 - x_1)y_1$ and $(x_3 - x_1)(1 - y_2)$, respectively, and the choice of y_1 and y_2 yields that the measure of the rectangle $[x_1, x_3] \times [y_1, y_2]$ is zero. We obtain that the measure of the rectangle $[x_1, x_3] \times [0, 1]$ is $(x_3 - x_1)(1 - y_2 + y_1) < x_3 - x_1$, which is impossible. We can now conclude that the support of the measure μ is the whole square $[0, 1]^2$. Consequently the measure of each set $[x, x'] \times [y, y']$ is equal to $(x' - x)(y' - y)$, which implies that the measure μ is the uniform measure on $[0, 1]^2$. This finishes the proof of the lemma. □

For the rest of the section, we fix the following elements $A_1 \in \mathcal{A}^{\tau_1}$ and $A_2 \in \mathcal{A}^{\tau_2}$.

$$A_1 = (\underline{1}2\underline{3}4 - \underline{1}4\underline{3}2) + (1\underline{2}3\underline{4} - 3\underline{2}1\underline{4}) + (\underline{2}3\underline{4}1 - \underline{2}1\underline{4}3) + (4\underline{1}2\underline{3} - 2\underline{1}4\underline{3})$$
$$A_2 = (\underline{3}2\underline{1}4 - \underline{3}4\underline{1}2) + (1\underline{4}3\underline{2} - 3\underline{4}1\underline{2}) + (\underline{4}3\underline{2}1 - \underline{4}1\underline{2}3) + (4\underline{3}2\underline{1} - 2\underline{3}4\underline{1})$$

These definitions are motivated by the next two lemmas, which show that a permuton μ satisfies the assumptions of Lemma 3.2 if for both $i = 1$ and $i = 2$, the value of A_i is zero for almost all τ_i-rooted permutons μ^{τ_i}.

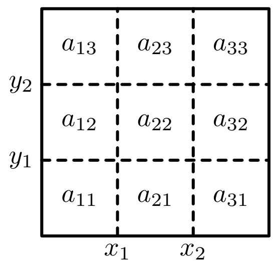


Figure 3.1: Notation used in the proof of Lemma 3.3.

Lemma 3.3. *Let μ be a permuton. If $h_\mu^{\tau_1}(A_1) = 0$ with probability 1, then*

$$\mu\left([x_1, x_2] \times [y_1, y_2]\right) = |x_2 - x_1| \cdot |y_2 - y_1|$$

for all points $(x_1, y_1), (x_2, y_2) \in supp\,(\mu)$ such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$.

Proof. Fix $(x_1, y_1), (x_2, y_2) \in \operatorname{supp}(\mu)$ such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$ and such that $h(A_1) = 0$ for the homomorphism $h : \mathcal{A}^{\tau_1} \to \mathbb{R}$ associated with the τ_1-rooted permuton $(\mu, (x_1, y_1), (x_2, y_2))$. Further let $(x_0, y_0) = (0, 0)$ and $(x_3, y_3) = (1, 1)$, and let

$$a_{ij} = \mu\left([x_{i-1}, x_i] \times [y_{j-1}, y_j]\right)$$

for $i, j \in [3]$. See Figure 3.1 for illustration of the just introduced notation.

Since $h(A_1) = 0$, the following holds:

$$a_{22}a_{33} - a_{23}a_{32} + a_{22}a_{11} - a_{12}a_{21} + a_{22}a_{31} - a_{21}a_{32} + a_{22}a_{13} - a_{12}a_{23} = 0\ .$$

We rewrite this expression using the property that μ has uniform marginals as follows:

$$\begin{aligned}
0 &= a_{22}a_{33} - a_{23}a_{32} + a_{22}a_{11} - a_{12}a_{21} + a_{22}a_{31} - a_{21}a_{32} + a_{22}a_{13} - a_{12}a_{23} \\
&= a_{22}(a_{11} + a_{13} + a_{31} + a_{33}) - (a_{21} + a_{23})(a_{12} + a_{32}) \\
&= a_{22}(1 - (x_2 - x_1) - (y_2 - y_1) + a_{22}) - (x_2 - x_1 - a_{22})(y_2 - y_1 - a_{22}) \\
&= a_{22} - (x_2 - x_1)(y_2 - y_1)
\end{aligned}$$

We conclude that the equality from the statement of the lemma holds for almost all points $(x_1, y_1), (x_2, y_2) \in \operatorname{supp}(\mu)$ such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$.

We next show that the equality in the statement of the lemma holds for all $(x_1, y_1), (x_2, y_2) \in \operatorname{supp}(\mu)$ such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$. Fix $(x_1, y_1), (x_2, y_2) \in$

supp (μ) such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$. If $x_1 = x_2$ or $y_1 = y_2$, then the equality holds since the measure μ has uniform marginals. Let $\varepsilon_0 = \min\{x_2 - x_1, y_2 - y_1\}$ and consider $\varepsilon \in (0, \varepsilon_0/2)$. Since all points in the ε-neighbourhood of (x_1, y_1) have both their coordinates smaller than all points in the ε-neighbourhood of (x_2, y_2), almost every point (x_1', y_1') in the intersection of supp (μ) and the ε-neighbourhood of (x_1, y_1) and almost every point (x_2', y_2') in the intersection of supp (μ) and the ε-neighbourhood of (x_2, y_2) satisfy the equality from the statement of the lemma, and it also holds that

$$\left|\mu\left([x_1, x_2] \times [y_1, y_2]\right) - \mu\left([x_1', x_2'] \times [y_1', y_2']\right)\right| \leqslant 4\varepsilon$$

because the measure μ has uniform marginals. Since both (x_1, y_1) and (x_2, y_2) are contained in supp (μ), the ε-neighbourhood of (x_1, y_1) has positive measure and the ε-neighbourhood of (x_2, y_2) also has positive measure, we conclude that

$$\left|\mu\left([x_1, x_2] \times [y_1, y_2]\right) - |x_2 - x_1| \cdot |y_2 - y_1|\right| \leqslant 8\varepsilon$$

for every $\varepsilon \in (0, \varepsilon_0/2)$. It follows that the equality from the statement of the lemma holds for all points $(x_1, y_1), (x_2, y_2) \in$ supp (μ) such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$. $\square$

A symmetric argument yields the following lemma.

Lemma 3.4. *Let μ be a permuton. If $h_\mu^{\tau_2}(A_2) = 0$ with probability 1, then*

$$\mu\left([x_1, x_2] \times [y_1, y_2]\right) = |x_2 - x_1| \cdot |y_2 - y_1|$$

for all points $(x_1, y_2), (x_2, y_1) \in$ supp (μ) such that $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$.

We are now ready to prove Theorems 3.5, 3.6, 3.7 and 3.8, which are based on the flag algebra method. We follow the standard path of applying the method by setting up appropriate SDP programs, i.e., supplying an objective as well as the types τ_1 and τ_2 and the integer $\ell = 4$ to the algorithm described at the end of Section 1.3. Solving these programs yields the positive semidefinite matrices M and vectors $A_1, \ldots, O_1$ and $A_2, \ldots, O_2$ used in the proofs of the four theorems.

Theorem 3.5. *Let $S = \{1234, 1243, 2134, 2143, 3412, 3421, 4312, 4321\}$. For every permuton μ,*

$$\sum_{\pi \in S} d(\pi, \mu) \geqslant \frac{1}{3}$$

and equality holds if and only if μ is uniform.

Proof. Let B_1, C_1, D_1 and E_1 be the following four elements of $\mathcal{A}^{\tau_1}$.

$$
\begin{aligned}
B_1 &= (1\underline{2}3\underline{4} - 3\underline{2}1\underline{4}) + (1\underline{23}4 - 4\underline{23}1) + (1\underline{24}3 - 3\underline{24}1) + (1\underline{2}4\underline{3} - 4\underline{2}1\underline{3})\\
C_1 &= (\underline{1}2\underline{3}4 - \underline{1}4\underline{3}2) + (1\underline{23}4 - 4\underline{23}1) + (\underline{2}1\underline{3}4 - \underline{2}4\underline{3}1) + (2\underline{13}4 - 4\underline{13}2)\\
D_1 &= (2\underline{1}4\underline{3} - 4\underline{1}2\underline{3}) + (1\underline{23}4 - 4\underline{23}1) + (2\underline{13}4 - 4\underline{13}2) + (1\underline{2}4\underline{3} - 4\underline{2}1\underline{3})\\
E_1 &= (\underline{2}1\underline{4}3 - \underline{2}3\underline{4}1) + (1\underline{23}4 - 4\underline{23}1) + (\underline{2}1\underline{3}4 - \underline{2}4\underline{3}1) + (1\underline{24}3 - 3\underline{24}1)
\end{aligned}
$$

Further, let B_2, C_2, D_2 and E_2 be the corresponding four elements of $\mathcal{A}^{\tau_2}$. For example, B_2 is the following element:

$$
B_2 = (1\underline{4}3\underline{2} - 3\underline{4}1\underline{2}) + (1\underline{32}4 - 4\underline{32}1) + (1\underline{42}3 - 3\underline{42}1) + (1\underline{3}4\underline{2} - 4\underline{3}1\underline{2})\,.
$$

Finally, let M be the following (positive definite) matrix.

$$
M = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 2 \end{bmatrix}
$$

A direct computation yields that

$$
\begin{aligned}
h_\mu\left(\llbracket w_1 M w_1^T \rrbracket_{\tau_1} + \llbracket w_2 M w_2^T \rrbracket_{\tau_2}\right) &= h_\mu\left(\frac{8}{9}\sum_{\pi\in S}\pi - \frac{2}{9}\sum_{\pi\in S_4\setminus S}\pi\right)\\
&= \frac{2}{3}\left(\sum_{\pi\in S} d(\pi,\mu) - \frac{1}{3}\right)
\end{aligned}
$$

where $w_1 = (B_1, C_1, D_1, E_1)$ and $w_2 = (B_2, C_2, D_2, E_2)$. Since the matrix M is positive semidefinite, $h_\mu\left(\llbracket w_1 M w_1^T \rrbracket_{\tau_1}\right) \geqslant 0$ and $h_\mu\left(\llbracket w_2 M w_2^T \rrbracket_{\tau_2}\right) \geqslant 0$, which implies that

$$
0 \leqslant \sum_{\pi\in S} d(\pi,\mu) - \frac{1}{3}.
$$

Moreover, the equality holds if and only if both $h_\mu^{\tau_1}(w_1 M w_1^T) = 0$ with probability 1 and $h_\mu^{\tau_2}(w_2 M w_2^T) = 0$ with probability 1. Since all the eigenvalues of the matrix M are positive, $h_\mu^{\tau_1}(w_1 M w_1^T) = 0$ if and only if $h_\mu^{\tau_1}(B_1) = 0$, $h_\mu^{\tau_1}(C_1) = 0$, $h_\mu^{\tau_1}(D_1) = 0$ and $h_\mu^{\tau_1}(E_1) = 0$. Since $A_1 = B_1 + C_1 - D_1 - E_1$, we conclude that if the equality holds, then $h_\mu^{\tau_1}(A_1) = 0$ with probability 1. A symmetric argument yields that if the equality holds, then $h_\mu^{\tau_2}(A_2) = 0$ with probability 1. The statement of the theorem now follows from Lemmas 3.2, 3.3 and 3.4. $\square$

We next prove the second main theorem of this section. Since the proofs of this theorem and the two subsequent to it are similar to the proof of Theorem 3.5, we will be brief in their parts that are analogous.

Theorem 3.6. *Let* $S = \{1234, 1432, 2143, 2341, 3214, 3412, 4123, 4321\}$. *For every permuton* μ,

$$\sum_{\pi \in S} d(\pi, \mu) \geqslant \frac{1}{3}$$

and equality holds if and only if μ *is uniform.*

Proof. Consider the following elements F_1 and G_1 of $\mathcal{A}^{\tau_1}$.

$$\begin{aligned} F_1 = {} & (1\underline{24}3 - 3\underline{24}1) + (4\underline{13}2 - 2\underline{13}4) + (\underline{1}24\underline{3} - \underline{1}42\underline{3}) + (\underline{2}31\underline{4} - \underline{2}13\underline{4}) \\ & + (\underline{13}24 - \underline{13}42) + (\underline{24}31 - \underline{24}13) + (31\underline{24} - 13\underline{24}) + (24\underline{13} - 42\underline{13}) \\ G_1 = {} & (\underline{12}43 - \underline{12}34) + (\underline{34}21 - \underline{34}12) + (\underline{14}32 - \underline{14}23) + (\underline{23}14 - \underline{23}41) \\ & + (43\underline{12} - 34\underline{12}) + (21\underline{34} - 12\underline{34}) + (32\underline{14} - 23\underline{14}) + (14\underline{23} - 41\underline{23}) \\ & + (\underline{1}43\underline{2} - \underline{1}34\underline{2}) + (\underline{3}21\underline{4} - \underline{3}12\underline{4}) + (\underline{1}32\underline{4} - \underline{1}23\underline{4}) + (\underline{2}14\underline{3} - \underline{2}41\underline{3}) \\ & + (3\underline{12}4 - 4\underline{12}3) + (1\underline{34}2 - 2\underline{34}1) + (2\underline{14}3 - 3\underline{14}2) + (4\underline{23}1 - 1\underline{23}4) \end{aligned}$$

Let F_2 and G_2 be the corresponding elements of $\mathcal{A}^{\tau_2}$ as in the proof of Theorem 3.5, and let M be the following (positive definite) matrix.

$$M = \begin{bmatrix} 5 & 0 & 3 \\ 0 & 9 & 0 \\ 3 & 0 & 4 \end{bmatrix}$$

Then

$$h_\mu \left(\llbracket w_1 M w_1^T \rrbracket_{\tau_1} + \llbracket w_2 M w_2^T \rrbracket_{\tau_2} \right) = 2 \left(\sum_{\pi \in S} d(\pi, \mu) - \frac{1}{3} \right)$$

where $w_1 = (A_1, F_1, G_1)$ and $w_2 = (A_2, F_2, G_2)$. This implies that

$$0 \leqslant \sum_{\pi \in S} d(\pi, \mu) - \frac{1}{3}$$

and equality holds if and only if both $h_\mu^{\tau_1}(w_1 M w_1^T) = 0$ with probability 1 and $h_\mu^{\tau_2}(w_2 M w_2^T) = 0$ with probability 1. Since all the eigenvalues of M are positive (the eigenvalues are 9 and $\frac{9 \pm \sqrt{37}}{2}$), it follows that equality holds if and only if $h_\mu^{\tau_1}(A_1) = 0$ with probability 1 and $h_\mu^{\tau_2}(A_2) = 0$ with probability 1. The statement of the theorem now follows from Lemmas 3.2, 3.3 and 3.4. □

We next prove the third main theorem of this section. Note that the set S from the statement of Theorem 3.7 is symmetric to the set $\{1324, 1423, 2314, 2413, 3142, 3241, 4132, 4231\}$ by a 90-degree rotation (i.e. the symmetry $\pi(i) = \sigma(k+1-i)$ listed in Section 3.2). Therefore, Theorem 3.7 proves that both the third and fourth sets from the statement of Theorem 3.1 are Σ-forcing.

Theorem 3.7. *Let* $S = \{1324, 1342, 2413, 2431, 3124, 3142, 4213, 4231\}$. *For every permuton* μ,

$$\sum_{\pi \in S} d(\pi, \mu) \leqslant \frac{1}{3}$$

and equality holds if and only if μ *is uniform.*

Proof. Let $\overline{S} = S_4 \setminus S$ and consider the following four elements of $\mathcal{A}^{\tau_1}$.

$$\begin{aligned}
H_1 &= (1\underline{2}3\underline{4} - 3\underline{2}1\underline{4}) + (\underline{2}3\underline{4}1 - \underline{2}1\underline{4}3) + (1\underline{2}4\underline{3} - 4\underline{2}1\underline{3}) + (\underline{2}4\underline{3}1 - \underline{2}1\underline{3}4) \\
I_1 &= (2\underline{1}4\underline{3} - 4\underline{1}2\underline{3}) + (\underline{1}4\underline{3}2 - \underline{1}2\underline{3}4) + (1\underline{2}4\underline{3} - 4\underline{2}1\underline{3}) + (\underline{2}4\underline{3}1 - \underline{2}1\underline{3}4) \\
J_1 &= (\underline{2}13\underline{4} - \underline{2}31\underline{4}) + (13\underline{24} - 31\underline{24}) + (3\underline{24}1 - 1\underline{24}3) + (\underline{24}13 - \underline{24}31) \\
&\quad + (4\underline{23}1 - 1\underline{23}4) + (14\underline{23} - 41\underline{23}) + (\underline{23}14 - \underline{23}41) + (\underline{2}14\underline{3} - \underline{2}41\underline{3}) \\
K_1 &= (24\underline{13} - 42\underline{13}) + (4\underline{13}2 - 2\underline{13}4) + (\underline{1}24\underline{3} - \underline{1}42\underline{3}) + (\underline{13}24 - \underline{13}42) \\
&\quad + (4\underline{23}1 - 1\underline{23}4) + (14\underline{23} - 41\underline{23}) + (\underline{23}14 - \underline{23}41) + (\underline{2}14\underline{3} - \underline{2}41\underline{3})
\end{aligned}$$

Further, let H_2, I_2, J_2 and K_2 be the corresponding elements of $\mathcal{A}^{\tau_2}$ as in the proof of Theorem 3.5, and let M be the following (positive definite) matrix.

$$M = \begin{bmatrix} 35 & 0 & 12 & 0 \\ 0 & 35 & 0 & -12 \\ 12 & 0 & 37 & 0 \\ 0 & -12 & 0 & 37 \end{bmatrix}$$

Then

$$h_\mu\left([\![w_1 M w_1^T]\!]_{\tau_1} + [\![w_2 M w_2^T]\!]_{\tau_2} \right) = 16 \left(\sum_{\pi \in \overline{S}} d(\pi, \mu) - \frac{2}{3} \right)$$

where $w_1 = (H_1, I_1, J_1, K_1)$ and $w_2 = (H_2, I_2, J_2, K_2)$. This implies that

$$0 \leqslant \sum_{\pi \in \overline{S}} d(\pi, \mu) - \frac{2}{3}$$

and equality holds if and only if both $h_\mu^{\tau_1}(w_1 M w_1^T) = 0$ with probability 1 and $h_\mu^{\tau_2}(w_2 M w_2^T) = 0$ with probability 1. Since all the eigenvalues of M are positive (the matrix has eigenvalues $36+\sqrt{145}$ and $36-\sqrt{145}$, each with multiplicity two), $h^{\tau_1}_\mu(w_1 M w_1^T)=0$ if and only if $h^{\tau_1}_\mu(H_1)=0$, $h^{\tau_1}_\mu(I_1)=0$, $h^{\tau_1}_\mu(J_1)=0$ and $h^{\tau_1}_\mu(K_1)=0$. Hence, if equality holds, then $h^{\tau_1}_\mu(A_1)=0$ with probability 1 (note that $A_1 = H_1 - I_1$). A symmetric argument yields that $h^{\tau_2}_\mu(A_2)=0$ with probability 1. The statement of the theorem now follows from Lemmas 3.2, 3.3 and 3.4. $\square$

Finally, we prove the last main theorem of this section.

