UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

TOPICS IN EXTREMAL AND PROBABILISTIC COMBINATORICS: TREES AND WORDS

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MODELACIÓN MATEMÁTICA

MATÍAS NICOLÁS PAVEZ SIGNÉ

PROFESORA GUÍA:
MAYA JAKOBINE STEIN
PROFESOR CO-GUÍA:
HIỆP HÀN

MIEMBROS DE LA COMISIÓN
JULIA BÖTTCHER
MARCOS KIWI KRAUSKOPF
ROBERT MORRIS
JOSÉ SOTO SAN MARTÍN

Este trabajo ha sido parcialmente financiado por CMM ANID PIA AFB170001 y la beca ANID-PFCHA/Doctorado Nacional/2017-21171132

SANTIAGO DE CHILE
2021

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MODELACIÓN MATEMÁTICA
POR: MATÍAS NICOLÁS PAVEZ SIGNÉ
FECHA: 2021
PROFESORA GUÍA: MAYA JAKOBINE STEIN
PROFESOR CO-GUÍA: HIỆP HÀN

TOPICS IN EXTREMAL AND PROBABILISTIC COMBINATORICS:
TREES AND WORDS

En esta tesis se estudia una serie de problemas en combinatoria extremal y probabilista relacionados a árboles y palabras. En la primera parte de este trabajo se estudian qué condiciones debe cumplir un grafo para que contenga a todos los árboles de cierto tamaño. Se prueban una serie de resultados que combinan condiciones de grado mínimo y máximo para contener a todos los árboles de cierto tamaño y grado acotado. También se logra un avance en la conjetura de Erdős–Sós [42] para árboles de grado acotado. Finalmente, se estudia el problema de contenimiento de árboles en el grafo aleatorio $G(n,p)$. Se prueba que incluso después de borrar una fracción de las aristas de $G(n,p)$ el grafo resultante sigue conteniendo árboles grandes con grado acotado.

En la segunda parte de esta tesis se estudian problemas extremales para palabras. Se determina el largo mínimo que debe tener una palabra para contener cada palabra de largo $k \in \mathbb{N}$. Además, se determina el umbral $n = n(k)$ de modo que, con alta probabilidad, una palabra aleatoria de largo $(1+\varepsilon)n$ contenga una copia de cada palabra de largo k. Finalmente, se estudia una noción de *cuasi-aleatoriedad* para palabras y se muestra una serie de propiedades equivalentes. Basados en esta noción de cuasi-aleatoriedad, se desarrolla una *teoría límite* para palabras finitas en el espíritu de lo que se ha hecho para grafos [82].

In this thesis, we study several problems in extremal and probabilistic combinatorics regarding trees and words. In the first part of this work we study which conditions a graph has to satisfy in order to contain every tree of certain size. We obtain a series of results regarding a combination of minimum and maximum degree that ensures the containment of every tree of certain size and bounded degree. We also make progress towards the Erdős–Sós conjecture [42] for trees with bounded maximum degree. Finally, we study the tree containment problem in the random graph $G(n,p)$ showing that even after a deletion of a fraction of the edges of $G(n,p)$ the resulting subgraph still contains large trees of bounded degree.

In the second part of this thesis we study extremal problems for words. We determine the minimum length of a word containing every word of length $k \in \mathbb{N}$, and the threshold $n = n(k)$ so that, with high probability, a random word of length $(1+\varepsilon)n$ contains a copy of every word of length k. Finally, we study a notion of *quasi-randomness* for words and we show a series of equivalent properties. Based on this quasi-random notion, we develop a *limit theory* for finite words in the spirit of what has been done for graphs [82].

A Sergio Santelices Arrospide, ex profesor del Instituto Nacional. Q.E.P.D.

Agradecimientos

Me gustaría comenzar agradeciendo a mis profesors guías Maya Stein y Hiệp Hàn por estos 4 años. Gracias por todo, aprendí muchísimo de ambos y guardo con mucho cariño estos 4 años de trabajo (espero haber estado a la altura). Me gustaría también agradecer Julia Bötcher, Marcos Kiwi, Rob Morris, y José Soto por haber accedido a ser parte de mi comité de tesis. En particular, me gustaría agradecer a Marcos Kiwi, quién fue casi mi tercer profesor guía, y a Rob Morris, que me recibió en el IMPA y me ha apoyado en diversas cosas durante este tiempo.

También me gustaría agradecer a Fábio Botler, Pedro Campos Araújo, Jan Corsten, Gwen McKinley, Patrick Morris, Guilherme Mota, Simón Piga, Daniel Quiroz, Nicolás Sanhueza Matamala (entre otros). Gracias por hacer las conferencias tan amigables, ahora cada conferencia es una gran instancia para poder ver a los amigos y amigas que he hecho durante estos años. En particular, me gustaría agradecer a Nico y Chopo, con quienes puedo pasar horas hablando de matemática, principalmente, y de otras cosas también. Espero que la vida nos llene de colaboraciones y teoremas nuevos.

Quiero agradecer a la comunidad DIM–CMM por todos estos años. Gracias a Eterin, Karen y Silvia por todas las cosas en que me ayudaron, y en particular a Natacha por haberme ayudado en básicamente todo, sin su ayuda no creo que hubiese podido llegar hasta aquí. Quiero agradecer también a los profesores y profesoras por la formación que recibí durante estos largos años. Gracias también a Rodolfo Gutierrez, con quien he compartido mi formación desde el comienzo y ha sido un gran amigo e inspiración durante estos años.

Esta tesis fue escrita entre medio de una revuelta social en Chile y una pandemia. No sé cómo logré superar este año 2020, pero estoy seguro que sin mi familia y amigos esto no hubiese sido posible. Gracias a todes ustedes por la vida que me han dado, lo principal son los afectos. En particular, gracias a mi madre por su apoyo incondicional y gracias a mi hermano por ser la mejor persona que conozco. Quisiera agradecer a Paula por todos estos años juntos, gracias por compartir las alegrías, acompañarnos en los momentos oscuros, y ayudarme con mis inseguridades. Agradezco poder crecer junto a ti. Finalmente, quiero dar gracias a mis compañeros animales Bongo, Botón y Filipo, quienes me han llenado de felicidad desde que llegaron a mi vida.

Contents

Part I

Tree embeddings and degree conditions

Chapter 1

Introduction

A central problem in graph theory consists of determining which conditions a graph G has to satisfy in order to ensure it contains a given substructure. For instance, one of the most important questions in extremal graph theory is the *Turán problem*, which asks for global degree conditions to force the containment of a graph or, more generally, a family of graphs. The *extremal number* of a graph H, denoted by $\mathrm{ex}(n, H)$, is defined as the maximum number of edges in a graph G on n vertices which does not contain H as a subgraph. Another important problem is to determine local degree conditions that force the containment of certain graphs. For instance, Dirac's theorem states that every graph on $n \geqslant 3$ vertices with minimum degree at least $\frac{n}{2}$ contains a *Hamilton cycle*, that is, a cycle that uses each vertex on the graph exactly once. For a general overview of this area, we refer to the recent survey of Simonovits and Szemerédi [103].

In this thesis, we will be interested in the class of graphs called *trees*. A tree is a *connected* graph without cycles. In this part of the thesis, we will focus on degree conditions that ensure the containment of all trees of a given size satisfying some condition on the maximum degree.

Let us start with an easy observation. Let $k \in \mathbb{N}$, let G be a graph with minimum degree at least k, and let T be a tree with k edges rooted at some vertex $r \in V(T)$. We claim that T embeds into G. Indeed, starting from any vertex $v \in V(G)$, we may map r to v and then greedily embed the children of r into unoccupied neighbours of v. We then repeat this argument with the children of r and so on until T is completely embedded. It is important to note that the minimum degree of G is at least as large as the remaining vertices of T at each step, which ensure that we can run this argument until T is completely embedded. Although this minimum degree condition is rather strong, we note that it is actually tight. Indeed, one can consider the union of several disjoint copies of K_k, which does not contain any tree with k edges.