Theorem 3.8. *Let* $S=\{1234,1243,1432,2134,2143,2341,3214,3412,3421,4123,4312,4321\}$. *For every permuton* μ,

$$\sum_{\pi\in S} d(\pi,\mu) \geqslant \frac{1}{2}$$

and equality holds if and only if μ *is uniform.*

Proof. Consider the following four elements of $\mathcal{A}^{\tau_1}$.

$$\begin{aligned}
L_1 &= (4\underline{2}1\underline{3} - 1\underline{2}4\underline{3}) + (4\underline{1}2\underline{3} - 2\underline{1}4\underline{3}) + (\underline{2}3\underline{4}1 - \underline{2}1\underline{4}3)\\
&\quad + (4\underline{23}1 - 1\underline{23}4) + (\underline{1}2\underline{3}4 - \underline{1}4\underline{3}2) + (3\underline{24}1 - 1\underline{24}3)\\
M_1 &= (\underline{2}1\underline{3}4 - \underline{2}4\underline{3}1) + (1\underline{23}4 - 4\underline{23}1) + (\underline{1}2\underline{3}4 - \underline{1}4\underline{3}2)\\
&\quad + (1\underline{24}3 - 3\underline{24}1) + (2\underline{13}4 - 4\underline{13}2) + (3\underline{24}1 - 1\underline{24}3)\\
N_1 &= (\underline{12}43 - \underline{12}34) + (21\underline{34} - 12\underline{34}) + (\underline{1}32\underline{4} - \underline{1}23\underline{4}) + (\underline{2}14\underline{3} - \underline{2}41\underline{3})\\
&\quad + (2\underline{14}3 - 3\underline{14}2) + (\underline{23}14 - \underline{23}41) + (32\underline{14} - 23\underline{14}) + (\underline{1}43\underline{2} - \underline{1}34\underline{2})\\
&\quad + (1\underline{34}2 - 2\underline{34}1) + (\underline{3}21\underline{4} - \underline{3}12\underline{4}) + (3\underline{12}4 - 4\underline{12}3) + (\underline{34}21 - \underline{34}12)\\
&\quad + (43\underline{12} - 34\underline{12}) + (\underline{14}32 - \underline{14}23) + (14\underline{23} - 41\underline{23}) + (4\underline{23}1 - 1\underline{23}4)\\
O_1 &= (\underline{1}42\underline{3} - \underline{1}24\underline{3}) + (\underline{13}42 - \underline{13}24) + (13\underline{24} - 31\underline{24}) + (\underline{24}13 - \underline{24}31)\\
&\quad + (42\underline{13} - 24\underline{13}) + (\underline{2}13\underline{4} - \underline{2}31\underline{4}) + (2\underline{13}4 - 4\underline{13}2) + (3\underline{24}1 - 1\underline{24}3)
\end{aligned}$$

Further, let L_2, M_2, N_2 and O_2 be the corresponding elements of $\mathcal{A}^{\tau_2}$ as in the proof of Theorem 3.5, and let M be the following (positive definite) matrix.

$$M = \begin{bmatrix} 1132 & -652 & -638 & 197 & 326\\ -652 & 774 & 516 & -68 & -326\\ -638 & 516 & 774 & 68 & -326\\ 197 & -68 & 68 & 172 & 0\\ 326 & -326 & -326 & 0 & 516 \end{bmatrix}$$

Then

$$h_\mu\left(\llbracket w_1 M w_1^T \rrbracket_{\tau_1} + \llbracket w_2 M w_2^T \rrbracket_{\tau_2}\right) = 172\left(\sum_{\pi\in S} d(\pi,\mu) - \frac{1}{2}\right)$$

where $w_1 = (A_1, L_1, M_1, N_1, O_1)$ and $w_2 = (A_2, L_2, M_2, N_2, O_2)$. This implies that

$$0 \leqslant \sum_{\pi\in S} d(\pi,\mu) - \frac{1}{2}$$

and equality holds if and only if both $h_\mu^{\tau_1}(w_1 M w_1^T) = 0$ with probability 1 and $h_\mu^{\tau_2}(w_2 M w_2^T) = 0$ with probability 1. Since all the eigenvalues of M are positive (because all the leading principal minors of M are positive), it follows that equality holds if and only if both $h_\mu^{\tau_1}(A_1) = 0$ with probability 1 and $h_\mu^{\tau_2}(A_2) = 0$ with probability 1. The statement of the theorem now follows from Lemmas 3.2, 3.3 and 3.4. □

3.4 Perturbations of the uniform permuton

In this section, we analyse pattern densities in step permutons obtained from the uniform permuton by a perturbation. This analysis will yield that most of the sets different from those listed in Theorem 3.1 are not Σ-forcing (see Lemma 3.15; the remaining cases will be dealt with individually in the proof of Lemma 3.16).

If A is a (non-negative) doubly stochastic square matrix of order n, i.e., each row sum and each column sum of A is equal to one, we can associate to it a *step permuton* $\mu[A]$ by setting

$$\mu[A](X) := \sum_{i,j\in[n]} A_{ij}\cdot n\cdot\left|X \,\cap\, \left[\frac{i-1}{n},\frac{i}{n}\right)\times\left[\frac{j-1}{n},\frac{j}{n}\right)\right|$$

for every Borel set $X \subseteq [0,1]^2$. A straightforward computation yields the following expression for the density of a k-permutation π in $\mu[A]$; we use $f : [k] \nearrow [n]$ to mean that f is a non-decreasing function from $[k]$ to $[n]$. Indeed, each of the summands corresponds to the probability that the $\mu[A]$-random permutation of order k is π and the k points defining π are sampled from the squares with coordinates $(f(i), g(\pi(i)))$, $i \in [n]$.

Lemma 3.9. *Let A be a doubly stochastic square matrix of order n, and π a k-permutation. Then*

$$d(\pi,\mu[A]) = \frac{k!}{n^k}\sum_{f,g:[k]\nearrow[n]} \frac{1}{\prod\limits_{i\in[n]} |f^{-1}(i)|!\cdot|g^{-1}(i)|!} \times \prod_{i\in[k]} A_{f(i),g(\pi(i))}.$$

For $i, j \in [n-1]$, let B^{ij} be the matrix such that

$$B^{ij}_{i'j'} = \begin{cases} +1 & \text{if either } i' = i \text{ and } j' = j \text{ or } i' = i+1 \text{ and } j' = j+1, \\ -1 & \text{if either } i' = i \text{ and } j' = j+1 \text{ or } i' = i+1 \text{ and } j' = j, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

In the following exposition, the order n of the matrices B^{ij} will always be clear from the context and so we use only the indices i and j to avoid unnecessarily complex notation.

For an integer n and a permutation π, we define a function $h_{\pi,n} : U_n \to \mathbb{R}$ on the cube $U_n := \{\vec{x} \in \mathbb{R}^{[n-1]^2} : \|\vec{x}\|_\infty \leqslant 1/4n\}$ around the origin as

$$h_{\pi,n}(x_{1,1}, \ldots, x_{n-1,n-1}) := d\left(\pi, \mu\left[A + \sum_{i,j\in[n-1]} x_{ij} B^{ij}\right]\right)$$

where A is the $n \times n$ matrix with all entries equal to $1/n$. Note that this is well-defined as $A + \sum_{i,j\in[n-1]} x_{ij} B^{ij}$ is doubly stochastic whenever $x_{i,j} \in U_n$ for all $i, j \in [n-1]$. More generally, if S is a set of permutations, we define $h_{S,n} : U_n \to \mathbb{R}$ as

$$h_{S,n}(\vec{x}) := \sum_{\pi\in S} h_{\pi,n}(\vec{x}) \ .$$

In this section, we are concerned with sets S that consist of 4-permutations only.

If S is a set of 4-permutations, we define the *cover matrix* of S to be a 4×4 matrix C^S such that C^S_{ij} is the number of permutations $\pi \in S$ such that $\pi(j) = i$. If the set S is clear from context, then we just write C for the cover matrix. We show that the gradient of $h_{S,n}$ at the origin is determined by the cover matrix of S.

Lemma 3.10. *Let n be an integer and S a set of 4-permutations with cover matrix C. Then*

$$\frac{\partial}{\partial x_{ij}} h_{S,n}(0, \ldots, 0) = \frac{4!}{n^7} \sum_{f,g:[4]\nearrow[n]} \frac{1}{\prod_{m\in[n]} |f^{-1}(m)|! \cdot |g^{-1}(m)|!} \times \left(\sum_{\substack{k\in f^{-1}(i)\\ \ell\in g^{-1}(j)}} C_{k,\ell} - \sum_{\substack{k\in f^{-1}(i+1)\\ \ell\in g^{-1}(j)}} C_{k,\ell} - \sum_{\substack{k\in f^{-1}(i)\\ \ell\in g^{-1}(j+1)}} C_{k,\ell} + \sum_{\substack{k\in f^{-1}(i+1)\\ \ell\in g^{-1}(j+1)}} C_{k,\ell} \right)$$

for every $i, j \in [n-1]$.

Proof. Since both the first derivative on the left hand side and the expression on the right hand side in the statement of the lemma are additive with respect to adding elements of the set S, it is enough to prove the lemma when S contains a single element π. In such case, the formula given in the statement of the lemma follows directly from Lemma 3.9. $\square$

Lemma 3.10 yields the following.

Lemma 3.11. *Let n be an integer and S a set of 4-permutations. If the cover matrix C is constant, then the gradient*

$$\nabla h_{S,n}(0,\dots,0) = \left(\frac{\partial}{\partial x_{ij}} h_{S,n}(0,\dots,0)\right)_{i,j\in[n-1]}$$

is zero.

Proof. We start by defining an operator on non-decreasing functions from [4] to $[n]$. Given $f : [4] \nearrow [n]$ and an index $k \in [n-1]$, we define $\tilde{f}^{(k)}$ as follows. Let Z be the image of f viewed as a multiset with every k replaced with $k+1$ and every $k+1$ replaced with k. Then $\tilde{f}^{(k)}$ is the unique non-decreasing function from [4] to $[n]$ whose image is Z. Informally speaking, we switch the values k and $k+1$ and reorder to obtain a non-decreasing function. Note that $f = \widetilde{(\tilde{f}^{(k)})}^{(k)}$ for all f and k, and $f = \tilde{f}^{(k)}$ if $|f^{-1}(k)| = |f^{-1}(k+1)|$.

We now analyse the individual summands inside the sum from the statement of Lemma 3.10. Fix two indices i and j, and a function $g : [4] \nearrow [n]$. If $f = \tilde{f}^{(i)}$, then the expression in the parenthesis evaluates to zero. If $f \neq \tilde{f}^{(i)}$, then the expressions for f and $\tilde{f}^{(i)}$ have opposite signs, in particular their contributions cancel out. We conclude that the sum is equal to zero if all the entries of the cover matrix C are the same. The lemma now follows. $\square$

Lemma 3.10 establishes that the gradient $\nabla h_{S,n}(0,\dots,0)$ for a set S of 4-permutations is a linear function of the entries of the cover matrix C of S. Analysing the matrix corresponding to this linear function for $n \in \{4,5\}$ yields that for such n the gradient $\nabla h_{S,n}(0,\dots,0)$ is zero if and only if the cover matrix of S is constant. Instead of providing this technical computation here, we give a more illustrative proof that the converse of Lemma 3.11 holds for large enough integers n, as this is sufficient for our exposition.

Lemma 3.12. *Let S be a set of 4-permutations whose cover matrix is not constant. Then there exists an integer n such that the gradient*

$$\nabla h_{S,n}(0,\ldots,0) = \left(\frac{\partial}{\partial x_{ij}} h_{S,n}(0,\ldots,0)\right)_{i,j\in[n-1]}$$

is non-zero.

Proof. We will assume that the gradient $\nabla h_{S,n}(0,\ldots,0)$ is zero and establish that the entries of the cover matrix satisfy $C_{k,\ell} - C_{k+1,\ell} - C_{k,\ell+1} + C_{k+1,\ell+1} = 0$ for all $k,\ell \in [3]$. We will then use this to show that the cover matrix C must be constant.

We start by analysing the partial derivative $\frac{\partial}{\partial x_{ij}} h_{S,n}(0,\ldots,0)$ for $i = 1$ and $j = 1$. Recall the notation $\tilde{f}^{(k)}$ from the proof of Lemma 3.11. If $|\mathrm{Im}(f) \cap \{1,2\}| \leqslant 1$ or $|\mathrm{Im}(g) \cap \{1,2\}| \leqslant 1$, then the summands in the expression given in Lemma 3.10 corresponding to (f,g), $(\tilde{f}^{(1)},g)$, $(f,\tilde{g}^{(1)})$ and $(\tilde{f}^{(1)},\tilde{g}^{(1)})$ sum to zero. Hence, we need to focus on the summands where $\{1,2\} \subseteq \mathrm{Im}(f)$ and $\{1,2\} \subseteq \mathrm{Im}(g)$. Note the number of summands such that f or g is not injective is $O(n^3)$, which yields the following.

$$\begin{aligned}\frac{\partial}{\partial x_{11}} h_{S,n}(0,\ldots,0) &= \frac{4!}{n^7}\left(\sum_{\substack{f,g:[4]\nearrow[n]\\ f(1)=1,f(2)=2,|\mathrm{Im}(f)|=4\\ g(1)=1,g(2)=2,|\mathrm{Im}(g)|=4}} (C_{11} - C_{12} - C_{21} + C_{22}) + O(n^3)\right)\\ &= \frac{4!}{n^7}\binom{n-2}{2}^2 (C_{11} - C_{12} - C_{21} + C_{22}) + O\left(\frac{1}{n^4}\right).\end{aligned}$$

If n is sufficiently large, the above expression can be zero only if $C_{11} - C_{12} - C_{21} + C_{22} = 0$. An analogous argument for $i = 1$ and $j = n-1$ yields that $C_{13} - C_{14} - C_{23} + C_{24} = 0$, for $i = n-1$ and $j = 1$ that $C_{31} - C_{32} - C_{41} + C_{42} = 0$, and for $i = n-1$ and $j = n-1$ that $C_{33} - C_{34} - C_{43} + C_{44} = 0$.

We next analyse the partial derivative $\frac{\partial}{\partial x_{ij}} h_{S,n}(0,\ldots,0)$ for $i = 1$ and $j = \lfloor n/2 \rfloor$. If $|\mathrm{Im}(f) \cap \{1,2\}| \leqslant 1$ or $|\mathrm{Im}(g) \cap \{\lfloor n/2\rfloor, \lfloor n/2\rfloor + 1\}| \leqslant 1$, then the summands in the expression given in Lemma 3.10 corresponding to (f,g), $(\tilde{f}^{(1)},g)$, $(f,\tilde{g}^{(\lfloor n/2\rfloor)})$ and $(\tilde{f}^{(1)},\tilde{g}^{(\lfloor n/2\rfloor)})$ sum to zero. Hence, we need to focus on the summands where $|\mathrm{Im}(f) \cap \{1,2\}| = 2$ and $|\mathrm{Im}(g) \cap \{\lfloor n/2\rfloor, \lfloor n/2\rfloor + 1\}| = 2$. Since the number of summands such that f or g is not injective is $O(n^3)$,

$$\frac{\partial}{\partial x_{1,\lfloor n/2\rfloor}} h_{S,n}(0,\ldots,0) = \frac{4!}{n^7}\left(\sum_{\substack{f,g:[4]\nearrow[n]\\ f(1)=1,f(2)=2,|\mathrm{Im}(f)|=4\\ g(1)=\lfloor n/2\rfloor,g(2)=\lfloor n/2\rfloor+1,|\mathrm{Im}(g)|=4}} (C_{11}-C_{12}-C_{21}+C_{22}) \right.$$
$$+ \sum_{\substack{f,g:[4]\nearrow[n]\\ f(1)=1,f(2)=2,|\mathrm{Im}(f)|=4\\ g(2)=\lfloor n/2\rfloor,g(3)=\lfloor n/2\rfloor+1,|\mathrm{Im}(g)|=4}} (C_{12}-C_{13}-C_{22}+C_{23})$$
$$\left. + \sum_{\substack{f,g:[4]\nearrow[n]\\ f(1)=1,f(2)=2,|\mathrm{Im}(f)|=4\\ g(3)=\lfloor n/2\rfloor,g(4)=\lfloor n/2\rfloor+1,|\mathrm{Im}(g)|=4}} (C_{13}-C_{14}-C_{23}+C_{24}) \right)$$
$$+ O\left(\frac{1}{n^4}\right).$$

Since the first and the third sum are equal to zero, we obtain that

$$\frac{\partial}{\partial x_{1,\lfloor n/2\rfloor}} h_{S,n}(0,\ldots,0) = (C_{12}-C_{13}-C_{22}+C_{23})\cdot\Theta\left(\frac{1}{n^3}\right) + O\left(\frac{1}{n^4}\right).$$

Hence, if n is large enough and this partial derivative is zero, it must hold that $C_{12}-C_{13}-C_{22}+C_{23}=0$. An analogous argument for $i=\lfloor n/2\rfloor$ and $j=1$ yields that $C_{21}-C_{22}-C_{31}+C_{32}=0$, for $i=n-1$ and $j=\lfloor n/2\rfloor$ that $C_{32}-C_{33}-C_{42}+C_{43}=0$, and for $i=\lfloor n/2\rfloor$ and $j=n-1$ that $C_{23}-C_{24}-C_{33}+C_{34}=0$.

Finally, we analyse the partial derivative $\frac{\partial}{\partial x_{ij}} h_{S,n}(0,\ldots,0)$ for $i=j=\lfloor n/2\rfloor$. As in the preceding two cases, we consider the functions $\tilde{f}^{(\lfloor n/2\rfloor)}$ and $\tilde{g}^{(\lfloor n/2\rfloor)}$ to conclude that the summands with $|\mathrm{Im}(f)\cap\{\lfloor n/2\rfloor,\lfloor n/2\rfloor+1\}|\leqslant 1$ or $|\mathrm{Im}(g)\cap\{\lfloor n/2\rfloor,\lfloor n/2\rfloor+1\}|\leqslant 1$ sum to zero. We next express the partial derivative as the sum of nine terms corresponding to injective mappings f and g with $\{\lfloor n/2\rfloor,\lfloor n/2\rfloor+1\}\subseteq\mathrm{Im}(f)$ and $\{\lfloor n/2\rfloor,\lfloor n/2\rfloor+1\}\subseteq\mathrm{Im}(g)$ (the terms are determined by the preimages of $\lfloor n/2\rfloor$ and $\lfloor n/2\rfloor+1$). Eight of these terms correspond to the sums of the entries of the cover matrix that we have already shown to be zero, which leads to the following expression for the considered partial derivative:

$$\frac{\partial}{\partial x_{\lfloor n/2\rfloor,\lfloor n/2\rfloor}} h_{S,n}(0,\ldots,0) = (C_{22}-C_{23}-C_{32}+C_{33})\cdot\Theta\left(\frac{1}{n^3}\right) + O\left(\frac{1}{n^4}\right).$$

Hence, if n is large enough and the partial derivative is zero, it must hold that

$C_{22} - C_{23} - C_{32} + C_{33} = 0$.

Since the cover matrix C satisfies that $C_{k,\ell} - C_{k,\ell+1} - C_{k+1,\ell} + C_{k+1,\ell+1} = 0$ for all $k, \ell \in [3]$, C is of the form

$$C = \begin{pmatrix} a & b & c & d \\ e & b+e-a & c+e-a & d+e-a \\ f & b+f-a & c+f-a & d+f-a \\ g & b+g-a & c+g-a & d+g-a \end{pmatrix}$$

for some integers $a, \ldots, g$. Since C is a cover matrix for a set S of 4-permutations, each row and each column must sum to $|S|$, i.e., the sums of the entries of each row are equal and the same holds for the columns of C. It follows that $b = c = d$ and $e = f = g$, so

$$C = \begin{pmatrix} a & b & b & b \\ e & b+e-a & b+e-a & b+e-a \\ e & b+e-a & b+e-a & b+e-a \\ e & b+e-a & b+e-a & b+e-a \end{pmatrix}.$$

It now follows that $b = e$ (otherwise, the sum of the second row and the second column would differ), which yields that the matrix C must be of the form

$$C = \begin{pmatrix} a & b & b & b \\ b & 2b-a & 2b-a & 2b-a \\ b & 2b-a & 2b-a & 2b-a \\ b & 2b-a & 2b-a & 2b-a \end{pmatrix}.$$

Hence, we get that $a + 3b = 7b - 3a$, which yields that $a = b$. We conclude that the matrix C is constant. □

The following lemma will be used to analyse sets of 4-permutations with constant cover matrix.

Lemma 3.13. *Let S be a set of 4-permutations such that the cover matrix C is constant. The Hessian matrix of the second order partial derivatives of $h_{S,5}$ at $(0, \ldots, 0)$ has both a positive and a negative eigenvalue, unless S is symmetric to one of the following sets of 4-permutations*

- $\{1234, 2143, 3412, 4321\}$,
- $\{1234, 1243, 2134, 2143, 3412, 3421, 4312, 4321\}$,
- $\{1234, 1432, 2143, 2341, 3214, 3412, 4123, 4321\}$,

- $\{1324, 1342, 2413, 2431, 3124, 3142, 4213, 4231\}$,
- $\{1342, 1423, 2314, 2431, 3124, 3241, 4132, 4213\}$,
- $\{1234, 1243, 1324, 2134, 2143, 2413, 3142, 3412, 3421, 4231, 4312, 4321\}$,
- $\{1234, 1243, 1342, 2134, 2143, 2431, 3124, 3412, 3421, 4213, 4312, 4321\}$,
- $\{1234, 1243, 1342, 2134, 2143, 2431, 3214, 3412, 3421, 4123, 4312, 4321\}$,
- $\{1234, 1243, 1432, 2134, 2143, 2341, 3214, 3412, 3421, 4123, 4312, 4321\}$,
- $\{1234, 1243, 1432, 2134, 2341, 2413, 3142, 3214, 3421, 4123, 4312, 4321\}$,
- $\{1234, 1243, 1432, 2143, 2314, 2341, 3214, 3412, 3421, 4123, 4132, 4321\}$,
- $\{1234, 1342, 1423, 2143, 2314, 2431, 3124, 3241, 3412, 4132, 4213, 4321\}$,
- $\{1234, 1342, 1423, 2314, 2413, 2431, 3124, 3142, 3241, 4132, 4213, 4321\}$,

or their complements.

Proof. For a 4-permutation π, let H_π be the Hessian matrix (of order sixteen)

$$\left(\frac{\partial^2}{\partial x_{ij}\partial x_{i'j'}} h_{\{\pi\},5}(0,\ldots,0)\right)_{i,j,i',j'\in[4]}.$$

The matrices H_π for all 4-permutations can be found in Appendix 1. For a set S of 4-permutations, let H_S be the corresponding Hessian matrix, i.e.,

$$H_S = \sum_{\pi\in S} H_\pi.$$

Note that $H_S = -H_{\overline{S}}$ where $\overline{S}$ is the complement of S with respect to the set of all 4-permutations. If the cover matrix of S is constant, then $|S|$ must be divisible by four. Up to symmetry, there are 12 sets S with 4 elements and 65 sets S with 8 elements whose cover matrix is constant. Up to symmetry and taking complements, there are 68 sets S with 12 elements whose cover matrix is constant. These sets are listed in Appendices 2–4 together with the corresponding matrices H_S and their largest and smallest eigenvalues. An inspection of these values yields the statement of the lemma (the sets S such that the matrix H_S does not have both positive and negative eigenvalues are highlighted by the bold font in Appendices 2–4). □

We are now ready to prove the main theorem of this section.

Theorem 3.14. *Let S be a set of 4-permutations. There exists an integer n and $\vec{x}, \vec{y} \in U_n$ such that $h_{S,n}(\vec{x}) < |S|/24 < h_{S,n}(\vec{y})$, unless S is symmetric to one of the sets of 4-permutations listed in Lemma 3.13, or to the complement of one of them.*

Proof. If the cover matrix C of S is not constant, then there exists an integer n such that the gradient $\nabla h_{S,n}(0, \dots, 0)$ is non-zero by Lemma 3.12. Hence, we can set $\vec{x} = -\varepsilon \nabla h_{S,n}(0, \dots, 0)$ and $\vec{y} = \varepsilon \nabla h_{S,n}(0, \dots, 0)$ for a sufficiently small positive ε. If the cover matrix C of S is constant, then $\nabla h_{S,n}(0, \dots, 0)$ is zero for every integer n by Lemma 3.11, in particular, for $n = 5$. However, unless S is symmetric to one of the sets of 4-permutations listed in Lemma 3.13 or to the complement of one of them, the Hessian matrix of the second partial derivatives of $h_{S,5}$ at $(0, \dots, 0)$ has both positive and negative eigenvalues. Hence, we can set $\vec{x}$ to be an ε-multiple of the eigenvector corresponding to a negative eigenvalue of the Hessian matrix and $\vec{y}$ to be an ε-multiple of the eigenvector corresponding to a positive eigenvalue for a sufficiently small positive ε. □

3.5 Non-Σ-forcing sets

We start this section with a lemma which asserts that in order to show that a set S of 4-permutations is not Σ-forcing, it is enough to find a permuton where the sum of pattern densities is smaller than $|S|/24$, and a permuton where the sum of pattern densities is larger than $|S|/24$.

Lemma 3.15. *Let S be a set of 4-permutations. If there exist permutons μ_1 and μ_2 such that*

$$\sum_{\pi \in S} d(\pi, \mu_1) < \frac{|S|}{24} \text{ and } \sum_{\pi \in S} d(\pi, \mu_2) > \frac{|S|}{24},$$

then there exists a non-uniform permuton μ such that

$$\sum_{\pi \in S} d(\pi, \mu) = \frac{|S|}{24}.$$

Proof. Define a permuton μ_λ for $\lambda \in (1, 2)$ as follows:

$$\begin{aligned} \mu_\lambda(X) = (2 - \lambda) \cdot \mu_1 &\left(\frac{1}{2 - \lambda} \times \left(X \cap [0, 2 - \lambda]^2 \right) \right) \\ &+ (\lambda - 1) \cdot \mu_2 \left(\frac{1}{\lambda - 1} \times \left(X \cap [2 - \lambda, 1]^2 - (2 - \lambda, 2 - \lambda) \right) \right), \end{aligned}$$

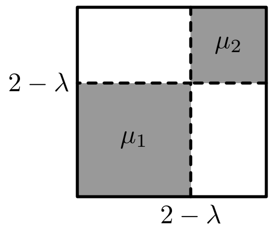


Figure 3.2: The permuton μ_λ in the proof of Lemma 3.15. The support of the permuton lies in the grey area.

where $\alpha \times X$ stands for $\{\alpha \cdot x, x \in X\}$ and $X - v$ for $\{x - v, x \in X\}$. The definition of a permuton μ_λ is illustrated in Figure 3.2. Note that μ_λ is μ_1 for $\lambda = 1$ and μ_2 for $\lambda = 2$. Next define a function $f : [1, 2] \to [0, 1]$ as

$$f(\lambda) = \sum_{\pi \in S} d(\pi, \mu_\lambda).$$

Observe that f is a continuous function on the interval $[1, 2]$. Hence, there exists $\lambda \in (1, 2)$ such that $f(\lambda) = |S|/24$. Since the permuton μ_λ is not uniform for any $\lambda \in (1, 2)$, the statement of the lemma follows. □

We are now ready to prove the main theorem of this section.

Theorem 3.16. *Let S be a set of 4-permutations. Then there exists a non-uniform permuton μ such that*

$$\sum_{\pi \in S} d(\pi, \mu) = \frac{|S|}{24}$$

unless the set S is one of the following sets of 4-permutations

- $\{1234, 1243, 2134, 2143, 3412, 3421, 4312, 4321\}$,
- $\{1234, 1432, 2143, 2341, 3214, 3412, 4123, 4321\}$,
- $\{1324, 1342, 2413, 2431, 3124, 3142, 4213, 4231\}$,
- $\{1324, 1423, 2314, 2413, 3142, 3241, 4132, 4231\}$,
- $\{1234, 1243, 1432, 2134, 2143, 2341, 3214, 3412, 3421, 4123, 4312, 4321\}$,

or their complements.

Proof. Fix a set S of 4-permutations that is not one of the sets listed in the statement of the lemma. We can assume that $|S| \leqslant 12$ by considering the complement of S if necessary. By Lemma 3.15, it suffices to find permutons μ_1 and μ_2 such that the sum of the pattern densities of the permutations contained in S for μ_1 is less than $|S|/24$ and for μ_2 is larger than $|S|/24$. If S is not symmetric to a set listed in the statement of Lemma 3.13, such permutons μ_1 and μ_2 exist by Theorem 3.14. Hence, we can assume that S is one of the 9 sets listed in the statement of Lemma 3.13 but not in the statement of Theorem 3.16.

We first consider the case $S = \{1342, 1423, 2314, 2431, 3124, 3241, 4132, 4213\}$. We choose μ_1 to be the *monotone increasing permuton*, i.e., the unique permuton such that $\operatorname{supp}(\mu)_1 = \{(x,x), x \in [0,1]\}$. The density of a pattern π in μ_1 is 1 if π is increasing and 0 otherwise; in particular, the sum of the pattern densities of the permutations from S is zero. Next, consider the following doubly stochastic matrix A

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and set $\mu_2 = \mu[A]$. A direct computation yields that the sum of the pattern densities of the permutation contained in S in μ_2 is $\frac{25}{72} > \frac{1}{3}$.

Each of the eight sets S that remain to be considered contain the permutation 1234. Hence, we set μ_2 to be the monotone increasing permuton. The permutons μ_1 for these sets can be chosen as step permutons corresponding to doubly stochastic matrices listed in Appendix 5. □

Chapter 4

Inducibility of Trees

4.1 Overview

In this chapter, we resolve three questions posed by Bubeck and Linial [22] about limit densities of subtrees in trees. Recall that the density $d(S,T)$ of a k-vertex tree S in a tree T is the number of embeddings of S in T divided by the number of k-vertex subtrees of T, and that the inducibility of S the maximum possible density of S in a large tree. Our main result is:

Theorem 4.1. *The inducibility of every tree S that is neither a star nor a path is at most* $1 - 10^{-35}$.

The theorem is proved in Section 4.6, and relies on several preliminary results in Sections 4.3–4.5. Theorem 4.1 is complemented in Section 4.7 by a lower bound on the inducibility of a class of trees that we call *sparklers*, which implies the following:

Theorem 4.2. *There are infinitely many trees with inducibility at least* $13/165$.

Lastly, in Section 4.8 we construct a universal sequence of trees.

Theorem 4.3. *There exists a sequence $(T_n)_{n\in\mathbb{N}}$ of trees in which the limit density $\lim_{n\to\infty} d(S,T_n)$ of every tree S is positive.*

4.2 Preliminaries

Given a vertex v in a tree T, a *branch* of T rooted at v is a subtree of T formed by a component of the graph $T \setminus v$ together with its edge to v. A branch is *non-trivial* if it is not a single edge; in other words, it does not correspond to a leaf of T. A non-trivial branch rooted at a vertex v is a *fork* if it is isomorphic to a star (note that

v must be a leaf of this star). The *order* of a fork is its number of (non-root) leaves. A branch is *major* if it is a non-trivial branch that is not a fork. A *caterpillar* is a tree T such that every vertex of T is the root of at most two non-trivial branches. Finally, a vertex of a caterpillar that is not a leaf is called *internal.* Observe that a tree is a caterpillar if and only if its internal vertices induce a path.

Czabarka, Székely and Wagner [38, Theorem 1 and Lemma 4] proved the following result about limit densities in trees of bounded radius.

Proposition 4.4 ([38])**.** *Let $(T_n)_{n\in\mathbb{N}}$ be a convergent sequence of trees with $|T_n| \to \infty$. If there exists an integer K such that the radius of each T_n is at most K, then*

$$\lim_{n\to\infty} d(S_k, T_n) = 1$$

for every $k \in \mathbb{N}$, where S_k is the k-vertex star.

As mentioned at the end of Section 1.4, the result below is proved independently in [21, Theorem 2] and [38, Theorem 1].

Proposition 4.5 ([21,38])**.** *Let $k \geqslant 4$ and let $(T_n)_{n\in\mathbb{N}}$ be a convergent sequence of trees with $|T_n| \to \infty$. If $\lim_{n\to\infty} d(P_k, T_n) = 0$, then $\lim_{n\to\infty} d(S_k, T_n) = 1$.*

A *center* of a tree T is a vertex v such that each branch rooted at v has at most $|T|/2$ edges. Every tree T has either one or two centers. Moreover, if T has two centers, then $|T|$ is even, the two centers are adjacent, each center has a branch rooted at it with exactly $|T|/2$ edges, and the other center is its neighbour in this branch. A *hub* of a tree T is a vertex v that is the only vertex on the path from v to the nearest center of T that is the root of at least three non-trivial branches. In particular, if a center of T is the root of at least three non-trivial branches, then it is a hub.

Proposition 4.6. *Every tree T that is not a caterpillar has at least one and at most two hubs.*

Proof. Let T' be the tree obtained from T by removing all of its leaves. The degree of a vertex v in T' is equal to the number of non-trivial branches rooted at v in T. Since T is not a caterpillar, T' is not a path. Therefore, T' contains a vertex of degree at least 3, so T has at least one hub.

Let W be the set of vertices of T' with degree at least 3. Suppose that T has a single center v_C. If v_C has degree at least three in T', then v_C is the only hub of T. Otherwise, the degree of v_C in T' is equal to 1 or 2 and there exists at least one and

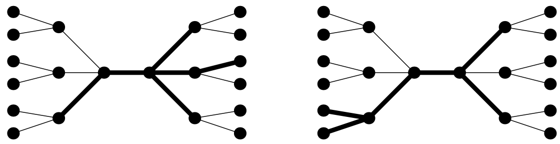


Figure 4.1: Two embeddings of 7-vertex trees that can be obtained from each other by moving two edges. The edges of the embeddings are in bold.

at most two vertices $w \in W$ such that there is no other vertex of W on the unique path between v_C and w. These vertices w are the hubs of T.

In the case that T has two centers v_C and v'_C, which are necessarily adjacent, then each center is a hub if its degree in T' is at least three. Otherwise, there exists at most one vertex $w \in W$ such that the unique path between v_C and w contains neither another vertex of W nor v'_C. Similarly, there exists at most one vertex $w \in W$ such that the unique path between v'_C and w contains neither another vertex of W nor v_C. Hence, T has at most two hubs. □

Let S_0 and S be embeddings of trees in a tree T with $|S_0| = |S| = n$, and let k be an integer less than n. (In fact, we only use $k \leqslant 3$). We say that S can be obtained from S_0 by *moving* k edges if the intersection of S_0 and S is a subtree of T with $n - k$ vertices (see Figure 4.1). In this sense, S is said to be obtained from S_0 by *removing* the edges of S_0 that are not contained in S, and then *adding* the edges of S that are not contained in S_0.