1.1 Average degree

Arguably, the most important problem regarding tree containment and degree conditions is a famous conjecture of Erdős and Sós from 1964, which suggests that it is possible to replace the minimum degree condition, that we discussed before, with a bound on the average degree.

Conjecture 1.1.1 (Erdős and Sós [42])**.** *Let $k \in \mathbb{N}$ and let G be a graph with average degree greater than $k-1$. Then G contains a copy of every tree with k edges as a subgraph.*

We observe that this conjecture is tight for every $k \in \mathbb{N}$, which can be seen by considering (again) the complete graph on k vertices. This graph has average degree exactly $k-1$ but it is too small to contain any tree with k edges. A structurally different example is the *balanced* complete bipartite graph on $2k-2$ vertices (where by balanced we mean that the bipartition classes have equal sizes). This graph has average degree $k-1$ but does not contain the star with k edges. In order to obtain examples of larger order, one can consider the disjoint union of copies of the two *extremal graphs* we just described.

To illustrate the importance of the Erdős–Sós conjecture in extremal graph theory, let us quickly consider the Turán problem for trees. We first note that for any fixed tree T with k edges, the minimum degree condition discussed before implies that $\mathrm{ex}(n,T) \leqslant (k-1)n$. Indeed, one can prove that any graph G on n vertices such that $e(G) > (k-1)n$ contains a subgraph of minimum degree at least k, and therefore contains a copy of T. On the other hand, the Erdős–Sós conjecture greatly improves this bound by a factor of $\frac{1}{2}$, that is, Conjecture 1.1.1 would imply

$$\mathrm{ex}(n,T) \leqslant \frac{(k-1)}{2}n. \tag{1.1}$$

The Erdős-Sós conjecture has further consequences in *Ramsey theory* for trees (see Section 4.5).

Let us now give some evidence for Conjecture 1.1.1. It is easy to see that the Erdős–Sós conjecture is true for stars and double stars (the latter are graphs obtained by joining the centres of two stars with an edge). A classical result of Erdős and Gallai [43] implies that it also holds for paths. In the early 90s, Ajtai, Komlós, Simonovits and Szemerédi announced a proof of the Erdős–Sós conjecture for large k. Nevertheless, many particular cases have been settled since then. For instance, Brandt and Dobson [28] proved that the Erdős–Sós conjecture is true for graphs with girth at least 5, and Saclé and Woźniak [100] proved it for C_4-free graphs. Goerlich and Zak [54] proved the Erdős–Sós conjecture for graphs of order $n = k + c$, where c is a given constant and k is sufficiently large depending on c. More recently, Rozhoň [99] gave an approximate version of the Erdős–Sós conjecture for trees with linearly bounded maximum degree and dense host graph. In this thesis, we will show that Conjecture 1.1.1 holds for trees with maximum degree bounded by any given constant and dense host graph. Namely, we prove the following theorem.

Theorem 1.1.2 (Besomi, P., and Stein [18])**.** *For all $\delta > 0$ and $\Delta \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that for each $n, k \in \mathbb{N}$ with $n \geqslant n_0$ and $n \geqslant k \geqslant \delta n$ the following holds. Let G be a graph on n vertices such that $d(G) > k-1$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.*

1.2 Spanning trees and minimum degree

In his classical book of extremal graph theory, Bollobás [24] conjectured that for any $\delta > 0$ and $\Delta \in \mathbb{N}$, there is $n_0 \in \mathbb{N}$ such that every graph on $n \geqslant n_0$ vertices and minimum degree at least $(\frac{1}{2} + \delta)n$ would contain every spanning tree with maximum degree bounded by Δ. This conjecture was proved by Komlós, Sárközy and Szemerédi [72] in 1995, and its proof strategy was a prototype version of what is now known as the "blow-up lemma". In 2001, the same authors [73] improved their earlier result in a different direction, showing that one can actually embed spanning trees with maximum degree of order $O(\frac{n}{\log n})$.

Theorem 1.2.1 (Komlós, Sárközy and Szemerédi [73])**.** *For all $\delta > 0$, there are positive constants n_0 and C such that for all $n \geqslant n_0$ the following holds. Let G be a graph on n vertices such that $\delta(G) \geqslant (\frac{1}{2} + \delta)n$. Then G contains a copy of every tree T on n vertices such that $\Delta(T) \leqslant C\frac{n}{\log n}$.*

They also show that the bound on the maximum degree is essentially best possible. Indeed, for a sufficiently large constant $C > 0$, let T be the tree consisting of a vertex r connected to $\frac{\log n}{C}$ vertices such that each child of r has $C\frac{n}{\log n}$ children. Note that T has a dominating set of size $\frac{\log n}{C}$. Let us consider the binomial random graph $G = G(n, p)$ with $p = 0.9$. It is easy to see that, with high probability, G has minimum degree greater than $0.8n$ and has no dominating set of size larger than $\frac{\log n}{C}$. Thus, with high probability, G does not contain T as a subgraph.

For trees with maximum degree bounded by a constant, Csaba, Levitt, Nagy-György, and Szemerédi [39] showed in 2010 that actually a minimum degree of at least $\frac{n}{2} + \Omega(\log n)$ suffices.

Theorem 1.2.2 (Csaba, Levitt, Nagy-György, and Szemerédi [39])**.** *For all $\Delta \geqslant 2$, there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ the following holds. Let G be a graph on n vertices and $\delta(G) \geqslant \frac{n}{2} + c\log n$. Then G contains a copy of every tree T on n vertices such that $\Delta(T) \leqslant \Delta$.*

Moreover, they proved that there exists a graph G with $\delta(G) \geqslant \frac{n}{2} + \frac{\log n}{17}$ such that G does not contain the *complete ternary tree*. A very interesting question is to understand what happens if the minimum degree is between $\frac{n}{2} + c\log n$ and $\frac{n}{2} + \delta n$. The following problem was asked by Rob Morris [90].

Problem 1.2.3. *Let $f : \mathbb{N} \to \mathbb{N}$ be a function such that there exist positive constants c and C such that $c\log n \leqslant f(n) \leqslant Cn$ for all $n \in \mathbb{N}$. Determine a function $g : \mathbb{N} \to \mathbb{N}$ so that for all large n, every graph G on n vertices with minimum degree $\delta(G) \geqslant \frac{n}{2} + f(n)$ contains a copy of every tree on n vertices with maximum degree bounded by $g(n)$.*

In view of Theorems 1.2.1 and 1.2.2, it is tempting to make the following conjecture.

Conjecture 1.2.4. *Let c and C be positive constants and let $f : \mathbb{N} \to \mathbb{N}$ be a function satisfying $c\log n \leqslant f(n) \leqslant Cn$. Then there exist $K > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ the following holds. Let G be a graph on n vertices and minimum degree $\delta(G) \geqslant \frac{n}{2} + f(n)$. Then G contains a copy of every tree T on n vertices such that $\Delta(T) \leqslant K\frac{f(n)}{\log n}$.*

1.3 Maximum and minimum degree

A new angle in the tree containment problem was introduced in 2016 by Havet, Reed, Stein, and Wood [61], who impose bounds on both the minimum *and* the maximum degree to force the containment of every tree of fixed size. More precisely, they suggest the following conjecture, which we call the $\frac{2}{3}$*–conjecture*.

Conjecture 1.3.1 ($\frac{2}{3}$–conjecture; Havet, Reed, Stein, and Wood [61])**.** *Let $k \in \mathbb{N}$ and let G be a graph with maximum degree at least k and minimum degree at least $\lfloor \frac{2k}{3} \rfloor$. Then G contains a copy of every tree with k edges as a subgraph.*

The following example shows that Conjecture 1.3.1 is essentially tight. Let k be divisible by 3 and consider a graph G consisting of two disjoint copies of $K_{\frac{2k}{3}-2}$ and an additional vertex v which is adjacent to every other vertex in G. Let T be the tree consisting of three paths, each of length $\frac{k}{3}$, sharing a common end point. It is easy to see that T cannot be embedded into G, since at least two of those paths must be embedded into one of the copies of $K_{\frac{2k}{3}-2}$.

Conjecture 1.3.1 is obviously true for stars and double stars. The following argument shows that it also holds for paths. If the host graph G has a 2-connected component of size at least $k+1$, then by a variant[1] of Dirac's theorem, this component contains a cycle of length at least k, and thus also a k-edge path (possibly using one edge that leaves the cycle). Otherwise, we can embed a vertex from the middle of the path into any cutvertex x of G, and then greedily embed the remainder of the path into two components of $G - x$, using the minimum degree of G. In [61], Havet, Reed, Stein, and Wood proved the following partial results towards Conjecture 1.3.1.

Theorem 1.3.2 (Havet, Reed, Stein, and Wood [61])**.** *There exist a function $f : \mathbb{N} \to \mathbb{N}$ and a small constant $\gamma > 0$ such that if a graph G satisfies either*

1. *$\Delta(G) \geqslant f(k)$ and $\delta(G) \geqslant \lfloor \frac{2k}{3} \rfloor$, or*
2. *$\Delta(G) \geqslant k$ and $\delta(G) \geqslant (1-\gamma)k$,*

then G contains a copy of every tree with k edges.

Even if the degree conditions in Theorem 1.3.2 are no the same as in Conjecture 1.3.1, it proves that the idea combining a maximum and minimum degree condition is morally correct for the tree containment problem. We remark that the function $f(k)$ of Theorem 1.3.2 is super-exponential in k, and so any improvement on $f(k)$ would be of great interest. Moreover, Reed and Stein recently showed in [94, 95] that Conjecture 1.3.1 holds for large k, in the case of spanning trees (that is, if we additionally assume that $|V(G)| = |V(T)| = k+1$). We prove an approximate version of Conjecture 1.3.1 for trees with certain bound on the maximum degree and dense host graph.

[1]This variant was already observed by Dirac [41]. It states that every 2-connected n-vertex graph G has a cycle of length at least $\min\{n, 2\delta(G)\}$.

Theorem 1.3.3 (Besomi, P., and Stein [19])**.** *For all $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that for each $n, k \in \mathbb{N}$ with $n \geqslant n_0$ and $n \geqslant k \geqslant \delta n$ the following holds. Let G be a graph on n vertices with minimum degree at least $(1+\delta)\frac{2k}{3}$ and maximum degree at least $(1+\delta)k$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant k^{\frac{1}{49}}$.*

Another natural question is whether a version of Theorem 1.2.1 holds for trees that are not necessarily spanning. That is, one might ask if a graph G with minimum degree $\delta(G) \geqslant \frac{k}{2}$ contains a copy of every tree with k edges (or at least each such tree of bounded degree). Clearly, this cannot work because of the examples showing the tightness of Conjecture 1.1.1 or Conjecture 1.3.1. However, as in Conjecture 1.3.1, we believe that if in addition to the minimum degree condition, we require G to have at least one vertex of large degree, then every tree with k edges should be contained in G. More precisely, we believe that the following holds.

Conjecture 1.3.4 ($2k$–$\frac{k}{2}$ conjecture; Besomi, P., and Stein [19])**.** *Let $k \in \mathbb{N}$ and let G be a graph of minimum degree at least $\frac{k}{2}$ and maximum degree at least $2k$. Then G contains a copy of every tree with k edges as a subgraph.*

Let us give a quick example showing that Conjecture 1.3.4 is essentially tight (an example with better bounds will be given in Section 3.1). For $\varepsilon > 0$ and $k \in \mathbb{N}$, let $G_{\varepsilon,k}$ be the graph consisting of two disjoint copies of the complete bipartite graph, with parts of size $(1-\varepsilon)k$ and $(1-\varepsilon)\frac{k}{2}$, and one vertex that is adjacent to every vertex in the parts of size $(1-\varepsilon)k$. It is easy to see that $G_{\varepsilon,k}$ does not contain the tree T_k consisting of $\sqrt{k}$ stars of size $\sqrt{k}$ whose centers are adjacent to the central vertex of T_k, provided that k is sufficiently large.

Similar as for the $\frac{2}{3}$-conjecture, one can see that Conjecture 1.3.4 is true for stars, double stars, and paths. As more evidence for Conjecture 1.3.4, we prove an approximate version for trees of bounded degree and dense host graphs.

Theorem 1.3.5 (Besomi, P., and Stein [19])**.** *For all $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that for each $n, k \in \mathbb{N}$ with $n \geqslant n_0$ and $n \geqslant k \geqslant \delta n$ the following holds. Let G be a graph on n vertices with minimum degree at least $(1+\delta)\frac{k}{2}$ and maximum degree at least $(1+\delta)2k$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant k^{\frac{1}{67}}$ as a subgraph.*

Moreover, if we consider trees whose maximum degree is bounded by an absolute constant, we can improve the bound on the maximum degree of the host graph given by Theorem 1.3.5 as follows.

Theorem 1.3.6 (Besomi, P., and Stein [19])**.** *For all $\delta > 0$ and $\Delta \geqslant 2$, there exists $n_0 \in \mathbb{N}$ such that for each $n, k \in \mathbb{N}$ with $n \geqslant n_0$ and $n \geqslant k \geqslant \delta n$ the following holds. Let G be a graph on n vertices with minimum degree at least $(1+\delta)\frac{k}{2}$ and maximum degree at least $2(\frac{\Delta-1}{\Delta} + \delta)k$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.*

Comparing the two variants of maximum/minimum degree conditions given by conjectures 1.3.1 and 1.3.4, it seems natural to ask whether one can allow for a wider spectrum of bounds for the maximum and the minimum degree of the host graph. We believe that it is possible to weaken the bound on the maximum degree given by the $2k$–$\frac{k}{2}$ conjecture, if simultaneously, the bound on the minimum degree is increased. Quantitatively speaking, we suggest the following.

Conjecture 1.3.7 (Intermediate range conjecture; Besomi, P., and Stein [17])**.** *Let $k \in \mathbb{N}$ and let $\alpha \in [0, \frac{1}{3})$. Let G be a graph with $\delta(G) \geqslant (1+\alpha)\frac{k}{2}$ and $\Delta(G) \geqslant (1-\alpha)2k$. Then G contains a copy of every tree with k edges as a subgraph.*

Note that for $\alpha = 0$, the bounds from Conjecture 1.3.7 coincide with the bounds from the $2k$–$\frac{k}{2}$ conjecture. In contrast, the case $\alpha = \frac{1}{3}$ is not included in Conjecture 1.3.7 as we believe that the appropriate value for the maximum degree is k and not $\frac{4k}{3}$ if the minimum degree is $\frac{2}{3}k$ (as suggested by the $\frac{2}{3}$-conjecture). We show in Section 3.1 that Conjecture 1.3.7 is asymptotically best possible for infinitely many values of α.

Again, Conjecture 1.3.7 holds for stars, for double stars, and for paths. In this thesis, we provide further evidence for the correctness of Conjecture 1.3.7 by proving an approximate version for bounded degree trees and large dense host graphs.