We next bound the number of vertices that can become a center of an embedding of a tree when at most three edges are moved.

Proposition 4.7. *Let S_0 be an embedding of a tree with at least 17 vertices in another tree T. There exists a set X of at most 8 vertices of T such that if three or fewer edges of S_0 are moved to produce an embedding S of a tree in T, then each center of S is contained in X.*

Proof. Let $n = |S_0| \geqslant 17$. Let X be the set of vertices v of S_0 such that each branch rooted at v has at most $n/2 + 3$ edges. We claim that X has the property given in the statement of the lemma. Indeed, if S is an embedding obtained from S_0 by moving at most three edges and w is a center of S, then each branch of S rooted at w has at most $n/2$ edges and so each branch of S_0 rooted at w has at most $n/2 + 3$ edges. Hence, w is contained in X.

It remains to estimate $|X|$. We call a branch B of S_0 *significant* if B is rooted at a center of S_0, has at least $n/2 - 3$ edges, and does not contain the other center (if another center exists). Every vertex $x \in X$ is either a center or is contained in a significant branch—otherwise, the branch rooted at x containing the center(s) has at least $(n-1)-(n/2-4)+1 = n/2+4$ edges. Since significant branches are edge-disjoint and $3(n/2-3) = (n-1)+(n/2-8) > n-1$, S_0 has at most two significant branches. Note that each significant branch has at most $\lfloor n/2 \rfloor$ edges since it is rooted at a center of S_0, so the other branches rooted at the same center contain at least $\lceil n/2 \rceil - 1$ edges in total.

Therefore, if n is odd, each significant branch has at most three vertices w such that the branch rooted at w containing the center(s) has at most $\lceil n/2 \rceil + 2 = \lfloor n/2 \rfloor + 3$ edges. If n is even and S_0 has two centers, then the branches rooted at each center that contain the other center have exactly $n/2$ edges. So again, each significant branch has at most three vertices w such that the branch rooted at w containing the center(s) has at most $n/2+3$ edges. Lastly, if n is even and S_0 has only one center, then we use the fact that there is at most one significant branch with exactly $n/2$ edges. This branch, if it exists, has at most four vertices w such that the branch rooted at w containing the center has at most $n/2+3$ edges; any other significant branch has at most three such vertices w. In each case, $|X| \leqslant 8$. □

Note that the bound on $|X|$ in Proposition 4.7 is best possible since it is attained when S_0 is a path with an even number of vertices.

We finish this section by bounding the number of vertices that can become a hub of an embedding of a tree when at most three edges are moved.

Proposition 4.8. *Let S be a non-caterpillar tree with at least 17 vertices, and fix an embedding of a tree S' with $|S'| = |S|$ in a tree T. There exists a set X of at most 144 vertices of T such that if an embedding of S in T can be obtained by moving three or fewer edges of S', then each hub of the obtained embedding of S is contained in X.*

Proof. Let X_0 be the set of the vertices from Proposition 4.7 applied with $S_0 = S'$, and let D be the set of distances between the hubs of S and the nearest center in S. By Proposition 4.6, S has at most two hubs, so $|D| \leqslant 2$.

For a vertex z in the embedding of S', define the *resistance* of z as the number of edges not incident with z that are contained in branches of S' rooted at z with the two largest branches excluded. Informally speaking, the resistance of z is the number of edges that must be removed from S' so that z is no longer the root of three non-trivial branches, and therefore not a candidate hub.

Consider a vertex $x \in X_0$ that is a center of an embedding of S in T obtained by moving at most three edges of S'. Observe that in this embedding of S, a vertex

$v \neq x$ of T can be a hub whose nearest center is x only if the following holds:

- v is an internal vertex of S',
- the distance d between v and x belongs to the set D, and
- the sum of the resistance of x and the resistances of the internal vertices on the path between v and x is at most 3.

Let X be the union of X_0 with the set of vertices v that satisfy these three conditions for some $x \in X_0$.

For a vertex $x \in X_0$, let Z_x be the union of $\{x\}$ with the set of internal vertices z of S' such that the sum of the resistance of x and the resistances of the internal vertices on the path between z and x is at most 3. Observe that if a vertex z belongs to Z_x, then all vertices on the path between z and x also belong to Z_x. Define S_Z to be the subtree of S' induced by Z_x, and note that the resistance of z is an upper bound on the number of leaves of S_Z lying in non-trivial branches rooted at z with the two largest branches excluded. Let δ be the number of branches of S_Z rooted at x. Since each of the δ branches of S_Z rooted at x has at most $4-\gamma \leqslant \min\{4, 6-\delta\}$ leaves, where $\gamma \geqslant \delta - 2$ is the resistance of x, the tree S_Z has at most 9 leaves. This implies that the number of vertices of Z_x lying at a distance contained in D from x is at most 18. Hence, the set X contains at most $8 \cdot 18 = 144$ vertices. □

4.3 Inducibility of trees with three large branches

In this section, we present a part of the proof of Theorem 4.1 for trees with three large branches rooted at a hub. For a k-vertex tree S and a host tree T, one approach would be to construct a function $f_{S,T}$ that maps each embedding of S in T to an embedding of a k-vertex subtree of T non-isomorphic to S such that at most α embeddings of S are mapped to the same subtree of T, where α is a constant independent of S and T. This would imply that the inducibility of S is at most $\alpha/(\alpha+1)$. An explicit construction of such a function $f_{S,T}$ is technical, so we prove its existence implicitly using a discharging argument.

Theorem 4.9. *Assume S is a k-vertex tree ($k \geqslant 17$) with a fixed hub v_S that is either adjacent to at most one leaf, or is the root of at least three major branches and at most one fork. If T is a tree with radius at least $4k$, then $d(S,T) \leqslant 1 - 10^{-7}$.*

Proof. Let v_T be a vertex of T such that there exist $(2k+1)$-vertex paths P_1 and P_2 starting at v_T that are disjoint except at v_T itself; such a choice is possible because the radius of T is at least $4k$. For every vertex v of T, fix a linear order $\preceq_v$ of the edges incident with v.

If there is at most one leaf adjacent to v_S, then we say that every non-trivial branch rooted at v_S is *important*; otherwise, a branch rooted at v_S is said to be *important* only if it is major. Note that there are at least three important branches regardless which of the two cases described in the statement of the theorem apply. In this proof, a *stub* is an embedding of a $(k-3)$-vertex tree S' in T with a distinguished vertex v' and three distinguished branches, together with a correspondence between the distinguished branches and three (isomorphism classes of) branches of S rooted at v_S such that it is possible to add a single leaf to each of the distinguished branches of S' so that there is an isomorphism from S' to S that maps $v_{S\to T}$ to v_S and the vertices of each of the distinguished branches of S' to the vertices of the corresponding branch of S. The three distinguished branches of the stub are referred to as *grafts*.

We next introduce a canonical way of obtaining a stub from an embedding of S in T. For an embedding of S in T, let $v_{S\to T}$ be the vertex of T corresponding to the hub v_S; if there are two possible choices for $v_{S\to T}$, we choose an arbitrary one. Consider the three important branches of the embedding rooted at $v_{S\to T}$ whose edges incident with $v_{S\to T}$ appear earliest in the linear order $\preceq_v$. These three branches will be denoted by S_A, S_B, and S_C; we will decide which branch is S_A, which is S_B, and which is S_C later in the proof. Consider the DFS traversal of the branches S_A, S_B, and S_C from $v_{S\to T}$ such that the edges at each vertex v of the branches are visited in the order given by $\preceq_v$; that is, a part of the branch joined by an edge earlier in the order $\preceq_v$ is explored first. We obtain the stub S' by removing the leaf of the embedding that appears last in the DFS traversal in the branch S_A, the leaf that appears last in the branch S_B, and the leaf that appears last in the branch S_C. The branches obtained from S_A, S_B, and S_C are the grafts of the stub S'.

Let v_A, v_B, and v_C be the vertices of the branches S_A, S_B, and S_C adjacent to the removed leaves of the embedding, and let d_A, d_B, and d_C be the degrees of v_A, v_B, and v_C in the embedding of S, respectively. Fix the indexing of the branches S_A, S_B, and S_C so that $d_A \geqslant d_B \geqslant d_C \geqslant 2$. The edges of T incident with v_A, v_B, and v_C, respectively, that appear in the orders $\preceq_{v_A}$, $\preceq_{v_B}$, and $\preceq_{v_C}$ after the edge of the embedding of S that is visited second-last by the DFS traversal are referred to as *active*, with the possible exception of the edge towards the vertex $v_{S\to T}$; that is, the edge incident with v_A, v_B, or v_C on the path to $v_{S\to T}$ is never active. In particular, if $d_A = 2$ then all edges incident with v_A except the one towards the vertex $v_{S\to T}$

are active. Observe that no active edges are contained in the stub S' and the only active edges contained in the embedding of S are the three edges incident with the removed leaves. Let s_A, s_B, and s_C be the number of active edges incident with v_A, v_B, and v_C, respectively.

Observe that if a stub S' is fixed, including the choice of the distinguished vertex and grafts, the vertices v_A, v_B and v_C are uniquely determined: they are the last vertices in the DFS traversal uniquely determined by the orders $\preceq_v$ whose distance is 1 less than the distance of the missing leaf. Hence, the same stub S' can be obtained from exactly $s_A s_B s_C$ embeddings of a tree S in the tree T.

Define Q to be the unique path in the tree T between the vertices $v_{S\to T}$ and v_T prolonged by P_1 if P_1 does not contain the edge from v_T towards $v_{S\to T}$ and prolonged by P_2 otherwise. Since P_1 and P_2 each have $2k+1$ vertices, Q has at least $k+1$ vertices not contained in the embedding of S. We next construct a set $\mathcal{S}$ of embeddings of several k-vertex trees non-isomorphic to S in T as follows.

- If $s_A \geqslant 3$ and $s_A \geqslant s_B$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding two active edges incident with v_A and an active edge incident with v_C.
- If $3 \leqslant s_A < s_B$ and $s_B \geqslant s_C$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_A and two active edges incident with v_B.
- If $3 \leqslant s_A < s_B < s_C$ and $d_B \neq d_C + 1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_A and two active edges incident with v_C.
- If $3 \leqslant s_A < s_B < s_C$ and $d_A - 1 \neq d_B = d_C + 1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding two active edges incident with v_B and an active edge incident with v_C.
- If $3 \leqslant s_A < s_B < s_C$ and $d_A - 1 = d_B = d_C + 1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_B and two active edges incident with v_C.
- If $s_A < 3$, $s_B \geqslant 3$ and $s_B \geqslant s_C$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_A and two active edges incident with v_B.
- If $s_A < 3 \leqslant s_C$, $s_B < s_C$ and $d_B \neq d_C + 1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_A and two active edges incident with v_C.

- If $s_A < 3 \leqslant s_C$, $s_B < s_C$ and $d_A \neq d_B = d_C+1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding an active edge incident with v_B and two active edges incident with v_C.

- If $s_A < 3 \leqslant s_C$, $s_B < s_C$ and $d_A = d_B = d_C+1$, then $\mathcal{S}$ contains all embeddings obtained from S' by adding three active edges incident with v_C.

- If s_A, s_B, and s_C are all less than 3, then $\mathcal{S}$ contains the unique tree that is obtained from S' by adding the first three edges of Q that are not already contained in S'.

The above ten cases cover all values of s_A, s_B, s_C, d_A, d_B, and d_C satisfying $s_A, s_B, s_C \geqslant 1$ and $d_A \geqslant d_B \geqslant d_C \geqslant 2$. In each case, the degree sequence of every tree in $\mathcal{S}$ is different from the degree sequence of S. For example, in the first case, trees in $\mathcal{S}$ contain more vertices of degree $d_A + 1$ than S, and in the last case, the tree in $\mathcal{S}$ has fewer leaves than S. Therefore, none of the embeddings in $\mathcal{S}$ is an embedding of a tree isomorphic to S. In all but the last case, $|\mathcal{S}| \geqslant s_A s_B s_C/12$. For example, in the first case,

$$|\mathcal{S}| = \binom{s_A}{2} s_C = \frac{s_A(s_A-1)s_C}{2} \geqslant \frac{s_A^2 s_C}{3} \geqslant \frac{s_A s_B s_C}{3} \geqslant \frac{s_A s_B s_C}{12}.$$

Cases 2–8 follow similarly (by permuting the letters A, B, and C in the argument above). In the second-last case, since $s_A \leqslant 2$,

$$|\mathcal{S}| = \binom{s_C}{3} \geqslant \frac{s_C(s_C-1)}{6} \geqslant \frac{s_C s_B}{6} \geqslant \frac{s_A s_B s_C}{12}.$$

In the last case, $|\mathcal{S}| = 1 \geqslant s_A s_B s_C/8$. This implies that $|\mathcal{S}|$ is at least the number of embeddings of S yielding the stub S' divided by 12.

Fix an embedding S'' of a k-vertex tree that is not isomorphic to S. We now estimate the number of stubs S' associated with an embedding of S whose corresponding set $\mathcal{S}$ contains S''. We will create a stub S' from S'' by following constructive steps that we next describe. The steps sometimes result in a tree that cannot be a stub of an embedding of the tree S, however, any stub S' associated with an embedding of S such that the corresponding set $\mathcal{S}$ contains S'' can be created by following the described steps.

The distinguished vertex of S' must be a hub $v_{S\to T}$ of an embedding of S that can be transformed into S'' by moving at most three edges. By Proposition 4.8, there are at most 144 choices for $v_{S\to T}$. Once the vertex $v_{S\to T}$ is chosen, it needs to be decided which three branches of S'' rooted at $v_{S\to T}$ correspond to grafts of the stub S'. Suppose S has at most one leaf adjacent to v_S, which implies that every non-trivial branch rooted at v_S is important. Every graft of S' corresponds to either a leaf (as an important branch can be become trivial after the removal of a single edge—note that at most three additional leaves can be created in this way), or to one of the first four non-trivial branches rooted at v_S in the order given by $\preceq_{v_S}$ (as a new non-trivial branch can be created by adding the first three edges of Q). Hence, there are at most eight branches of S'' that could possibly be grafts in S' when $v_{S\to T}$ is chosen. Similarly, if S has at most one fork rooted at v_S, every graft of S' corresponds to either a fork (as an important branch can become a fork after the removal of a single edge—again, at most three additional forks can be created in this way), or to one of the first four major branches rooted at v_S in the order given by $\preceq_{v_S}$. Again, there are at most eight branches of S'' that could possibly be grafts in S'.

Next, fix a triple among the at most eight branches that could be the three grafts of S'. Observe that the degree of $v_{S\to T}$ in S'' is the same as the degree of v_S in S unless a new branch at $v_{S\to T}$ was created by adding the first three edges on Q; that is, these three edges form a branch rooted at $v_{S\to T}$ in S''. In the latter case, remove the three edges of Q that have been added to get the same number of branches in the embedding as in S'. The correspondence between the branches of the embedding and S' different from the grafts is given by their isomorphism to the branches of S rooted at v_S. Three branches of S remain unmatched in this way and these can correspond in $3! = 6$ ways to the grafts. When the correspondence of these three branches and the grafts is fixed, it is uniquely determined which edges of S'' need to be removed to get the stub S'. For example, if one of the branches of S'' has two additional edges but not two additional leaves compared to the corresponding branch of S, then the last of the ten cases applied (that is, three edges from Q were added), and we just remove the three edges of Q to get S'. Otherwise, the difference between the number of edges in the three branches of S'' chosen as grafts and the corresponding branches of S determine the number of edges to be removed to get S', and the correct edges to be removed are uniquely determined by the linear orders $\preceq_v$.

We conclude that for every k-vertex tree S'', there are at most

$$144 \cdot \binom{8}{3} \cdot 6 \leqslant 48\,384$$

stubs S' such that S' is associated with an embedding of S that the corresponding set $\mathcal{S}$ contains S''; the estimate follows from the fact that there are at most 144 choices of $v_{S\to T}$, each of which leads to at most $\binom{8}{3}$ choices of grafts and at most six ways in which the grafts can correspond to the branches of S rooted at v_S.

The bound on the density of S in T is obtained as follows. Assign a charge of $48\,384 \cdot 12 = 580\,608$ to each embedding of a k-vertex tree S'' in T that is not isomorphic to S. Each such embedding sends 12 units of charge to each of the at most 48 384 stubs S' associated with an embedding of S whose corresponding set $\mathcal{S}$ contains S''. In this way, every stub S' receives at least $s_A s_B s_C$ units of charge, where s_A, s_B and s_C are defined as above (note that the quantities s_A, s_B and s_C are uniquely determined by the stub S'). Finally, the stub S' sends one unit of charge to each embedding of S in T whose associated stub is S'. Since every embedding of S in T receives at least one unit of charge, the density of S in T is at most $1 - 580\,609^{-1} \leqslant 1 - 10^{-7}$. □

4.4 Inducibility of trees with forks

In this section, we analyse the inducibility of non-caterpillar trees that are not covered by Theorem 4.9. A similar argument applies to a large class of caterpillars and so we formulate a single theorem to cover all cases.

Theorem 4.10. *Let S be a k-vertex tree ($k \geqslant 17$) that has a fixed vertex v_S satisfying one of the following:*

- *S is not a caterpillar, and v_S is a hub of S that is the root of at least one fork and is adjacent to at least two leaves,*
- *S is a caterpillar with at least four internal vertices, and v_S is the root of a fork of order at least two and is adjacent to a leaf, or*
- *S is a caterpillar with exactly three internal vertices, and v_S is the root of a fork and is adjacent to a leaf.*

If T is a tree with radius at least $4k$, then $d(S,T) \leqslant 1 - 10^{-4}$.

Proof. The assumptions guarantee that v_S is the root of at least two non-trivial branches, and that there is at most one vertex $v'_S \neq v_S$ of S such that S has an automorphism mapping the vertex v_S to v'_S. Let ℓ be the maximum order of a fork rooted at v_S; since v_S is the root of a fork, $\ell > 0$ is well-defined.