Theorem 1.3.8 (Besomi, P., and Stein [17])**.** *For all $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that for each $\alpha \in [0, \frac{1}{3})$ and $n, k \in \mathbb{N}$ with $n \geqslant n_0$ and $n \geqslant k \geqslant \delta n$ the following holds. Let G be a graph on n vertices with minimum degree at least $(1+\delta)(1+\alpha)\frac{k}{2}$ and maximum degree at least $(1+\delta)(1-\alpha)2k$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant k^{\frac{1}{67}}$.*

1.4 Trees in random graphs

The binomial random graph $G(n,p)$ is a graph on n vertices where each of the possible $\binom{n}{2}$ edges appears, independently, with probability p. A *graph property* $\mathcal{P}$ is a family of graphs closed under isomorphisms. Given a graph property $\mathcal{P}$, the typical question in this area is to determine the probability of the event that $G(n,p) \in \mathcal{P}$. For many properties, such as *monotone properties*, this probability shows a *phase transition* as p grows from 0 to 1, meaning that $\mathbb{P}[G(n,p) \in \mathcal{P}]$ changes abruptly from 0 to 1 as p passes some threshold. We say that a graph property $\mathcal{P}$ has a *threshold* $p^* = p^*(n)$ if

$$\lim_{n\to\infty} \mathbb{P}[G(n,p) \in \mathcal{P}] = \begin{cases} 0 & \text{if } p = o(p^*), \\ 1 & \text{if } p = \omega(p^*). \end{cases}$$

If $\lim_{n\to\infty} \mathbb{P}(G(n,p) \in \mathcal{P}) = 1$ then we say that $\mathcal{P}$ holds with high probability. A classical result due to Erdős and Rényi [46] states that the property of being connected has a *sharp threshold* at $p^* = \frac{\log n}{n}$. This means that for any fixed $\varepsilon > 0$, for $p \geqslant (1+\varepsilon)\frac{\log n}{n}$ the random graph $G(n,p)$ is connected with high probability, and for $p \leqslant (1-\varepsilon)\frac{\log n}{n}$ the random graph $G(n,p)$ has an isolated vertex with high probability. In particular, this result states that if $p \geqslant (1+\varepsilon)\frac{\log n}{n}$ then, with high probability, $G(n,p)$ contains a spanning tree. A reasonable question is therefore to ask which spanning trees appear in $G(n,p)$ at this probability. The following conjecture was posed by Kahn.

Conjecture 1.4.1 (Kahn conjecture). *For every $\Delta \geqslant 2$ there exists a positive constant C such that the following holds. For any tree T with n vertices and $\Delta(T) \leqslant \Delta$, the random graph $G(n, C\frac{\log n}{n})$ contains a copy of T with high probability.*

This conjecture has received a lot of attention over the last 20 years and was recently solved in a much stronger form. Indeed, in 2018 Montgomery [88] showed that $G(n, C\frac{\log n}{n})$ contains a copy of every spanning tree with bounded degree at the same time. Montgomery's proof relies on the *absorption method* for random graphs. A radically different proof was recently found by Frankston, Kahn, Narayanan, and Park [50], who proved a fractional version of the *expectation-threshold conjecture of Kahn and Kalai* [67] which, among many other results, implies Conjecture 1.4.1.

Regarding almost spanning trees, Alon, Krivelevich, and Sudakov [5] proved that for any $\alpha \in (0,1)$ and $\Delta \geqslant 2$, there exists a constant $C > 0$ such that, with high probability, the random graph $G(n, \frac{C}{n})$ contains a copy of every tree T on $(1-\alpha)n$ vertices and $\Delta(T) \leqslant \Delta$. In 2014, Balogh, Csaba, and Samotij [13] showed that even by deleting a $(\frac{1}{2} - \delta)$-fraction of the edges incident to each vertex from $G(n, \frac{C}{n})$, the resulting subgraph still contains a copy of every almost spanning tree of bounded degree.

Theorem 1.4.2 (Balogh, Csaba, and Samotij [13]). *For every $\Delta \geqslant 2$ and $\alpha, \delta \in (0,1)$, there exists $C > 0$ such that if $p \geqslant \frac{C}{n}$, then $G = G(n,p)$, with high probability, has the following property. Let $G' \subseteq G$ be a subgraph with $\delta(G') \geqslant (\frac{1}{2} + \delta)pn$, then G' contains a copy of every tree T on $(1-\alpha)n$ vertices such that $\Delta(T) \leqslant \Delta$.*

This result is best possible in some sense. For instance, the value of p is tight up to a constant factor since for $p = o(\frac{1}{n})$ the size of the largest connected component of $G(n,p)$ is sublinear. The maximum degree condition is tight too, since for $p = O(\frac{1}{n})$ the degree of a typical vertex of $G(n,p)$ is roughly $pn = O(1)$. Finally, the constant $\frac{1}{2}$ is also sharp since one can delete a $(\frac{1}{2} + \delta)$-fraction of edges to every vertex in $G(n, \frac{C}{n})$ so that the largest connected component of the resulting graph has about $\frac{n}{2}$ vertices (see [13] for details).

Let us note that Theorem 1.4.2 is a random analogue of Komlós–Sárközy–Szemerédi's theorem (Theorem 1.2.1) for very sparse graphs. In this thesis we prove a *global* version of Theorem 1.4.2, which may be seen as a sparse random analogue of the Erdős–Sós conjecture. Namely, we prove the following theorem.

Theorem 1.4.3 (Araújo, Moreira, and P. [10]). *For every $r, \Delta \geqslant 2$ and $\delta \in (0,1)$, there exists $C > 0$ such that if $p \geqslant \frac{C}{n}$, then $G = G(n,p)$, with high probability, has the following property. Let $G' \subseteq G$ be a subgraph such that $e(G') \geqslant \left(\frac{1}{r} + \delta\right) p\binom{n}{2}$, then G' contains a copy of every tree T with $\frac{n}{r}$ edges such that $\Delta(T) \leqslant \Delta$.*

We point out that Theorem 1.4.3 is best possible in the same ways as Theorem 1.4.2 is tight. Again, the value of p is tight up to a constant factor since for smaller values of p the largest connected component has sublinear size, and one cannot hope to find trees of higher degree for $p = O(\frac{1}{n})$. Moreover, the $\frac{1}{r}$ factor cannot be improved. Indeed, one can partition the vertex set in $r+1$ parts such that the smaller part has at most r vertices and the others r parts have the same number of vertices, and thus fewer than $\frac{n}{r}$. Then, with high probability, the graph $G' \subseteq G(n,p)$ obtained by removing edges between parts has $(\frac{1}{r} - o(1))p\binom{n}{2}$ edges but every connected component of G' has less than $\frac{n}{r}$ vertices.

We wonder if Theorem 1.4.3 holds for smaller trees as well. It is tempting to conjecture that, for a reasonable p and a tree T with bounded degree, if $G' \subseteq G(n,p)$ is a subgraph with $e(G') \geqslant (1+o(1))p|T|\frac{n}{2}$, then G' contains a copy of T. We observe that Theorem 1.4.3 shows that this conjecture holds for trees with linear size, however, we believe that for smaller trees this problem might be quite hard. A more tractable question is the following.

Problem 1.4.4. *Let $f : \mathbb{N} \to \mathbb{N}$ be a function such that $f(n) = o(n)$, and let T be a tree on $f(n)$ vertices and maximum degree bounded by some fixed constant Δ. Determine $p = p(n)$ and a constant $C > 0$ so that, with high probability, every subgraph $G' \subseteq G(n,p)$ with $e(G') \geqslant Cpf(n)n$ contains a copy of T.*

When $f(n)$ is a constant function, we believe that this problem follows by an application of the *hypergraph container method* (see [15] for a survey). However, it is not clear what happens if $f(n)$ grows with n.

Chapter 2

Preliminaries

2.1 Basic notation

Given a positive integer $\ell \in \mathbb{N}$, we write $[\ell] = \{1, \ldots, \ell\}$. Also, we will write $a \ll b$ to indicate that given b, we choose a significantly smaller than b. The value for such a can be explicitly calculated from the proofs, but sometimes we will prefer to omit it for clarity of the presentation. For real numbers a, b, x, we write $a = b \pm x$ if $a \in [b - x, b + x]$. Given a set S and an integer $0 \leqslant k \leqslant |S|$, we denote by $\binom{S}{k}$ the collection of all subsets of S of size k.

A *graph* is a pair $G = (V, E)$, where V is the set of vertices of G and $E \subseteq \binom{V}{2}$ is the set of edges of G. If it is not specified, we write $V(G)$ and $E(G)$ for vertex set and edge set of G, respectively. We say that a graph G is *bipartite* if there exists a partition $V(G) = A \cup B$ such that each edge $e \in E(G)$ has one endpoint in A and the other in B. If G is bipartite, we will write $G = (A, B)$ to refer that G has a bipartition $V(G) = A \cup B$.

Given a graph H, we write $|H| = |V(H)|$ for its number of vertices and $e(H) = |E(H)|$ for the number of edges of H. We write $\delta(H)$, $d(H)$, and $\Delta(H)$, for the minimum, average, and maximum degree of H, respectively. As usual, $\deg_H(x)$ denotes the degree of a vertex $x \in V(H)$, and we write $N_H(x)$ for the set of neighbours of x. Moreover, given a set $S \subseteq V(H)$, we write $N_H(x, S) = N_H(x) \cap S$ for the neighbourhood of x in S and $\deg_H(x, S)$ for the respective degree. For two disjoint sets $X, Y \subseteq V(H)$, we write $E_H(X, Y)$ for the set of edges $xy \in E(H)$ such that $x \in X$ and $y \in Y$, and we set $e_H(X, Y) := |E_H(X, Y)|$. In all of the above, we omit the subscript H if it is clear from the context.

Given a set $U \subseteq V(H)$, we write $H[U]$ for the graph induced in H by the set U, that is, the vertex set of $H[U]$ is U and the edge set corresponds to all edges having both endpoints in U. For two disjoint sets $X, Y \subseteq V(H)$, we write $H[X, Y]$ for the bipartite graph induced in H by X and Y. We say a vertex x *sees* a set $U \subseteq V(H)$ if it sends at least one edge to U.

Given a collection of sets $\mathcal{F}$, we write $\bigcup \mathcal{F}$ for the union of all members of $\mathcal{F}$. For instance, if $\mathcal{F} = \{F_1, \ldots, F_m\}$ then $\bigcup \mathcal{F} = F_1 \cup \cdots \cup F_m$. If $\mathcal{G}$ is a collection of graphs, then $\bigcup \mathcal{G}$ denotes the graph which is the union of all graphs in $\mathcal{G}$.

Given two graphs F and G, a *homomorphism* from F to G is a function $\varphi : V(F) \to V(G)$ which preserves adjacency, that is, if for every edge $e \in E(F)$ we have $\varphi(e) \in E(G)$. An *embedding* φ of F into G is an injective homomorphism from F to G and we say that F *embeds* into G if there exists an embedding $\varphi : V(F) \to V(G)$. Moreover, we say that F is a *subgraph* of G if F embeds into G.

2.2 Trees

Let us go through some notation for trees. We will write (T, r) for a tree T rooted at a vertex $r \in V(T)$. Given a rooted tree (T, r) and vertices $x, y \in V(T)$, we say that x is *below* y (resp. y is *above* x) if x lies on the unique path from y to r (our trees grow from the top to the bottom). If in addition $xy \in E(T)$, we say y is a *child* of x, and x is the *parent* of y. We note that this defines a partial order on the vertex set of T. The *tree induced by* x, denoted by $T(x)$, is the subtree of T induced by all vertices above x. For $i \geqslant 0$, the i-th level of T, denoted by L_i, consists of all vertices at distance i from r.

We say that a vertex of a tree is a *leaf* if it has degree 1. A *bare path* in a tree is a path all whose internal vertices have degree 2 in the tree. The next lemma has been extensively used in the literature of tree embeddings, as it states that the structure of any given tree satisfies a certain dichotomy. Namely, each tree contains either a large number of leaves or a large number of bare paths of some fixed constant length (we refer to [76, 88] for a more general statement and a proof, and note that here and elsewhere, the length of a path is its number of edges).

Lemma 2.2.1 (Lemma 2.1 from [88]). *Let $\ell > 2$ and let T be a tree. Then either T has at least $|T|/4\ell$ leaves or it has at least $|T|/4\ell$ vertex disjoint bare paths, each of length ℓ.*

Trees are bipartite graphs whose bipartition classes may be as imbalanced as possible. For instance, a path of length k has colour classes of size differing in at most 1, and a star with k edges has a colour class of size 1 and the other class of size k. Nevertheless, for trees having maximum degree bounded by a constant one can guarantee that both colour classes have linear size.

Fact 2.2.2. *Let $\Delta \geqslant 2$ and let T be a tree with bipartition $V(T) = C \cup D$ and maximum degree $\Delta(T) \leqslant \Delta$. Then* $\min\{|C|, |D|\} \geqslant \frac{k}{\Delta}$.

2.2.1 Basic results on tree embedding

As we mentioned before, a greedy argument shows that every k-edge tree can be embedded into any graph of minimum degree at least k. We now give two lemmas that generalise this simple observation.

Lemma 2.2.3. *Let $\Delta, h, k \in \mathbb{N}$, let (T, r) be a tree with $k - h$ edges and $\Delta(T) \leqslant \Delta$, and let G be a graph satisfying* *(i)* $\delta(G) \geqslant \Delta + h$*;*

(ii) there are at most h *vertices* $x \in V(G)$ *with* $\deg(x) < k$*.*

Then T *can be embedded in* G*. Moreover, any vertex* $v \in G$ *can be chosen as the image of* r*.*

Proof. We construct an embedding ϕ as follows. We set $\phi(r) := v$. Since $\deg(v) \geqslant \Delta + h$, we can embed each neighbour of r into a neighbour of v that has degree at least k. Since T has $k - h$ vertices, we can then embed the rest of T levelwise using only vertices of degree at least k at each step. □

Observe that for $h = 0$ Lemma 2.2.3 recovers the greedy procedure we mentioned above. Moreover, if the host graph G is bipartite, one can relax the minimum degree condition for one side of the bipartition of G. The proof of the following result is a straightforward modification of the proof of Lemma 2.2.3.

Lemma 2.2.4. *Let* $\Delta, h, k_1, k_2 \in \mathbb{N}$*, let* (T, r) *be a tree with colour classes* C, D *of sizes* $k_1 - h$ *and* $k_2 - h$*, respectively, and* $\Delta(T) \leqslant \Delta$*. Let* $G = (A, B)$ *be a bipartite graph such that*

(i) $\delta(G) \geqslant \Delta + h$*;*

(ii) there are at most h *vertices* $a \in A$ *with* $\deg(x) < k_2$*;*

(iii) there are at most h *vertices* $b \in B$ *with* $\deg(x) < k_1$*.*

Then T *can be embedded into* G *with* C *going to* A *and* D *going to* B*. Moreover, if* $r \in C$ *(resp.* D*), then any vertex* $a \in A$ *(resp.* $b \in B$*) can be chosen as the image of* r*.*

2.2.2 Cutting trees

In this section, we present some results regarding how to cut a tree into small pieces. The first two results allow us to find a cut vertex which split a tree into subtrees of controlled sizes. On the other hand, the last result states that any large enough tree can be decomposed into a bounded family of small subtrees.