Notation. Fix a tree T with radius at least $4k$. Let v_T be a vertex of T such that there exist $(2k+1)$-vertex paths P_1 and P_2 starting at v_T that are disjoint except at v_T itself; such a choice is possible because the radius of T is at least $4k$. We show that the density of S in T is at most $1-10^{-4}$ using a discharging argument that assigns each embedding of a k-vertex tree non-isomorphic to S a charge of 9 999 units and redistributes this charge to embeddings of S so that each one receives at least one unit of charge.

Consider an embedding of S in T and let $v_{S\to T}$ be the vertex of T corresponding to v_S; if there are two valid choices, choose $v_{S\to T}$ arbitrarily among them. Let R_0 be the set of leaves of the embedding of S adjacent to $v_{S\to T}$ and let R_i be the neighbours of $v_{S\to T}$ that are contained in a fork of order i for $i \in \{1,\dots,\ell\}$. Note that $R_0 \neq \varnothing$ and $R_\ell \neq \varnothing$. In addition, observe that $\ell > 1$ or $|R_0| > 1$; in the last case described in the statement of the lemma, this is because $|S| \geqslant 17$. Set $R = R_0 \cup \dots \cup R_\ell$. Let α be the number of edges of T incident with $v_{S\to T}$ that are not contained in S, and for a vertex $v \in R$, let β_v be the number of edges incident with v that are not contained in S. Finally, define Q to be the unique path in T between $v_{S\to T}$ and v_T prolonged by P_1 if P_1 does not contain the edge from v_T towards $v_{S\to T}$ and prolonged by P_2 otherwise. Since P_1 and P_2 each have $2k+1$ vertices, Q has at least $k+1$ vertices not contained in the embedding of S.

Definition of correspondence. We next define sets $\mathcal{S}_A$, $\mathcal{S}_B$ and $\mathcal{S}_C$ of embeddings of trees non-isomorphic to S, and in some cases, we also define a set $\mathcal{S}_D$. Each of the embeddings contained in $\mathcal{S}_A$, $\mathcal{S}_B$ and $\mathcal{S}_C$ can be obtained from the embedding of S by moving an edge, and some of these sets can be empty.

Let $\mathcal{S}_A$ be the set of embeddings obtained by removing a leaf adjacent to $v_{S\to T}$ and adding a leaf to a fork of order ℓ rooted at $v_{S\to T}$. Note that the number of leaves of S that are adjacent to a vertex that is the root of exactly one non-trivial branch (that is, the number of leaves contained in a fork) is one fewer than the number of such leaves in the obtained embedding. Hence, the trees in $\mathcal{S}_A$ are not isomorphic to S. Observe that

$$|\mathcal{S}_A| = |R_0| \cdot \sum_{v\in R_\ell} \beta_v,$$

and let

$$\varepsilon_A = \frac{|\mathcal{S}_A|}{(\ell+1)(\alpha+1)} = \frac{|R_0| \cdot \sum_{v\in R_\ell} \beta_v}{(\ell+1)(\alpha+1)}.$$

Let $\mathcal{S}_B$ be the set of embeddings obtained by removing a leaf adjacent to $v_{S\to T}$ and adding a leaf to a fork of order $\ell-1$ if $\ell \geqslant 2$ or adding a leaf to another leaf adjacent to $v_{S\to T}$ if $\ell = 1$. Since the number of leaves of S that are adjacent to a vertex that is the root of exactly one non-trivial branch is one fewer than the number of such leaves in the obtained embedding, the trees in $\mathcal{S}_B$ are not isomorphic to S. Observe that if $\ell \neq 1$, then

$$|\mathcal{S}_B| = |R_0| \cdot \sum_{w \in R_{\ell-1}} \beta_w,$$

and if $\ell = 1$, then

$$|\mathcal{S}_B| = (|R_0| - 1) \cdot \sum_{w \in R_{\ell-1}} \beta_w.$$

Finally, let

$$\varepsilon_B = \frac{|\mathcal{S}_B|}{(|R_\ell| + 1)\ell(\alpha + 1)} \geqslant \frac{|R_0| \cdot \sum_{w \in R_{\ell-1}} \beta_w}{2(|R_\ell| + 1)\ell(\alpha + 1)};$$

the inequality holds since $\ell > 1$ or $|R_0| > 1$.

Next, let $\mathcal{S}_C$ be the set of embeddings obtained by removing a leaf of a fork of order ℓ rooted at $v_{S \to T}$ and adding a leaf adjacent to $v_{S \to T}$. Unless $\ell = 1$, the number of leaves of S that are adjacent to a vertex that is the root of exactly one non-trivial branch is one more than the number of such leaves in the obtained embedding. If $\ell = 1$, then the number of leaves of S is one less than the number of leaves in the obtained embedding. In both cases, the trees contained in $\mathcal{S}_C$ are non-isomorphic to S. Observe that

$$|\mathcal{S}_C| = |R_\ell| \cdot \ell \cdot \alpha,$$

and let

$$\varepsilon_C = \sum_{v \in R_\ell} \frac{\ell \cdot \alpha}{(|R_0| + 1)\left(\beta_v + 1 + \sum_{w \in R_{\ell-1}} \beta_w\right)}.$$

If $\ell \geqslant 2$, then we also define $\mathcal{S}_D$ to be the set of embeddings obtained by removing a leaf adjacent to $v_{S \to T}$ and a leaf of a fork of order ℓ rooted at $v_{S \to T}$, and adding the first two edges of Q not contained in the embedded tree. Since the number of leaves of S is at least one more than the number of leaves in the obtained embedding, the obtained embedding is not an embedding of S. Unlike in the previous three cases, the embeddings obtained in this way need not all be embeddings of the same tree since one of the removed edges can be contained in Q and then added back. Observe that

$$|\mathcal{S}_D| = |R_0| \cdot |R_\ell| \cdot \ell,$$

and let

$$\varepsilon_D = \sum_{v \in R_\ell} \frac{|R_0| \cdot \ell}{(\alpha + 1)\left(\beta_v + 1 + \sum_{w \in R_{\ell-1}} \beta_w\right)}.$$

Discharging argument. Given an embedding S' of a k-vertex tree in T, the number of choices of a vertex $v_{S\to T}$ in S' such that S' is contained in one of the sets $\mathcal{S}_A$, $\mathcal{S}_B$, $\mathcal{S}_C$, and $\mathcal{S}_D$ for an embedding of S with the vertex v_S mapped to $v_{S\to T}$ is at most 144 by Proposition 4.8 if S is not a caterpillar. If S is a caterpillar, then the number of choices of a vertex $v_{S\to T}$ in S' such that S' is contained in $\mathcal{S}_A$, $\mathcal{S}_B$, or $\mathcal{S}_C$ for an embedding of S with the vertex v_S mapped to $v_{S\to T}$ is at most 4: if S' is a caterpillar then $v_{S\to T}$ must be its first, second, second-last, or last internal vertex, and if S' is not a caterpillar, then it can only be contained in $\mathcal{S}_B$ and $v_{S\to T}$ is its unique vertex with two forks. Furthermore, the number of choices of a vertex $v_{S\to T}$ in S' such that S' is contained in $\mathcal{S}_D$ for an embedding of S with the vertex v_S mapped to $v_{S\to T}$ is at most 4: there are at most two choices of edges that could have been added as part of Q and, once these edges are chosen and removed, $v_{S\to T}$ is either its second or second-last internal vertex (assuming the tree is a caterpillar). We conclude that if S is a caterpillar, then the number of choices of a vertex $v_{S\to T}$ in S' such that S' is contained in one of the sets $\mathcal{S}_A$, $\mathcal{S}_B$, $\mathcal{S}_C$ and $\mathcal{S}_D$ for an embedding of S with the vertex v_S mapped to $v_{S\to T}$ is at most 8.

For each choice of $v_{S\to T}$, the embedding S' distributes 10 units of its charge equally to the embeddings of S with $v_{S\to T}$ such that S' is in the corresponding set $\mathcal{S}_A$, 10 units of its charge equally to the embeddings such that S' is in the corresponding set $\mathcal{S}_B$, 10 units of its charge equally to the embeddings such that S' is in the corresponding set $\mathcal{S}_C$, and, if $\ell \geqslant 2$, an additional 10 units of its charge equally to the embeddings such that S' is in the corresponding set $\mathcal{S}_D$. In this way, the embedding S' distributes at most $8 \cdot 40 = 320$ charge if S is a caterpillar, and at most $144 \cdot 40 = 5\,760$ units of charge if it is not. We remark that there will be additional charge distributed by S' by rules described later in the proof. Each embedding of S receives at least $10(\varepsilon_A + \varepsilon_B + \varepsilon_C)$ units of charge and, if $\ell \geqslant 2$, at least $10(\varepsilon_A + \varepsilon_B + \varepsilon_C + \varepsilon_D)$ units of charge. In particular, the considered embedding receives at least one unit of charge unless ε_A, ε_B, ε_C, and ε_D are all less than $1/10$.

We next show that one of ε_A, ε_B and ε_C is at least $1/10$ unless $\alpha = 0$ or $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$. Suppose that $\alpha \neq 0$. Let B be the maximum value of β_v for $v \in R_\ell$. If $B > \sum_{w\in R_{\ell-1}} \beta_w$, then

$$\begin{aligned}\varepsilon_A\varepsilon_C &= \frac{|R_0| \cdot \sum_{v\in R_\ell} \beta_v}{(\ell+1)(\alpha+1)} \cdot \sum_{v\in R_\ell} \frac{\ell \cdot \alpha}{(|R_0|+1)\left(\beta_v + 1 + \sum_{w\in R_{\ell-1}} \beta_w\right)} \\ &\geqslant \frac{|R_0| \cdot \sum_{v\in R_\ell} \beta_v}{4\ell\alpha} \cdot \frac{|R_\ell| \cdot \ell \cdot \alpha}{2(|R_0|+1)B} \geqslant \frac{|R_0|}{8(|R_0|+1)} \geqslant \frac{1}{16}.\end{aligned}$$

Hence, ε_A or ε_C is at least 1/10. If $B \leqslant \sum_{w\in R_{\ell-1}} \beta_w$ and $\sum_{w\in R_{\ell-1}} \beta_w \neq 0$, then

$$\begin{aligned}\varepsilon_B\varepsilon_C &\geqslant \frac{|R_0| \cdot \sum_{w\in R_{\ell-1}} \beta_w}{2(|R_\ell|+1)\ell(\alpha+1)} \cdot \sum_{v\in R_\ell} \frac{\ell\cdot\alpha}{(|R_0|+1)\left(\beta_v + 1 + \sum_{w\in R_{\ell-1}} \beta_w\right)} \\ &\geqslant \frac{|R_0| \cdot \sum_{w\in R_{\ell-1}} \beta_w}{8|R_\ell|\ell\alpha} \cdot \sum_{v\in R_\ell} \frac{\ell\cdot\alpha}{3(|R_0|+1)\sum_{w\in R_{\ell-1}} \beta_w} \\ &= \frac{|R_0|}{24(|R_0|+1)} \geqslant \frac{1}{48}.\end{aligned}$$

Hence, ε_B or ε_C is at least 1/10. We conclude that if $\alpha \neq 0$, then one of ε_A, ε_B and ε_C is at least 1/10 unless $\sum_{w\in R_{\ell-1}} \beta_w = 0$ and $\sum_{v\in R_\ell} \beta_v = 0$. So, we need to analyse the cases when $\alpha = 0$ or when $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$.

Analysis of non-caterpillars. Suppose that S is not a caterpillar and $\alpha = 0$. Let S' be obtained from the embedding of S by removing any two leaves adjacent to $v_{S\to T}$ and adding the first two edges on Q not contained in the embedding. Since S' has at least one less leaf than S, it is not isomorphic to S. The embedding S' sends one unit of charge to the considered embedding of S. Note that the embedding S' sends by this rule at most $2 \cdot 144$ units of charge in addition to the charge sent earlier: when S' is fixed, there are at most two choices of edges that could have been added as part of Q, and at most 144 choices of $v_{S\to T}$ by Proposition 4.8. The leaves adjacent to $v_{S\to T}$ that were removed are uniquely determined since $\alpha = 0$.

Suppose that S is not a caterpillar, $\alpha > 0$ and $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$. If $\ell \geqslant 2$, then

$$\varepsilon_C = \frac{|R_\ell|\ell\alpha}{|R_0|+1} \qquad \text{and} \qquad \varepsilon_D = \frac{|R_\ell|\ell|R_0|}{\alpha+1}.$$

It follows that $\varepsilon_C\varepsilon_D \geqslant 1/4$; that is, ε_C or ε_D is at least 1/10. If $\ell = 1$ and $|R_1| \geqslant 2$, then let S' be obtained from the embedding of S by removing a leaf from two forks rooted at $v_{S\to T}$ and adding the first two edges of Q. Since the embedding S' has fewer leaves contained in forks, S' is not isomorphic to S. The embedding S' sends one unit of charge to the considered embedding of S. Each embedding S' sends in this way at most $2 \cdot 144$ units of charge in addition to the charge sent earlier: when S' is fixed, there are at most two choices of edges that could have been added as part of Q, and at most 144 choices of $v_{S\to T}$ by Proposition 4.8. The leaves adjacent to $v_{S\to T}$ to be changed to a fork are uniquely determined and so are the edges to be added since $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$.

If $\ell = |R_1| = 1$, then let e be the edge incident with the leaf of the fork rooted at $v_{S\to T}$. If $v_{S\to T}$ has a neighbour w in T that is not contained in S and that has degree at least two in T, then remove the edge e and add the edge $v_{S\to T}w$

to obtain an embedding S'. The embedding S' has more leaves than S and so is not isomorphic to S. The embedding S' sends one unit of charge to the considered embedding of S. Each embedding S' sends in this way at most $2 \cdot 144$ units of charge in addition to the charge sent earlier as there are at most two leaves adjacent to $v_{S\to T}$ that can be changed to a fork with the unique edges to be added (as $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$). Hence, we can assume that all neighbours of $v_{S\to T}$ in T are leaves except its neighbours that are contained in the non-trivial branches of S.

If e is not contained in Q, then let S' be the embedding obtained from S by removing e and adding the first edge of Q not contained in S, and let S'' be the embedding obtained from S by removing the fork containing e and adding the first two edges of Q not contained in S. Observe that S' or S'' is not isomorphic to S since at least one of them has a different number of leaves from S. The embedding that is not isomorphic sends one unit of charge to S and each embedding sends at most $4 \cdot 144$ units of charge in this way (it can appear in the role of S' and S'', there are at most two choices of edges that could have been added as part of Q, and there are at most 144 choices of $v_{S\to T}$, each determining the embedding S uniquely).

If e is contained in Q, then let $\mathcal{S}_1$ be the set of embeddings obtained by removing a leaf adjacent to $v_{S\to T}$ and adding the edge of Q following the edge e; note that $|\mathcal{S}_1| = |R_0|$. Let $\mathcal{S}_2$ be the set of embeddings obtained by removing the edge e and adding an edge incident with $v_{S\to T}$ not contained in S; note that $|\mathcal{S}_2| = \alpha$. Since the embeddings in $\mathcal{S}_1$ and $\mathcal{S}_2$ have different numbers of leaves than S, they are not isomorphic to S. Each embedding in $\mathcal{S}_1$ distributes one unit of charge equally among all $\alpha + 1$ embeddings of S that can be obtained in this way, and each embedding in $\mathcal{S}_2$ distributes one unit of charge equally among all $|R_0| + 1$ embeddings of S that can be obtained in this way (note that $v_{S\to T}$ is uniquely determined as the vertex in the embedding with degree greater than 2 that is closest to the added edges of Q, and the fork of an embedding of S is created only by adding an edge to a leaf adjacent to $v_{S\to T}$ whose degree in T is 2). We conclude that the embedding of S receives at least

$$\frac{|R_0|}{\alpha+1} + \frac{\alpha}{|R_0|+1} \geqslant \frac{1}{2}\left(\frac{|R_0|}{\alpha} + \frac{\alpha}{|R_0|}\right) \geqslant 1$$

units of charge.

Analysis of caterpillars. We next analyse the case when S is a caterpillar. If $\alpha = 0$ and $|R_0| \geqslant 2$, then let S' be obtained from the embedding of S by removing any two leaves adjacent to $v_{S\to T}$ and adding the first two edges on Q not contained in the embedding. Since S' has fewer leaves than S, S' is not isomorphic to S. The embedding S' sends one unit of charge to the considered embedding of S. Note that in this way the embedding S' sends at most $2 \cdot 2$ units of charge in addition to the charge sent earlier: there are at most two choices of edges that could have been added as part of Q and, once these edges are chosen and removed, the vertex $v_{S\to T}$ is either the second or second-last internal vertex of the resulting caterpillar.

If $\alpha = 0$, $|R_0| = 1$ and $\ell \geqslant 2$, then we derive along the lines used in the general case that $\varepsilon_A\varepsilon_D \geqslant \frac{|R_0|^2}{4(\alpha+1)^2} = \frac{1}{4}$ or $\varepsilon_B\varepsilon_D \geqslant \frac{|R_0|^2}{12(\alpha+1)^2} = \frac{1}{12}$ unless $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$. However, if $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$, then $\varepsilon_D = \frac{|R_0|\cdot|R_\ell|\cdot\ell}{\alpha+1} \geqslant 2$.

If $\alpha = 0$ and $|R_0| = \ell = 1$, then $\varepsilon_A \geqslant 1/2$ unless $\sum_{v\in R_\ell} \beta_v = 0$. If $\alpha = 0$, $|R_0| = \ell = 1$, and $\sum_{v\in R_\ell} \beta_v = 0$, and more generally whenever $\ell = 1$ and $\sum_{v\in R_\ell} \beta_v = 0$, then we are in the third case from the statement of the lemma, and $|R_\ell| = 2$. In other words, S is a star with two different edges subdivided. Consider S' obtained from the embedding of S by removing the edges that are incident with the leaves of the two forks of S and adding the first two edges on Q not contained in the embedding. Observe that S' is not isomorphic to S. The embedding S' sends one unit of charge to the considered embedding of S. Note that in this way the embedding S' sends at most one unit of charge in addition to the charge sent earlier; the edges of S' that were added as a part of Q are the unique edges whose removal creates a star, the vertex $v_{S\to T}$ is the internal vertex of this star, and the remaining two edges of the embedding of S are uniquely determined since $\sum_{v\in R_\ell} \beta_v = 0$.