Lemma 2.2.5. *Let* T *be a tree on* $k+1$ *vertices, and let* x *be a leaf of* T*. Then* T *has a vertex* z *such that every component of* $T - z$ *has at most* $\lfloor \frac{k}{2} \rfloor$ *vertices, except the one containing* x*, which has at most* $\lceil \frac{k}{2} \rceil$ *vertices.*

Proof. Let z be a maximal vertex, with respect to the order in (T, x), such that $|T(z)| > \lfloor \frac{k}{2} \rfloor$. Then every component of $T - z$ has at most $\lfloor \frac{k}{2} \rfloor$ vertices: This is obvious (from the definition of z) for the components not containing x, while the component that contains x only has $|T| - |T(z)| \leqslant k + 1 - (\lfloor \frac{k}{2} \rfloor + 1) = \lceil \frac{k}{2} \rceil$ vertices. □

The following lemma is a generalisation of Lemma 2.2.5, stating that one can control even more the size of at least one of the components (see [61] for other variants).

Lemma 2.2.6. *For all $0 < \gamma \leqslant 1$ and for all $k \geqslant \frac{200}{\gamma}$, any given tree T with k edges has a subtree (T^*, t^*) such that*

(i) $\frac{\gamma k}{2} \leqslant |V(T^*)| \leqslant \gamma k$; *and*

(ii) every component of $T - T^*$ *is adjacent to* t^*.

Proof. Let $r \in V(T)$ be an arbitrary vertex and let $t^* \in V(T)$ be a maximal vertex, with respect to the order in (T, r), such that $|T(t^*)| \geqslant \frac{\gamma k}{2}$. Note that, by maximality of t^*, every child u of t^* satisfies $|T(u)| < \frac{\gamma k}{2}$. Let U be a minimal subset of the children of t^* such that $|\bigcup_{u \in U} T(u)| \geqslant \gamma k/2$. Then the tree T^* induced by t^* and $\bigcup_{u \in U} T(u)$ satisfies the desired properties. □

By iteratively applying Lemma 2.2.6, we can show that any tree can be decomposed into a bounded family of arbitrary small subtrees. Versions of this result have already appeared in earlier literature on tree embeddings, see for instance [2].

Proposition 2.2.7. *Let $\beta \in (0, 1)$ and let (T, r) be a rooted tree with $k > \beta^{-1}$ edges. Then there exists a set $S \subseteq V(T)$ and a family $\mathcal{P}$ of disjoint rooted trees such that*

1. $r \in S$;
2. $\mathcal{P}$ *consists of the components of* $T - S$, *and each* $P \in \mathcal{P}$ *is rooted at the vertex closest to the root of* T;
3. $|P| \leqslant \beta k$ *for each* $P \in \mathcal{P}$; *and*
4. $|S| < \frac{1}{\beta} + 2$.

The vertices from S will be called seeds, and the components from $\mathcal{P}$ will be called the pieces of the decomposition.

Proof. We iteratively construct the set S, starting with $T^0 := T$ and $S^0 := \emptyset$. At step $i + 1$, let s_{i+1} be a maximal vertex of T^i (with respect to the order induced by r) such that

$$|T^i(s_{i+1})| > \beta k.$$

Note that by the maximality of s_{i+1} the trees in $T^i(s_{i+1}) - s_{i+1}$ each cover at most βk vertices. We obtain S^{i+1} by adding s_{i+1} to S^i and set $T^{i+1} = T^i - T^i(s_{i+1})$. If at some step j there is no vertex s_{j+1} with $|T^j(s_{j+1})| > \beta k$, then $|T^j| \leqslant \beta k$ and we end the process. We set $S := S^j \cup \{r\}$ and let $\mathcal{P}$ be the set of connected components of $T - S$. Note that Properties (1)–(3) clearly hold. For (4), we observe that $|T^{i+1}| < |T^i| - \beta k$ and hence

$$0 \leqslant |T^m| < |T^0| - j \cdot \beta k,$$

which implies that $|S| = j + 1 \leqslant \frac{|T|}{\beta k} + 1 < \frac{1}{\beta} + 2.$ □

2.3 The Regularity Lemma

The celebrated *Szemerédi's regularity lemma* [105] is one of the most powerful tools in extremal graph theory. It states that the vertex set of every large enough graph can be partitioned into finitely many parts so that most of the pairs of these parts induces a bipartite *quasi-random* graph. In order to state this result, let us begin with some definitions.

Let $H = (A, B)$ be a bipartite graph with density

$$d(A, B) := \frac{e(A, B)}{|A||B|}.$$

For a fixed $\varepsilon > 0$, the pair (A, B) is said to be *ε-regular* if for any $X \subseteq A$ and $Y \subseteq B$, with $|X| \geqslant \varepsilon|A|$ and $|Y| \geqslant \varepsilon|B|$, we have

$$|d(X, Y) - d(A, B)| \leqslant \varepsilon.$$

Moreover, an ε-regular pair (A, B) is called (ε, η)-regular if $d(A, B) \geqslant \eta$. Given an ε-regular pair (A, B), we say that $X \subseteq A$ is *ε-significant* if $|X| \geqslant \varepsilon|A|$, and similarly for subsets of B. A vertex $x \in A$ is called *ε-typical* to a significant set $Y \subseteq B$ if $\deg(x, Y) \geqslant (d(A, B) - \varepsilon)|Y|$. We simply write *regular*, *significant* or *typical* if ε is clear from the context.

It it well known that regular pairs behave, in many ways, like random bipartite graphs with the same edge density. The next well known fact (see for instance [74]) states that in a regular pair almost every vertex is typical to any given significant set, and also that regularity is inherited by subpairs.

Fact 2.3.1. *Let (A, B) be an ε-regular pair with density η. Then the following holds:*

1. *For any ε-significant $Y \subseteq B$, all but at most $\varepsilon|A|$ vertices from A are ε-typical to Y.*
2. *Let $\alpha \in (0, 1)$. For any subsets $X \subseteq A$ and $Y \subseteq B$, with $|X| \geqslant \alpha|A|$ and $|Y| \geqslant \alpha|B|$, the pair (X, Y) is $\frac{2\varepsilon}{\alpha}$-regular with density between $\eta - \varepsilon$ and $\eta + \varepsilon$.*

The regularity lemma of Szemerédi states that, for any given $\varepsilon > 0$, the vertex set of any large enough graph can be partitioned into a bounded number of sets, also called *clusters*, such that the graph induced by almost any pair of these clusters is ε-regular. We will make use of the well known degree form of the regularity lemma (see for instance [77]). Call a vertex partition $V(G) = V_1 \cup \cdots \cup V_\ell$ an *(ε, η)-regular partition* if

1. $|V_1| = |V_2| = \cdots = |V_\ell|$;
2. V_i is independent for all $i \in [\ell]$; and
3. for all $1 \leqslant i < j \leqslant \ell$, the pair (V_i, V_j) is ε-regular with density either $d(V_i, V_j) \geqslant \eta$ or $d(V_i, V_j) = 0$.

Lemma 2.3.2 (Lemma 41 from [77]). *For all $\varepsilon > 0$ and $m_0 \in \mathbb{N}$ there are N_0, M_0 such that the following holds for all $\eta \in [0, 1]$ and $n \geqslant N_0$. Any n-vertex graph G has a subgraph G', with $|G| - |G'| \leqslant \varepsilon n$ and $\deg_{G'}(x) \geqslant \deg_G(x) - (\eta + \varepsilon)n$ for all $x \in V(G')$, such that G' admits an (ε, η)-regular partition $V(G') = V_1 \cup \cdots \cup V_\ell$, with $m_0 \leqslant \ell \leqslant M_0$.*

The (ε, η)-*reduced graph* $\mathcal{R}$ corresponding to the (ε, η)-regular partition given by Lemma 2.3.2 has vertex set $V(\mathcal{R}) = \{V_i : i \in [\ell]\}$ in which V_iV_j is an edge if and only if $d(V_i, V_j) \geqslant \eta$. Henceforth, we will use calligraphic letters to refer to the reduced graph, or to subsets of its vertex set. Moreover, given $\mathcal{C} \subseteq V(\mathcal{R})$, we write $|\mathcal{C}|$ for the number of clusters in $\mathcal{C}$. In contrast, we write $|\bigcup \mathcal{C}|$ for the number of vertices of the subgraph $\bigcup \mathcal{C}$ of G. Now we state some useful facts about the reduced graph (see [74] for a proof).

Fact 2.3.3. *Let G be a graph on n vertices and let $\mathcal{R}$ be an (ε, η)-reduced graph of G. Then the following holds.*

(i) Given a cluster $C \in V(\mathcal{R})$ we have

$$\deg_{\mathcal{R}}(C) \geqslant \frac{1}{|C|} \sum_{v \in C} \deg(v) \cdot \frac{|\mathcal{R}|}{n}.$$

In particular, summing over all clusters we have $d(\mathcal{R}) \geqslant d(G) \cdot \frac{|\mathcal{R}|}{n}$.

(ii) Let $\mathcal{Y}$ be a collection of significant sets in G and let $C \in V(\mathcal{R})$. Then

$$|\{Y \in \mathcal{Y} : v \textit{ is typical to } Y\}| \geqslant (1 - \sqrt{\varepsilon})|\mathcal{Y}|$$

for all but at most $\sqrt{\varepsilon}|C|$ vertices $v \in C$.

We close this section with a lemma that illustrates why regularity is useful for embedding trees. It states that a tree will always fit into a regular pair, if the tree is small enough.

Lemma 2.3.4. *Let $0 < \beta \leqslant \varepsilon \leqslant \frac{1}{25}$. Let (A, B) be a $(\varepsilon, 5\sqrt{\varepsilon})$-regular pair with $|A| = |B| = m$, and let $X \subseteq A, Y \subseteq B, Z \subseteq A \cup B$ be such that $\min\{|X \setminus Z|, |Y \setminus Z|\} > \sqrt{\varepsilon}m$. Then any tree T on at most βm vertices can be embedded into $(X \cup Y) \setminus Z$. Moreover, for each $v \in V(T)$ there are at least $2\varepsilon m$ vertices from $(X \cup Y) \setminus Z$ that can be chosen as the image of v.*

Proof. We construct the embedding $\phi : V(T) \to X \cup Y$ levelwise, starting with the root, which is embedded into a typical vertex of $(X \cup Y) \setminus Z$. At each step i we ensure that all vertices of level i are embedded into vertices of $X \setminus Z$ (or $Y \setminus Z$) that are typical with respect to the unoccupied vertices of $Y \setminus Z$ (or $X \setminus Z$). This is possible, because at each step i, and for each vertex v of level i, the degree of a typical vertex into the unoccupied vertices on the other side is at least $4\varepsilon m$, and there are at most εm non typical vertices and at most $|T| \leqslant \beta m$ already occupied vertices. $\square$

Chapter 3

Embedding trees with maximum and minimum degree conditions

This chapter is based on joint work with Guido Besomi and Maya Stein [17, 19].

In this chapter, we will prove a series of results regarding maximum and minimum degree conditions that ensure the containment of every tree with maximum degree bounded by certain function. Namely, we will prove Theorems 1.3.5,1.3.6, 1.3.3, and 1.3.8. Most of our results rely on our *key embedding lemma* (Lemma 3.5.3) and thus let us start by describing this lemma, which will be stated and proved in Section 3.5.

Lemma 3.5.3 provides an embedding of any tree T with maximum degree bounded by $k^{\frac{1}{r}}$, where r is a constant, into any host graph G of suitable minimum degree, as long as G contains one of several favourable scenarios explicitly described in the statement of Lemma 3.5.3. The scenarios contemplated by the lemma cover the situation where, after applying the regularity lemma to G, the corresponding reduced graph has a large[1] component, but also cover a number of situations where there is no large component. In these latter situations, we will have to use a maximum degree vertex x of G, as well as a suitable cut vertex z of T, and embed the components of $T - z$ into components of $G - x$. Several possible shapes and sizes of components possibly seen by x are taken into account in Lemma 3.5.3.

Once we have Lemma 3.5.3, the proof of Theorems 1.3.5, 1.3.6 and 1.3.3 will be fairly easy. We only need to regularise the host graph G and then show that we are in one of the situations as described in Lemma 3.5.3. This is done in Section 3.6.

Let us now sketch the proof of our key embedding lemma. There are two crucial ingredients for the proof of Lemma 3.5.3. One of these ingredients is some work that we accomplish in Section 3.2. In that section, we prove some useful results on cutting trees, the most important ones being Lemma 3.2.1 and Proposition 3.2.5. These two auxiliary results allow us to cut a tree at some convenient cut vertex z and then group the components of $T - z$ into two or three groups (as necessary), so that the union of the components from these groups form sets of convenient sizes. Moreover, we show that it is possible to 2-colour the vertices

[1]That is, large enough to accommodate T.

of $T - z$ in a way that the resulting colour classes are not too unbalanced. This will be very important when, in the context of Lemma 3.5.3, we wish to embed several components of $T - z$ into a single bipartite component of the reduced graph of $G - x$.

The other crucial ingredient for the proof of Lemma 3.5.3 is the preparatory work accomplished in Sections 3.3 and 3.4. There we show how to embed a tree into a host graph that, after an application of the regularity lemma, has a reduced graph with a large connected component. For this, we cut the tree into tiny subtrees and few connecting vertices, and then embed these trees into suitable edges of the reduced graph. This approach has been used earlier in the literature, see for instance [2]. The only remaining problem is how to make the connections between the tiny trees.

For these connections, we use paths in the reduced graph. For this argument to work, we have to bound the maximum degree of the tree we wish to embed in terms of the diameter of the reduced graph of G (another argument will allow us to relax the bound later, see below). Also, we have to distinguish two cases, namely whether the large component of the reduced graph is bipartite or not. If it is bipartite, we embed the larger colour class of the tree into the larger side of the component and, since the smaller colour class of the tree is smaller than the minimum degree of G, each tiny tree can be embedded into an *unsaturated* edge. If the component is non-bipartite, we can find a large connected matching (see Lemma 3.3.8) that can be filled, in a balanced way, with tiny trees.

The two cases will be treated in Propositions 3.3.1 and 3.3.9, respectively. In the remainder of Section 3.3, we deduce some corollaries from these propositions, which will come in handy later when, in the proof of Lemma 3.5.3, we need to embed parts of the tree into parts of the host graph that correspond to different components of its reduced graph.

In Section 3.4, we unify and improve the results from Section 3.3. Namely, in Proposition 3.4.3 we provide an embedding result for trees into large connected components of the reduced graph of G, where the bound on the maximum degree of the tree no longer depends on the diameter of the reduced graph of the host graph, but instead is $k^{\frac{1}{r}}$, where r is an absolute constant. The idea for the proof of this result is that we first try to follow the embedding scheme from the previous section, but only using paths of bounded length for the connections. If this fails, then the only possible reason is that we could not reach suitable free space at a bounded distance from the cluster C we were currently embedding into. In this case, we abort our mission, and we are able to prove that it is possible to embed the tree into a ball of appropriate radius centered at C.

Finally, in Section 3.7 we prove Theorem 1.3.8. The proof of Theorem 1.3.8 is more involved and does not directly follows from Lemma 3.5.3. The proof is based on a structural result for graphs with minimum degree above $\frac{k}{2}$ and maximum degree above $\frac{4k}{3}$ avoiding some tree with k edges and bounded degree (Theorem 3.7.2).