The final case to consider is when $\alpha > 0$, $\sum_{w\in R_{\ell-1}} \beta_w = \sum_{v\in R_\ell} \beta_v = 0$, and $\ell \geqslant 2$ (note that the case $\ell = 1$ is covered in the previous paragraph). As in the non-caterpillar case, it follows that $\varepsilon_C\varepsilon_D \geqslant \ell^2|R_\ell|^2/4 \geqslant 1/4$; that is, ε_C or ε_D is at least $1/10$.

Conclusion. According to the rules set above, each embedding of a tree non-isomorphic to S distributes at most $320 + 4 + 1 = 325$ units of charge if S is a caterpillar, and at most $5\,760 + 10 \cdot 144 + 2 = 7\,202$ units of charge if it is not. Thus, the density of S in T is at most $1 - 7203^{-1} \leqslant 1 - 10^{-4}$. □

4.5 Inducibility of caterpillars

In this section, we complete the analysis of the inducibility of caterpillars. We start with caterpillars whose second or second-last internal vertex is the root of a fork of order 1 and is adjacent to a leaf.

Lemma 4.11. *Let S be a non-path caterpillar with $k \geqslant 10$ vertices that has at least four internal vertices and has a fixed vertex v_S that is the root of a fork of order 1 and is adjacent to a leaf. If T is a tree with radius at least $4k$, then $d(S,T) \leqslant 1 - 10^{-3}$.*

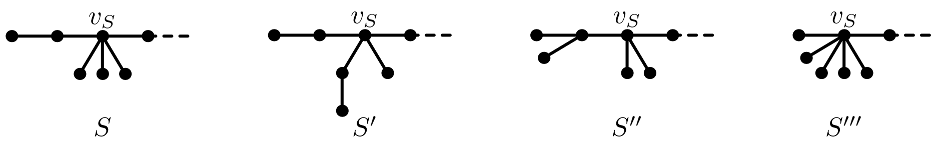


Figure 4.2: The trees S, S', S'' and S''' used in the proof of Lemma 4.11 when $\ell = 3$.

Proof. Let $\ell > 0$ be the number of leaves adjacent to v_S. Since S is a caterpillar with at least four internal vertices, v_S is the root of exactly one fork; the order of this fork is 1 by the assumption of the lemma.

Let v_T be a vertex of T such that there exist $(2k+1)$-vertex paths P_1 and P_2 starting at v_T that are disjoint except at v_T itself; such a choice is possible because the radius of T is at least $4k$.

We next define trees S_0, S', S'' and S''' (see Figure 4.2). Let S_0 be the tree obtained from S by removing the fork and all leaves adjacent to v_S. If $\ell \geqslant 2$, then let S' be the tree obtained from S by removing a leaf adjacent to v_S and turning another leaf adjacent to v_S into a fork of order 1; if $\ell = 1$, then S' is not defined. Let S'' be the tree obtained from S by removing a leaf adjacent to v_S and adding a leaf to the fork rooted at v_S. Finally, let S''' be the tree obtained from S be removing the leaf of the fork rooted at v_S and adding a leaf adjacent to v_S. The trees S, S' (if defined), S'', and S''' are mutually non-isomorphic, since S' is the only one that is not a caterpillar, S''' has fewer internal vertices than S and S'', and the numbers of leaves adjacent to the first and last internal vertices of S and S'' differ.

In this proof, a *stub* is an embedding of S_0 with a distinguished vertex $v_{S\to T}$ such that the embedding of S_0 can be extended to an embedding of S with $v_{S\to T}$ corresponding to v_S. Let Q be the unique path in T between the vertices $v_{S\to T}$ and v_T prolonged by P_1 if P_1 does not contain the edge from v_T towards $v_{S\to T}$ and prolonged by P_2 otherwise. Let $D \geqslant \ell + 1$ be the degree of $v_{S\to T}$ in T minus 1 and let $d_1, \dots, d_D$ be the degrees of its neighbours not contained in the embedding of S_0 minus 1. The numbers of ways that the embedding of S_0 can be extended (with $v_{S\to T}$ corresponding to v_S) to an embedding of S, S', S'', and S''' are

$$\sum_{1\leqslant i\leqslant D} d_i \binom{D-1}{\ell}, \sum_{1\leqslant i<j\leqslant D} d_i d_j \binom{D-2}{\ell-2}, \sum_{1\leqslant i\leqslant D} \binom{d_i}{2}\binom{D-1}{\ell-1}, \text{ and } \binom{D}{\ell+2},$$

respectively. Let N, N', N'' and N''' be these numbers. We claim that $N \leqslant 54(N'+N''+N'''+N_Q)$, where N_Q is the number of extensions of S_0 to an embedding of a tree S_Q that is defined later. Note that $N_Q > 0$ only if $\sum_{1\leqslant i\leqslant D} d_i = 1$ or

$\ell = d_1 = d_2 = 1$; we set $N_Q = 0$ otherwise.

The following three paragraphs concern the case when $\sum_{1\leqslant i\leqslant D} d_i = 1$; note that $N' = N'' = 0$ in this case. Fix an embedding of S and denote the internal vertices of the embedding by $v_1, \dots, v_m$ so that $v_2 = v_S$. Consider the first edge of Q not contained in this embedding. If this edge is incident with v_m, then let S_Q be the embedding obtained from S by removing one of the leaves adjacent to $v_{S\to T}$ and adding the first unused edge of Q. The sum of the degrees of the first and last internal vertices of S_Q is greater than that of S, so S_Q is not isomorphic to S. Furthermore, the number N_Q of ways that the embedding of S_0 can be extended to an embedding of S_Q (with the added edge of Q fixed) is $\binom{D-1}{\ell-1}$.

Now suppose that the first edge of Q not contained in the embedding of S is incident with a vertex other than v_m; note that this edge is not incident with v_2 because $\sum_{1\leqslant i\leqslant D} d_i = 1$. Let S_Q be the embedding obtained from S by removing the leaf of the fork rooted at $v_{S\to T}$ and one of the leaves adjacent to $v_{S\to T}$, and then adding the first two unused edges of Q. The resulting embedding is not isomorphic to S: if the first unused edge is incident with one of $v_3, \dots, v_{m-1}$ or their adjacent leaves, then S_Q is not a caterpillar, and if the first unused edge is incident with a leaf of v_1 or v_m, then S_Q has more internal vertices than S. Again, the number N_Q of ways that the embedding of S_0 can be extended to an embedding of S_Q (with the edge v_1v_2 and the added edges of Q fixed) is $\binom{D-1}{\ell-1}$.

It follows that

$$
\begin{aligned}
N &= \binom{D-1}{\ell} = \binom{D-1}{\ell-1}\frac{D-\ell}{\ell},\\
N''' &= \binom{D}{\ell+2} = \binom{D-1}{\ell-1}\frac{D(D-\ell)(D-\ell-1)}{(\ell+2)(\ell+1)\ell}, \text{ and}\\
N_Q &= \binom{D-1}{\ell-1}.
\end{aligned}
$$

If $D \leqslant 2\ell$, then $N \leqslant N_Q$. If $D \geqslant 2\ell+1$, then $D-\ell-1 \geqslant \ell \geqslant \frac{1}{2}(\ell+1)$ and $D \geqslant \ell+2$, so $N \leqslant 2N'''$. Therefore, regardless of the relationship between D and ℓ, we have $N \leqslant 2(N''' + N_Q)$.

In the rest of the proof, we analyse the case when $\sum_{1\leqslant i\leqslant D} d_i \geqslant 2$. If $D = \ell + 1$, then the numbers of ways that the embedding of S_0 can be extended to an embedding of S, S' and S'' (note that $N''' = 0$) are

$$
N = \sum_{1\leqslant i\leqslant D} d_i, \quad N' = \sum_{1\leqslant i<j\leqslant D} (\ell-1)d_i d_j, \text{ and } N'' = \sum_{1\leqslant i\leqslant D} \ell\binom{d_i}{2},
$$

respectively. If $\ell \geqslant 2$, then $N \leqslant 2N'$ unless only one of the d_i is non-zero, in which case $N \leqslant N''$. If $\ell = 1$, then $N = d_1 + d_2$, $N' = 0$ and $N'' = \binom{d_1}{2} + \binom{d_2}{2}$, in which case $N \leqslant 2N'' + 2 \leqslant 4N''$ unless $d_1 = d_2 = 1$. Finally, if $\ell = d_1 = d_2 = 1$, then consider the embedding S_Q defined in the same way as in the case $\sum_{1 \leqslant i \leqslant D} d_i = 1$ unless the first edge of Q not contained in the embedding of S is incident with the leaf adjacent to $v_{S \to T}$; note that the embedding S_Q is well-defined as the first edge of Q not contained in the embedding of S cannot be incident with v_2 as $D = \ell + 1$. If the first edge of Q not contained in the embedding of S is incident with the leaf adjacent to $v_{S \to T}$, then remove the fork rooted at $v_{S \to T}$ and add the first two edges of Q not contained in S; the resulting embedding S_Q is a caterpillar with diameter greater than that of S and so is non-isomorphic to S (note that the embedding S_Q is the same for both embeddings of S that can be obtained from the same stub). In all cases describe above, the embedding S_Q is uniquely determined by the embedding of the stub, so $N_Q = 1$, which implies that $N = 2 \leqslant 2N_Q$.

Next assume that $D \geqslant \ell + 2$. If $\ell \geqslant 2$, then

$$\begin{aligned}
N &\leqslant 2\binom{D-2}{\ell-2} \sum_{1 \leqslant i \leqslant D} d_i \frac{D(D-\ell)}{\ell^2}, \\
N' &\geqslant \binom{D-2}{\ell-2} \sum_{1 \leqslant i < j \leqslant D} d_i d_j, \\
N'' &\geqslant \binom{D-2}{\ell-2} \sum_{1 \leqslant i \leqslant D} \binom{d_i}{2}, \text{ and} \\
N''' &\geqslant \frac{1}{8}\binom{D-2}{\ell-2} \frac{D^2(D-\ell)^2}{\ell^4}
\end{aligned}$$

Hence,

$$N' + N'' \geqslant \frac{1}{4}\binom{D-2}{\ell-2} \left(\sum_{1 \leqslant i \leqslant D} d_i \right)^2,$$

which implies that $N \leqslant 8(N' + N'' + N''')$ by the AM-GM inequality.

If $\ell = 1$, then

$$\begin{aligned}
N &\leqslant \sum_{1 \leqslant i \leqslant D} D d_i \leqslant D^2 + \sum_{\substack{1 \leqslant i \leqslant D \\ d_i \geqslant 2}} D(d_i - 1) \leqslant D^2 + D^3 + \sum_{\substack{1 \leqslant i \leqslant D \\ d_i \geqslant 2}} (d_i - 1)^2, \\
N'' &= \sum_{1 \leqslant i \leqslant D} \binom{d_i}{2} \geqslant \sum_{\substack{1 \leqslant i \leqslant D \\ d_i \geqslant 2}} \frac{(d_i - 1)^2}{2}, \text{ and}
\end{aligned}$$

$$N''' = \binom{D}{3} \geqslant \frac{D^3}{27}.$$

It follows that

$$N \leqslant D^2 + D^3 + \sum_{\substack{1 \leqslant i \leqslant D \\ d_i \geqslant 2}} (d_i - 1)^2 \leqslant 54(N'' + N''').$$

In all cases, we have proved that $N \leqslant 54(N' + N'' + N''' + N_Q)$. Each embedding of a k-vertex tree non-isomorphic to S sends 72 charge to each stub that it extends. In this way, each stub receives at least $54(N' + N'' + N''' + N_Q) \geqslant N$ charge, which it can then distribute to its N extensions into embeddings of S.

To complete the discharging argument, it remains to bound the total amount of charge that each embedding of a k-vertex tree non-isomorphic to S sends. If the embedding is isomorphic to S', then $v_{S \to T}$ is the unique vertex that is the root of two forks. If the embedding is isomorphic to S'', then $v_{S \to T}$ is either the second or second-last internal vertex. If the embedding is isomorphic to S''', then $v_{S \to T}$ is either the first or last internal vertex. Finally, if the embedding is isomorphic to S_Q, then there are at most two choices of edges that could have been added as part of Q, and once these edges are chosen and removed, $v_{S \to T}$ is either the first, second, second-last, or last internal vertex of the resulting caterpillar. When the edge(s) added from Q are removed from the embedding and $v_{S \to T}$ is chosen, the edges of the stub can be recovered by removing the forks and leaves rooted at $v_{S \to T}$ from the embedding. Hence, the embedding sends at most $54 \cdot (\max\{1, 2, 2\} + 2 \cdot 4) \leqslant 999$ charge, and the density of S in T is at most $1 - 10^{-3}$. ☐

The next lemma deals with caterpillars S that are not covered by Theorem 4.10 and Lemma 4.11.

Lemma 4.12. *Let S be a caterpillar with $k \geqslant 10$ vertices that is not a path such that the path $v_1, \ldots, v_m$ formed by its internal vertices satisfies either $m = 2$, or $m \geqslant 3$ and the degrees of v_2 and v_{m-1} equal 2. If T is a tree with radius at least $4k$, then $d(S, T) \leqslant 1 - 10^{-3}$.*

Proof. Let $\alpha > 0$ and $\beta > 0$ be the number of leaves adjacent to v_1 and v_m respectively. By symmetry, we can assume that $\alpha \leqslant \beta$. Let S' be the caterpillar obtained from S by removing the α leaves adjacent to v_1 and β leaves adjacent to v_m.

Let v_T be a vertex of T such that there exist $(2k+1)$-vertex paths P_1 and P_2 starting at v_T that are disjoint except at v_T itself; such a choice is possible because the radius of T is at least $4k$.

In this proof, a *stub* is an embedding of S' in T together with a choice of orientation for the longest path in the embedding. (The length of this path is the same as the distance between v_1 and v_m.) Given a stub, let $v'_1, \ldots, v'_m$ be the vertices of the longest path in the embedding, ordered according to the chosen orientation, and let A and B be the degrees of v'_1 and v'_m minus 1. Let Q be the unique path in T between the vertices $v_{S\to T}$ and v_T prolonged by P_1 if P_1 does not contain the edge from v_T towards $v_{S\to T}$ and prolonged by P_2 otherwise.

We analyse the case when $\alpha = \beta = 1$ separately at the end of the proof, so for now suppose that $\beta \geqslant 2$. The number of ways the embedding of S' can be extended to an embedding of S with each v'_i corresponding to v_i is $\binom{A}{\alpha}\binom{B}{\beta}$. We will associate to each embedding of S' in T a set of N embeddings of k-vertex trees non-isomorphic to S so that

$$\binom{A}{\alpha}\binom{B}{\beta} \leqslant 9N.$$

Note that if $A < \alpha$ or $B < \beta$, then there is nothing to prove, so we assume that $A \geqslant \alpha$ and $B \geqslant \beta$.

If $B = \beta$, then we consider extensions obtained from the embedding of S' by adding α leaves to v'_1, $\beta - 2$ leaves to v'_m and then the first two edges of Q not contained in the embedding. If the obtained embedding is a caterpillar, then the sum of the degrees of its first and last internal vertices is less than the sum of the degrees of the first and last internal vertices of S. Therefore, the $N = \binom{A}{a} = \binom{A}{\alpha}\binom{B}{\beta}$ obtained embeddings are not isomorphic to S.

If $A = \alpha \geqslant 2$, then consider extensions obtained from the embedding of S' by adding $\alpha - 2$ leaves to v'_1, β leaves to v'_m and then the first two edges of Q not contained in the embedding. Again, if the obtained embedding is a caterpillar, then the sum of the degrees of its first and last internal vertices is less than the sum of the degrees of the first and last internal vertices of S. Therefore, the obtained embeddings are not isomorphic to S, and their number is $N = \binom{B}{\beta}$, which is equal to $\binom{A}{\alpha}\binom{B}{\beta}$.

If $A = \alpha = 1$ but $2 \leqslant \beta < B$, then the number of extensions to S is $\binom{B}{\beta}$, and we consider extensions of S' obtained by either adding $\beta + 1$ leaves to v'_2 or adding $\beta - 1$ leaves to v'_2 and the first two edges of Q not contained in the embedding; the number N such extensions is

$$N = \binom{B}{\beta - 1} + \binom{B}{\beta + 1} \geqslant \binom{B}{\beta}.$$

Hence, we can assume that $A > \alpha$ and $B > \beta$ in the remainder of the analysis of the case $\beta \geqslant 2$. We first deal with the case when $\alpha \neq \beta - 1$. Consider extensions obtained from the embedding of S' by either adding $\alpha + 1$ leaves to v_1' and $\beta - 1$ leaves to v_m' or adding $\alpha - 1$ leaves to v_1' and $\beta + 1$ leaves to v_m'; the number of such extensions is

$$\begin{aligned}
N &= \binom{A}{\alpha+1}\binom{B}{\beta-1} + \binom{A}{\alpha-1}\binom{B}{\beta+1} \\
&= \binom{A}{\alpha-1}\binom{B}{\beta-1}\left(\frac{(A-\alpha+1)(A-\alpha)}{\alpha(\alpha+1)} + \frac{(B-\beta+1)(B-\beta)}{\beta(\beta+1)}\right) \\
&\geqslant \frac{1}{4}\binom{A}{\alpha-1}\binom{B}{\beta-1}\left(\frac{(A-\alpha+1)^2}{\alpha^2} + \frac{(B-\beta+1)^2}{\beta^2}\right) \\
&\geqslant \frac{1}{2}\binom{A}{\alpha-1}\binom{B}{\beta-1}\frac{A-\alpha+1}{\alpha}\frac{B-\beta+1}{\beta} = \frac{1}{2}\binom{A}{\alpha}\binom{B}{\beta}.
\end{aligned}$$

It remains to analyse the case $\alpha = \beta - 1$. Consider extensions obtained from the embedding of S' by either adding $\alpha + 2$ leaves to v_1' and $\beta - 2$ leaves to v_m' or adding $\alpha - 1$ leaves to v_1' and $\beta + 1$ leaves to v_m'; the number of such extensions is

$$N = \binom{A}{\alpha+2}\binom{B}{\beta-2} + \binom{A}{\alpha-1}\binom{B}{\beta+1},$$

unless $A = \alpha + 1$. We next argue that $\binom{A}{\alpha}\binom{B}{\beta} \leqslant 9N$. If $\binom{A}{\alpha-1}\binom{B}{\beta+1} \leqslant \frac{1}{9}\binom{A}{\alpha}\binom{B}{\beta}$, then $9 \leqslant \frac{(\beta+1)(A-\alpha+1)}{\alpha(B-\beta)}$ and

$$\begin{aligned}
\binom{A}{\alpha+2}\binom{B}{\beta-2} &= \frac{\beta(\beta-1)(A-\alpha)(A-\alpha-1)}{(\alpha+2)(\alpha+1)(B-\beta+2)(B-\beta+1)}\binom{A}{\alpha}\binom{B}{\beta} \\
&\geqslant \frac{2^2}{3^4 \cdot 6^2}\left(\frac{(\beta+1)(A-\alpha+1)}{\alpha(B-\beta)}\right)^2\binom{A}{\alpha}\binom{B}{\beta} \\
&\geqslant \frac{9^2}{3^6}\binom{A}{\alpha}\binom{B}{\beta} = \frac{1}{9}\binom{A}{\alpha}\binom{B}{\beta},
\end{aligned}$$

so $N \geqslant \frac{1}{9}\binom{A}{\alpha}\binom{B}{\beta}$.

Finally, we deal with the case $A = \alpha + 1$. Consider extensions obtained from the embedding of S' by adding $\alpha - 1$ leaves to v_1', $\beta - 1$ leaves to v_m' and then the first two edges of Q not contained in the embedding, or adding $\alpha - 1$ leaves to v_1' and $\beta + 1$ leaves to v_m'; as before, these extensions are not isomorphic to S, and the number of such extensions is

$$\binom{A}{\alpha-1}\binom{B}{\beta-1}+\binom{A}{\alpha-1}\binom{B}{\beta+1}\geqslant\binom{A}{\alpha+1}\binom{B}{\beta-1}+\binom{A}{\alpha-1}\binom{B}{\beta+1},$$

which is at least $\binom{A}{\alpha}\binom{B}{\beta}/2$ as established in the case $\alpha \neq \beta - 1$.

We complete the case $\beta \geqslant 2$ by a discharging procedure similar to that used in the proof of Lemma 4.11. Each embedding of a k-vertex tree non-isomorphic to S sends a charge of 9 to each stub that it extends. In this way, each stub receives $9N \geqslant \binom{A}{\alpha}\binom{B}{\beta}$ charge, which is then distributed to the extensions of the stub into embeddings of S.

To finish the analysis of the discharging argument, it remains to bound the total amount of charge that each embedding of a k-vertex tree non-isomorphic to S sends. Fix an embedding of a k-vertex tree non-isomorphic to S and suppose that it can be obtained from a stub by at least one of the above processes. If this process *does not* involve adding two unused edges of Q, then the embedding is a caterpillar and the vertex that corresponds to v_1' is either the first or last internal vertex of the embedding, or the unique leaf of the first internal vertex if the first internal vertex has degree 2, or the unique leaf of the last internal vertex if the last internal vertex has degree 2. For each choice of v_1' there is at most one valid choice of v_m' at the correct distance in T, and the stub is then determined by removing the leaves of the embedding adjacent to v_1' and v_m'. Thus, the embedding sends at most $9 \cdot 4$ units of charge in this way. If the embedding can be obtained from a stub by a process that *does* involve adding two unused edges of Q, then there are at most two places where these edges could have been added. After choosing which of these sets of two edges to remove from the embedding, we follow the same procedure as above to obtain the possible stubs, so the embedding sends at most $9 \cdot 2 \cdot 4$ additional charge in this way. In total, each embedding of a k-vertex tree non-isomorphic to S sends at most $9 \cdot 3 \cdot 4 < 999$ units of charge, so the density of S in T is at most $1 - 10^{-3}$.

Finally, we return to the case $\alpha = \beta = 1$; the argument is again based on a discharging procedure. Since S is not a path, we have $m \geqslant 5$. Consider extensions of the embedding of S' obtained by either adding two leaves to the vertex v_1' or adding two leaves to the vertex v_m'. The number of such extensions is $\binom{A}{2} + \binom{B}{2} \geqslant \frac{1}{4}AB$ unless $A = B = 1$. If $A = B = 1$, then $\binom{A}{\alpha}\binom{B}{\beta} = 1$ and we consider the embedding obtained from S' by adding the first two edges of Q that are not contained in S'. If the first of these two edges attaches to a vertex of S' other than v_1', v_2', v_{m-1}', or v_m', then the obtained embedding is not a caterpillar. If it attaches to v_2' or v_{m-1}', then the second or second-last internal vertex of the obtained caterpillar has degree 3.

Thus the obtained embedding is not isomorphic to S unless the added edges attach to v'_1 or v'_m. By symmetry, we assume that they attach to v'_m. Any isomorphism between S and the obtained embedding maps the vertex v_{m-1} of S to the vertex v'_3 of the obtained embedding. In particular, if the two trees are isomorphic, then the degree of v'_3 in S' is 2, and we instead consider the embedding obtained from S' by adding a leaf to v'_2 and a leaf to v'_m, if such an embedding exists. If such an embedding does not exist, we instead consider the embedding obtained from S' by removing the edge $v'_1v'_2$ and adding the first three edges of Q. In all cases above from the analysis of the case $A = B = 1$, the obtaining embedding is non-isomorphic to S.

Thus, when each embedding of a k-vertex tree non-isomorphic to S sends 4 units of charge to each stub that it extends, each stub such that $A \neq 1$ or $B \neq 1$ receives at least AB units of charge, which it can then redistribute to its extensions into embeddings of S. We next estimate additional charge because of embeddings considered in the case $A = B = 1$; for this analysis, fix an embedding of a k-vertex tree non-isomorphic to S.

- If the embedding can be obtained from the process of adding the first two unused edges of Q, then there are at most two places where these edges could have been added. After choosing which of these two choices of two edges to remove, the vertices v'_1 and v'_m must correspond to the unique leaves of the first and last internal vertices, and the stub is determined once a choice for the correspondence is made.

- If the embedding can be obtained from the process of adding a leaf to v'_2 and v'_{m-1}, then either the first or last internal vertex of the embedding has two leaves, and this vertex corresponds to either v'_2 or v'_{m-1}. One of the two leaves then corresponds to either v'_1 or v'_m, and the stub is determined by choosing the correspondence, removing the other leaf, and removing the unique leaf of the terminal internal vertex at the other end of the caterpillar.

- If the embedding can be obtained from the process of removing the edge $v'_1v'_2$ and adding the first three unused edges of Q, there are at most two places where these edges could have been added. After choosing which of these two options of three edges to remove, v'_3 is the first or last internal vertex of the resulting caterpillar, v'_2 is the unique leaf of v'_3 in the caterpillar, and v'_1 is the unique neighbour of v'_2 in T that is not in the caterpillar.

Each embedding of a k-vertex tree non-isomorphic to S sends at most 12 additional units of charge and so it sends at most $16 < 99$ units of charge in total. We conclude that if $\alpha = \beta = 1$, then $d(S,T) \leqslant 1 - 10^{-2}$. □

The next theorem summarises our analysis of caterpillars.

Theorem 4.13. *Every caterpillar S with $|S| \geqslant 17$ that is neither a star nor a path has inducibility at most $1 - 10^{-4}$.*

Proof. The limit density of S in any sequence of trees with bounded radius is 0 by Proposition 4.4, so it remains to investigate the density of S in trees with unbounded radius. Since S is not a star, S has more than one internal vertex. The case where S has two internal vertices is covered by Lemma 4.12, the case where S has three internal vertices is covered by Lemma 4.12 and the third case of Theorem 4.10 (depending on whether the middle internal vertex has degree 2), and the case where S has four or more internal vertices is covered by Lemma 4.11, Lemma 4.12, and the second case of Theorem 4.10 (depending on the order of forks rooted at the second or second-last internal vertex of S and their degrees). □

4.6 Inducibility bounded away from 1

As detailed in the proof of Theorem 4.1 below, the results in the preceding sections show that the inducibility of every tree with at least 17 vertices is at most $1 - 10^{-8}$. On the other hand, the inducibility of every k-vertex tree X that is neither a path nor a star is less than 1, since for every convergent sequence of trees $(T_n)_{n\to\infty}$ with $|T_n| \to \infty$, if $\lim_{n\to\infty} d(X, T_n) = 1$, then $\lim_{n\to\infty} d(P_k, T_n) = \lim_{n\to\infty} d(S_k, T_n) = 0$, contradicting Proposition 4.5. In particular, trees with at most 16 vertices have inducibility bounded away from 1. These two results combined imply that the inducibility of every tree is at most $1 - \varepsilon$ for some fixed constant $\varepsilon > 0$. In the interest of obtaining an explicit value of ε, we provide a crude upper bound on the inducibility of small trees.

Lemma 4.14. *For $k \geqslant 5$, the inducibility of every k-vertex tree S that is neither a path nor a star is at most $1 - k^{-(2k-3)}$.*

Proof. Since the inducibility of S is defined with respect to $d(S,T)$ where $|T| \to \infty$, it suffices to show that $d(S,T) \leqslant 1 - k^{-(2k-3)}$ for every tree T with $|T| \geqslant k^k$. We prove the bound by a discharging argument. Every embedding of a k-vertex tree in T that is not isomorphic to S begins with one unit of charge and distributes the charge according to the following rules.

Every embedding S'' of a k-vertex tree that neither non-isomorphic to S nor a star distributes its charge equally among all embeddings S' of k-vertex trees that share an edge with S'' such that the maximum degree of a vertex of S' in T is at most k. Since there are at most $(k-1)!(k-1)^{k-2} \leqslant k^{2k-3}$ such embeddings S' for every embedding S'', each embedding S' of a k-vertex tree such that the maximum degree of a vertex of S' in T is at most k receives at least $k^{-(2k-3)}$ units of charge.

Every embedding S'' of a k-vertex star distributes its charge as follows. Let v be the center of S'' and d its degree in T. The embedding S'' distributes its charge equally among all embeddings S' of k-vertex trees that share an edge with S'' such that the maximum degree of the vertices of S' in T is at most d. Each such embedding S' receives charge from at least $\binom{d-1}{k-2}$ embeddings of stars and each embedding of a k-vertex star centered at v sends charge to at most $(k-1)!k(d-1)^{k-2} \leqslant k!d^{k-2}$ embeddings of S. Hence, each embedding S' of a k-vertex tree such that the maximum degree of the vertices of S' in T is $d > k$ receives at least

$$\binom{d-1}{k-2}\frac{1}{k!d^{k-2}} \geqslant \frac{1}{k!k^{k-2}} \geqslant k^{-(2k-3)}$$

units of charge.

Since every embedding of a k-vertex tree in T (regardless whether the embedding is of S or not) has at least $k^{-(2k-3)}$ units of charge at the end of the process described above, it follows that $d(S,T) \leqslant 1 - k^{-(2k-3)}$. □

We now combine the results of Sections 4.3–4.6 to prove the main result of this chapter. As mentioned earlier, we do not attempt to optimise the upper bound on the inducibility presented in the theorem.

Proof of Theorem 4.1. Since S is neither a star nor a path, $|S| \geqslant 5$. If $|S| \leqslant 16$, then the inducibility of S is at most $1 - 16^{-29} \leqslant 1 - 10^{-35}$ by Lemma 4.14. Now assume that $|S| \geqslant 17$. If S is a caterpillar, then the inducibility of S is at most $1 - 10^{-4}$ by Theorem 4.13. If S is not a caterpillar, then consider an arbitrary hub v of S. By the definition of a hub, v is the root of three non-trivial branches. If v is adjacent to at most one leaf, or is the root of at most one fork and at least three major branches, then the inducibility of S is at most $1-10^{-7}$ by Theorem 4.9. Otherwise, v is adjacent to at least two leaves, and additionally is either the root of at least two forks or at most two major branches. In either case, v is the root of a fork since every non-trivial branch is either a fork or is major. Theorem 4.10 then guarantees that the inducibility of S is at most $1 - 10^{-4}$. □

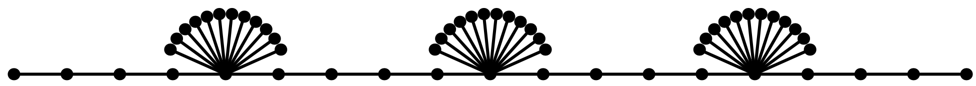


Figure 4.3: The tree T_3 constructed for $k = 4$ in the proof of Theorem 4.15.

4.7 Inducibility bounded away from 0

A *sparkler* is a graph obtained from a star by subdividing one of its edges once. The following result shows that sparklers are an infinite class of trees with inducibility bounded away from 0. This implies Theorem 4.2 and answers Problem 4 of Bubeck and Linial [22] in the affirmative.

Theorem 4.15. *The inducibility of every sparkler with at least four edges is at least* 13/165*.*

Proof. Fix $k \geqslant 4$, and let S'_k be the sparkler with k edges; that is, the graph obtained from the star with $k-1$ leaves by subdividing one of its edges. We will construct a a sequence $(T_n)_{n\in\mathbb{N}}$ of trees with $|T_n| \to \infty$ such that $d(S'_k, T_n) \geqslant 13/165$, which implies the theorem.

As illustrated in Figure 4.3, let T_n be the tree obtained from a path with $n(k+1)+k$ vertices (called the *spine*) by adding $3k$ leaves to its $(j(k+1))$-th vertex for $j \in \{1, \ldots, n\}$; each of the n vertices to which the leaves are attached is called a *vertebra*.

Observe that the number of copies of S'_k in T_n is

$$2n\binom{3k+1}{k-2}.$$

We next count the number of all k-edge subtrees of T_n. Each k-edge subtree of T_n contains exactly one of the vertebrae. The number of k-edge subtrees that contain exactly j edges from the spine of T_n for $j \in \{0, \ldots, k\}$ is

$$(j+1)n\binom{3k}{k-j}.$$

Thus the total number of k-edge subtrees of T_n is

$$n\sum_{j=0}^{k}(j+1)\binom{3k}{k-j}. \tag{4.1}$$

Observe that

$$\frac{\binom{3k}{k-j}}{\binom{3k}{k-j-1}} = \frac{2k+j+1}{k-j} \geqslant 2$$

for every $j \in \{0, \ldots, k-1\}$, which can be used iteratively on (4.1) to bound the number of k-edge subtrees of T_n:

$$n \sum_{j=0}^{k} (j+1) \binom{3k}{k-j} \leqslant \binom{3k}{k} n \sum_{j=0}^{k} \frac{j+1}{2^j} \leqslant \binom{3k}{k} n \sum_{j=0}^{\infty} \frac{j+1}{2^j}\,.$$

The arithmetico–geometric series in the last expression sums to 4, so it follows that the density of S_k' in T_n is at least

$$\frac{2\binom{3k+1}{k-2}n}{4\binom{3k}{k}n} = \frac{(3k+1)k(k-1)}{2(2k+3)(2k+2)(2k+1)} \geqslant \frac{13}{165},$$

where the last inequality holds since $k \geqslant 4$. □

We remark that the construction from Theorem 4.15 can be optimised by adding $\lceil \alpha k \rceil$ leaves instead of adding $3k$ to the vertebrae for $\alpha \approx 2.8507$, which yields that the inducibility of sufficiently large sparklers is at least 0.19004, while the bound presented in the proof converges to 3/16 for k tending to infinity.

4.8 Universal sequence of trees

In this section, we construct a universal sequence of trees, i.e., a sequence of trees in which the limit density of every tree S is positive.

Proof of Theorem 4.3. To describe the construction, we first define a gluing operation on trees, which we denote by $\oplus$; this operation has already been used in the context of tree profiles in [22]. If T and T' are trees, then $T \oplus T'$ is any tree obtained from the disjoint union of T and T' by joining a vertex of T and a vertex of T' by an edge. The resulting tree depends, of course, on which vertices are chosen to be joined by an edge, but the choice will not influence our arguments as long as the maximum degree of the resulting tree is controlled when we do a sequence of these operations. In particular, if we always choose a leaf of T and a leaf of T', then the maximum degree does not increase (unless $T \cong K_2$ or $T' \cong K_2$).

Observe that the number of k-vertex trees containing the gluing edge is at most $(k-1)^{k-1}\left(\Delta(T \oplus T') - 1\right)^{k-1} \leqslant \left(k(\Delta(T \oplus T') - 1)\right)^{k-1}$, where $\Delta(T \oplus T')$ denotes the maximum degree of $T \oplus T'$ (start with the gluing edge and then add $k-1$ edges

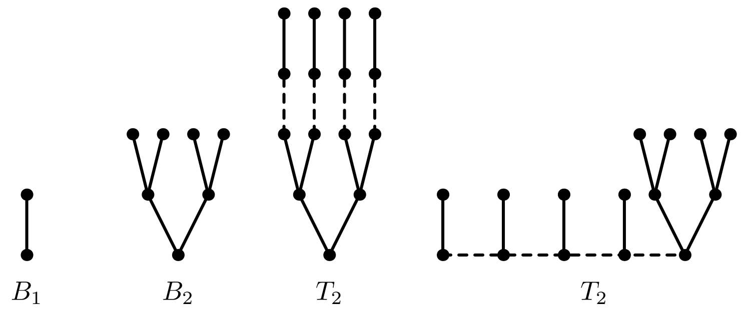


Figure 4.4: The trees B_1, B_2, and two possible choices of T_2. Edges added by the operation $\oplus$ are dashed.

iteratively, having at most $(k-1)\,(\Delta(T \oplus T') - 1)$ at each iteration). This yields that

$$Z_k(A) + Z_k(B) \leqslant Z_k(A \oplus B) \leqslant Z_k(A) + Z_k(B) + \big(k(\Delta(T \oplus T') - 1)\big)^{k-1} \tag{4.2}$$

We further define an iterative version of the gluing operation $\oplus$ by setting $T^{\oplus 1} = T$ and $T^{\oplus \ell} = T^{\oplus(\ell-1)} \oplus T$ for $\ell \geqslant 2$.

Let B_d be the complete d-ary tree of depth d; that is, B_d is the rooted tree such that every internal vertex has d children and every leaf is at distance d from the root. Observe that $|B_d| = 1 + d + d^2 + \cdots + d^d = \frac{d^{d+1}-1}{d-1} \leqslant d^{d+1}$, the maximum degree of B_d is $d+1$ if $d \geqslant 2$ and 1 if $d = 1$, and the tree B_d contains a copy of every tree with d vertices. We now define the sequence $(T_n)_{n \in \mathbb{N}}$ in the statement of the theorem. The tree T_n is obtained by gluing copies of the trees $B_1, \ldots, B_n$ in a ratio such that a significant proportion of the k-vertex subtrees in the resulting tree T_n arises from copies of $B_1, \ldots, B_k$. Formally, set $T_1 = B_1$, and for $n \geqslant 2$, define

$$T_n = B_n \oplus \left(T_{n-1}^{\oplus n^2}\right),$$

where the gluing operation is performed so that $\Delta(T_n) \leqslant n + 1$. Observe that T_n consists of $\left(\frac{n!}{d!}\right)^2$ copies of B_d for $d \in \{1, \ldots, n\}$. See Figure 4.4 for an illustration.

Fix a k-vertex tree S with $k \geqslant 3$ for the rest of the proof, and note that

$$d(S, T_n) \geqslant \frac{1}{Z_k(T_n)} \cdot \left(\frac{n!}{k!}\right)^2 \tag{4.3}$$

for every $n \geqslant k$. We next upper bound the number of k-vertex subtrees in T_n by using (4.2):

$$\begin{aligned}
Z_k(T_n) &\leqslant Z_k(B_n) + Z_k(T_{n-1}^{\oplus n^2}) + (kn)^{k-1} \\
&\leqslant Z_k(B_n) + n^2 Z_k(T_{n-1}) + n^2(kn)^{k-1} \\
&= Z_k(B_n) + n^2 Z_k(T_{n-1}) + k^{k-1}n^{k+1}.
\end{aligned}$$

Iterating the inequality, we obtain that

$$Z_k(T_n) \leqslant \left[\sum_{d=k+1}^{n} \left(\frac{n!}{d!}\right)^2 \left(Z_k(B_d) + k^{k-1}d^{k+1}\right)\right] + \left(\frac{n!}{k!}\right)^2 Z_k(T_k).$$

We next analyse the sum from the above expression:

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum_{d=k+1}^{n} \left(\frac{n!}{d!}\right)^2 \left(Z_k(B_d) + k^{k-1}d^{k+1}\right)}{\left(\frac{n!}{k!}\right)^2} &= \sum_{d=k+1}^{\infty} \left(\frac{k!}{d!}\right)^2 \left(Z_k(B_d) + k^{k-1}d^{k+1}\right) \\
&\leqslant \sum_{d=k+1}^{\infty} \left(\frac{k!}{d!}\right)^2 \left(d^{d+1}k^{k-1}d^{k-1} + k^{k-1}d^{k+1}\right) \\
&\leqslant 3(k!)^2 k^{k-1} \sum_{d=k+1}^{\infty} \frac{2d^{d+k}}{(d!)^2} \\
&\leqslant 6(k!)^2 k^{k-1} \sum_{d=k+1}^{\infty} \frac{d^{d+k}e^{2d-2}}{d^{2d}} \\
&\leqslant 6(k!)^2 k^{k-1} e^{2k}.
\end{aligned}$$

This combines with (4.3) to imply that

$$d(S, T_n) \geqslant \frac{1}{6(k!)^2 k^{k-1} e^{2k} + Z_k(T_k)} > 0.$$

Considering a convergent subsequence of (T_n) if necessary, we deduce that there exists a convergent sequence of trees in which the limit density of every tree S is positive. □

We remark that the choice of the vertices for the gluing operation permits creating sequences of trees with different “shapes”. For example, the trees can be grown to the depth as the left tree T_2 in Figure 4.4 or along a path as the right tree T_2 in Figure 4.4.

Bibliography

[1] M. H. Albert, M. D. Atkinson, C. C. Handley, D. A. Holton and W. Stromquist: *On packing densities of permutations*, Electron. J. Combin. **9** (2002).

[2] J. Balogh, P. Hu, B. Lidický, O. Pikhurko, B. Udvari and J. Volec: *Minimum number of monotone subsequences of length 4 in permutations*, Combin. Probab. Comput. **24** (2015), 658–679.

[3] I. Benjamini and O. Schramm: *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001).

[4] W. Bergsma and A. Dassios: *A consistent test of independence based on a sign covariance related to Kendall's tau*, Bernoulli **20** (2014), 1006–1028.

[5] G. Blekherman, A. Raymond, M. Singh and R. R. Thomas: *Simple graph density inequalities with no sum of squares proofs*, Combinatorica **40** (2020), 455–471.

[6] B. Bollobás: *Relations between sets of complete subgraphs*, Proc. Fifth British Combinatorial Conference (1975), 79–84.

[7] B. Bollobás: Modern Graph Theory, 4th edition, 1998, Springer.

[8] B. Bollobás: Extremal Graph Theory, 2004, Dover, reprint of the 1978 edition.

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

[10] B. Borchers: *CSDP* (1997), available at `https://github.com/coin-or/csdp/`.

[11] C. Borgs, J. Chayes and D. Gamarnik: *Convergent sequences of sparse graphs: A large deviations approach*, Random Structures Algorithms **51** (2017), 52–89.

[12] C. Borgs, J. Chayes, J. Kahn and L. Lovász: *Left and right convergence of graphs with bounded degree*, Random Structures Algorithms **42** (2013), 1–28.

[13] C. Borgs, J. Chayes and L. Lovász: *Moments of two-variable functions and the uniqueness of graph limits*, Geom. Funct. Anal. **19** (2010), 1597–1619.

[14] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy and K. Vesztergombi: *Graph limits and parameter testing*, Proc. 38th Annual ACM Symposium on Theory of Computing (2006), 261–270.

[15] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi: *Counting graph homomorphisms*, in: Topics in Discrete Mathematics (2006), 315–371.

[16] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi: *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), 1801–1851.

[17] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi: *Convergent sequences of dense graphs II: Multiway cuts and statistical physics*, Ann. of Math. (2) (2012), 151–219.

[18] F. Brandt, M. Brill and P. Harrenstein: *Tournament solutions*, in: Handbook of Computational Social Choice (2016), 57–84.

[19] A. Brauer and I. C. Gentry: *On the characteristic roots of tournament matrices*, Bull. Amer. Math. Soc. **74** (1968), 1133–1135.

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

[21] S. Bubeck, K. Edwards, H. Mania and C. Supko: *On paths, stars and wyes in trees* (2016), preprint arXiv:1601.01950.

[22] S. Bubeck and N. Linial: *On the local profiles of trees*, J. Graph Theory **81** (2016), 109–119.

[23] M. Bucić, E. Long, A. Shapira and B. Sudakov: *Tournament quasirandomness from local counting*, preprint arXiv:1910.09936 (2019).

[24] A. Cayley: *On the theory of the analytical forms called trees*, Philosophical Magazine, Series 4 **13** (1857), 172–176.

[25] T. F. N. Chan, A. Grzesik, D. Král' and J. A. Noel: *Cycles of length three and four in tournaments*, J. Combin. Theory Ser. A **175** (2020).

[26] T. F. N. Chan, D. Kráľ, B. Mohar and D. R. Wood: *Inducibility and universality for trees*, preprint arXiv:2102.02010 (2020).

[27] T. F. N. Chan, D. Kráľ, J. A. Noel, Y. Pehova, M. Sharifzadeh and J. Volec: *Characterization of quasirandom permutations by a pattern sum*, Random Structures Algorithms **57** (2020), 920–939.

[28] F. R. K. Chung and R. L. Graham: *Quasi-random hypergraphs*, Random Structures Algorithms **1** (1990), 105–124.

[29] F. R. K. Chung and R. L. Graham: *Quasi-random set systems*, J. Amer. Math. Soc. **4** (1991), 151–196.

[30] F. R. K. Chung and R. L. Graham: *Quasi-random tournaments*, J. Graph Theory **15** (1991), 173–198.

[31] F. R. K. Chung and R. L. Graham: *Quasi-random subsets of* Z_n, J. Combin. Theory Ser. A **61** (1992), 64–86.

[32] F. R. K. Chung, R. L. Graham and R. M. Wilson: *Quasi-random graphs*, Combinatorica **9** (1989), 345–362.

[33] J. Cooper and A. Petrarca: *Symmetric and asymptotically symmetric permutations*, preprint arXiv:0801.4181 (2008).

[34] J. N. Cooper: *Quasirandom permutations*, J. Combin. Theory Ser. A **106** (2004), 123–143.

[35] L. N. Coregliano and A. A. Razborov: *On the density of transitive tournaments*, J. Graph Theory **85** (2017), 12–21.

[36] É. Czabarka, A. A. Dossou-Olory, L. A. Székely and S. Wagner: *Inducibility of d-ary trees*, Discrete Math. **343** (2020).

[37] É. Czabarka, L. A. Székely and S. Wagner: *Inducibility in binary trees and crossings in random tanglegrams*, SIAM J. Discrete Math. **31** (2017), 1732–1750.

[38] É. Czabarka, L. A. Székely and S. Wagner: *Paths vs. stars in the local profile of trees*, Electron. J. Combin. **24** (2017).

[39] R. Diestel: Graph Theory, 5th edition, 2016/17, Springer.

[40] A. A. V. Dossou-Olory and S. Wagner: *Inducibility of topological trees*, Quaest. Math. **42** (2018), 749–764.

[41] A. A. V. Dossou-Olory and S. Wagner: *On the inducibility of small trees*, Discrete Math. Theor. Comput. Sci. **21** (2019).

[42] G. Elek: *Note on limits of finite graphs*, Combinatorica **27** (2007), 503–507.

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

[44] P. Erdős: *Some theorems on graphs*, Riveon Lematematika **9** (1955), 13–17.

[45] P. Erdős and G. Szekeres: *A combinatorial problem in geometry*, Compos. Math. (1935).

[46] C. Even-Zohar and N. Linial: *A note on the inducibility of 4-vertex graphs*, Graphs Combin. **31** (2015), 1367–1380.

[47] D. C. Fisher: *Lower bounds on the number of triangles in a graph*, J. Graph Theory **13** (1989), 505–512.

[48] D. H. Fremlin: *Volume 1: The Irreducible Minimum*, in: Measure Theory (2010).

[49] A. Frieze and R. Kannan: *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–200.

[50] K. Fujisawa, M. Fukuda, Y. Futakata, K. Kobayashi, M. Kojima, K. Nakata, M. Nakata and M. Yamashita: *SDPA Project* (1995), available at `http://sdpa.sourceforge.net/`.

[51] F. R. Gantmacher: The Theory of Matrices, Volume 1, 1959, AMS Chelsea, translated from Russian by K. A. Hirsch.

[52] M. Goemans and F. Rendl: *Combinatorial optimization*, in: Handbook of Semidefinite Programming (2000), 343–360.

[53] A. W. Goodman: *On sets of acquaintances and strangers at any party*, Amer. Math. Monthly **66** (1959), 778–783.

[54] W. T. Gowers: *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), 143–184.

[55] W. T. Gowers: *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), 897–946.

[56] W. T. Gowers: *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), 363–387.

[57] A. Grzesik: *Flag algebras in extremal graph theory*, Ph.D. thesis, Jagiellonian University (2014).

[58] R. Hancock, A. Kabela, D. Král', T. Martins, R. Parente, F. Skerman and J. Volec: *No additional tournaments are quasirandom-forcing*, preprint arXiv:1912.04243 (2019).

[59] H. Hatami, J. Hirst and S. Norine: *The inducibility of blow-up graphs*, J. Combin. Theory Ser. B **109** (2014), 196–212.

[60] H. Hatami, L. Lovász and B. Szegedy: *Limits of locally–globally convergent graph sequences*, Geom. Funct. Anal. **24** (2014), 269–296.

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

[62] J. Haviland and A. Thomason: *Pseudo-random hypergraphs*, Discrete Math. **75** (1989), 255–278.

[63] D. Hefetz and M. Tyomkyn: *On the inducibility of cycles*, J. Combin. Theory Ser. B **133** (2018), 243–258.

[64] J. Hirst: *The inducibility of graphs on four vertices*, J. Graph Theory **75** (2014), 231–243.

[65] W. Hoeffding: *A non-parametric test of independence*, Ann. Math. Statist. **19** (1948), 546–557.

[66] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth and R. M. Sampaio: *Limits of permutation sequences*, J. Combin. Theory Ser. B **103** (2013), 93–113.

[67] T. Huynh and D. R. Wood: *Tree densities in sparse graph classes*, preprint arXiv:2009.12989 (2020).

[68] D. E. Knuth: The Art of Computer Programming, Volume 1: Fundamental Algorithms, 3rd edition, 1997, Addison-Wesley.

[69] Y. Kohayakawa, V. Rödl and J. Skokan: *Hypergraphs, quasi-randomness, and conditions for regularity*, J. Combin. Theory Ser. A **97** (2002), 307–352.

[70] D. Král', S. Norin and J. Volec: *A bound on the inducibility of cycles*, J. Combin. Theory Ser. A **161** (2019), 359–363.

[71] D. Král' and O. Pikhurko: *Quasirandom permutations are characterized by 4-point densities*, Geom. Funct. Anal. **23** (2013), 570–579.

[72] N. Linial and A. Morgenstern: *On the number of 4-cycles in a tournament*, J. Graph Theory **83** (2016), 266–276.

[73] H. Liu, O. Pikhurko and K. Staden: *The exact minimum number of triangles in graphs of given order and size*, Forum Math., Pi **8** (2020), article no. E8.

[74] L. Lovász: Large Networks and Graph Limits, *Colloquium Publications*, volume 60, 2012, American Mathematical Society.

[75] L. Lovász and M. Simonovits: *On the number of complete subgraphs of a graph II*, in: Studies in Pure Mathematics (1983), 459–495.

[76] L. Lovász and B. Szegedy: *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), 933–957.

[77] L. Lovász and B. Szegedy: *Testing properties of graphs and functions*, Israel J. Math. **178** (2010), 113–156.

[78] L. Lovász and B. Szegedy: *Random graphons and a weak positivstellensatz for graphs*, J. Graph Theory **70** (2012), 214–225.

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

[80] J. Nešetřil and P. Ossona de Mendez: *A unified approach to structural limits and limits of graphs with bounded tree-depth*, Mem. Amer. Math. Soc. **263** (2020).

[81] J. Nešetřil and P. Ossona de Mendez: *A model theory approach to structural limits*, Comment. Math. Univ. Carolin. **53** (2012), 581–603.

[82] O. Pikhurko and A. Razborov: *Asymptotic structure of graphs with the minimum number of triangles*, Combin. Probab. Comput. **26** (2017), 138–160.

[83] N. Pippenger and M. C. Golumbic: *The inducibility of graphs*, J. Combin. Theory Ser. B **19** (1975), 189–203.

[84] C. B. Presutti and W. Stromquist: *Packing rates of measures and a conjecture for the packing density of 2413*, in: Permutation Patterns (2010), 287–316.

[85] A. A. Razborov: *Flag algebras*, J. Symb. Log. **72** (2007), 1239–1282.

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

[87] A. A. Razborov: *Flag algebras: An interim report*, in: The Mathematics of Paul Erdös II (2013), 207–232.

[88] L. Rédei: *Ein kombinatorischer satz*, Acta Litteraria Szeged **7** (1934), 39–43.

[89] V. Rödl: *On universality of graphs with uniformly distributed edges*, Discrete Math. **59** (1986), 125–134.

[90] A. Sidorenko: *A correlation inequality for bipartite graphs*, Graphs Combin. **9** (1993), 201–204.

[91] J. Skokan and L. Thoma: *Bipartite subgraphs and quasi-randomness*, Graphs Combin. **20** (2004), 255–262.

[92] J. Sliačan and W. Stromquist: *Permpack* (2017), available at `https://github.com/jsliacan/permpack`.

[93] J. Sliačan and W. Stromquist: *Improving bounds on packing densities of 4-point permutations*, Discrete Math. Theor. Comput. Sci. **19** (2018).

[94] E. Szemerédi: *Regular partitions of graphs*, Colloq. Int. CNRS **260** (1978), 399–401.

[95] A. Thomason: *Pseudorandom graphs*, Ann. Discrete Math. **33** (1987), 307–331.

[96] P. Turán: *On an extremal problem in graph theory*, Matematikai és Fizikai Lapok **48** (1941), 436–452.

[97] E. R. Vaughan: *Flagmatic 2.0* (2013), available at `https://github.com/jsliacan/flagmatic`.

[98] J. Volec: *Analytic methods in combinatorics*, Ph.D. thesis, University of Warwick and Université Paris Diderot (2014).

[99] J. Širáň: *A new lower bound for the inducibility of a graph*, Math. Slovaca **34** (1984), 365–370.

[100] T. Yanagimoto: *On measures of association and a related problem*, Ann. Inst. Statist. Math. **22** (1970), 57–63.

[101] R. Yuster: *On the exact maximum induced density of almost all graphs and their inducibility*, J. Combin. Theory Ser. B **136** (2019), 81–109.

[102] E. Zhang: *On quasirandom permutations* (2018), presented at the MIT PRIMES Conference. Slides available at `https://math.mit.edu/research/highschool/primes/materials/2018/\conf/9-2%20Zhang.pdf`.