3.1 Sharpness of Conjecture 1.3.7

This section is devoted to show the asymptotical tightness of our conjecture for infinitely many values of α. Namely, we will prove the following result.

Proposition 3.1.1. *For all odd $\ell \in \mathbb{N}$ with $\ell \geqslant 3$, and for all $\gamma > 0$ there are $k \in \mathbb{N}$, a k-edge tree T, and a graph G with $\delta(G) \geqslant (1 + \frac{1}{\ell} - \gamma)\frac{k}{2}$ and $\Delta(G) \geqslant 2(1 - \frac{1}{\ell} - \gamma)k$ such that T does not embed in G.*

In order to be able to prove Proposition 3.1.1, let us consider the following example.

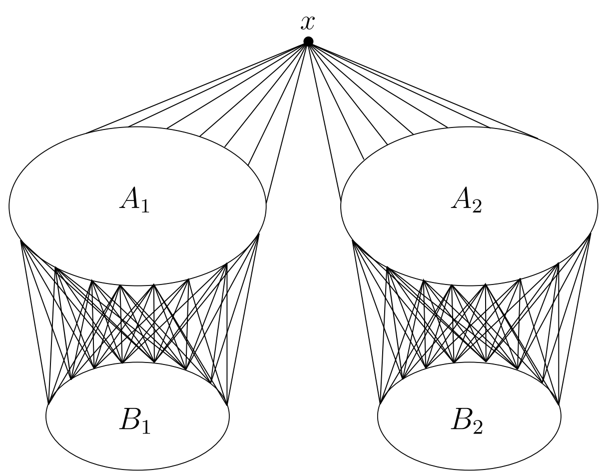

Figure 3.1: The graph $H_{k,\ell,c}$ from Example 3.1.2

Example 3.1.2. *Let $\ell, k, c \in \mathbb{N}$ with $1 \leqslant c \leqslant \frac{k}{\ell(\ell+1)}$ such that $\ell \geqslant 3$ is odd and divides k. For $i \in \{1, 2\}$, we define $H_i = (A_i, B_i)$ to be the complete bipartite graph with*

$$|A_i| = (\ell - 1)\left(\frac{k}{\ell} - 1\right) \text{ and } |B_i| = \frac{k}{2} + \frac{(c-1)(\ell+1)}{2} - 1.$$

We obtain $H_{k,\ell,c}$ by adding a new vertex x to $H_1 \cup H_2$, and adding all edges between x and $A_1 \cup A_2$. Observe that

$$\delta(H_{k,\ell,c}) = \min\{|A_1|, |B_1| + 1\} = |B_1| + 1 = \frac{k}{2} + \frac{(c-1)(\ell+1)}{2}$$

and

$$\Delta(H_{k,\ell,c}) = |A_1 \cup A_2| = 2(\ell - 1)\left(\frac{k}{\ell} - 1\right).$$

Let $T_{k,\ell}$ be the tree formed by ℓ stars of order $\frac{k}{\ell}$ and an additional vertex v connected to the centres of the stars.

We will use Example 3.1.2 to prove Proposition 3.1.1. However, a similar proposition (with slightly weaker bounds) could be obtained by replacing one of the graphs H_i from Example 3.1.2 with a small complete graph. See Example 3.1.5 near the end of this section.

Let us now show that the graph $H_{k,\ell,c}$ from Example 3.1.2 does not contain the tree $T_{k,\ell}$.

Lemma 3.1.3. *For all $\ell, k, c \in \mathbb{N}$ with $1 \leqslant c \leqslant \frac{k}{\ell(\ell+1)}$ such that $\ell \geqslant 3$ is odd and divides k, the tree $T_{k,\ell}$ from Example 3.1.2 does not embed in the graph $H_{k,\ell,c}$.*

Proof. Observe that we cannot embed $T_{k,\ell}$ in $H_{k,\ell,c}$ by mapping v into x, since then, one of the sets B_i would have to accommodate all leaves of at least $\frac{\ell+1}{2}$ of the stars of order $\frac{k}{\ell}$. But these are at least

$$\frac{\ell+1}{2} \cdot \left(\frac{k}{\ell} - 1\right) = \frac{k}{2} + \frac{1}{2\ell}(k - \ell(\ell+1)) \geqslant \frac{k}{2} + \frac{1}{2}(c-1)(\ell+1) > |B_i|$$

leaves in total, so they will not fit into B_i.

Moreover, we cannot map v into one of the H_i, because then we would have to embed at least $\ell - 1$ stars into H_i. The leaves of these stars would have to go to the same side as v, but together these are

$$(\ell-1)\left(\frac{k}{\ell} - 1\right) + 1 > |A_i| \geqslant |B_i|$$

vertices (note that we count v), so this, too, is impossible. We conclude that the tree $T_{k,\ell}$ does not embed in $H_{k,\ell,c}$. □

Before we prove Proposition 3.1.1, let us state a weaker result which, in particular, proves the tightness of Conjecture 1.3.4.

Proposition 3.1.4. *For all $\alpha \in (0, \frac{1}{2})$ there are $k \in \mathbb{N}$, a k-edge tree T, and a graph G with $\delta(G) = \frac{k}{2}$ and $\Delta(G) \geqslant 2(1-\alpha)k$ such that T does not embed in G.*

Proof. Given $\alpha \in (0, 1)$, we set $\ell := 2\lceil \frac{1}{\alpha} \rceil - 1$. Then $\ell \geqslant 3$ is odd, and we can consider the tree $T_{k,\ell}$ and the graph $H_{k,\ell,c}$ from Example 3.1.2, where we take $k := \ell(\ell+1)$ and $c := 1$. By Lemma 3.1.3, we know that $T_{k,\ell}$ does not embed in $H_{k,\ell,c}$.

Observe that $\delta(H_{k,\ell,c}) = \frac{k}{2}$ and, by our choice of k we have

$$\Delta(H_{k,\ell,c}) = 2(\ell-1)\Big(\frac{1}{\ell} - \frac{1}{k}\Big)k = 2\Big(1 - \frac{2}{\ell+1}\Big)k,$$

and therefore, $\Delta(H_{k,\ell,c}) \geqslant 2(1-\alpha)k$, which is as desired. □

Let us now prove Proposition 3.1.1. For this, we will let the constant c go to infinity.

Proof of Proposition 3.1.1. Let ℓ and γ be given. For any fixed integer $c \geqslant 1$, set

$$k := c\ell(\ell+1),$$

and consider the tree $T_{k,\ell}$ and the host graph $H_{k,\ell,c}$ from Example 3.1.2 for parameters k, ℓ and c. Observe that

$$\delta(H_{k,\ell,c}) > \Big(1 + \frac{(c-1)(\ell+1)}{k}\Big)\frac{k}{2} = \Big(1 + \frac{c-1}{c\ell}\Big)\frac{k}{2} = \Big(1 + \frac{1}{\ell} - \frac{1}{c\ell}\Big)\frac{k}{2}$$

and

$$\Delta(H_{k,\ell,c}) = 2\Big(1 - \frac{1}{\ell} - \frac{\ell-1}{k}\Big)k > 2\Big(1 - \frac{1}{\ell} - \frac{1}{c\ell}\Big)k.$$

So, for any given γ we can choose c large enough such that

$$\delta(H_{k,\ell,c}) \geqslant \Big(1 + \frac{1}{\ell} - \gamma\Big)\frac{k}{2} \quad \text{and} \quad \Delta(H_{k,\ell,c}) \geqslant 2\Big(1 - \frac{1}{\ell} - \gamma\Big)k,$$

which is as desired, since by Lemma 3.1.3, we know that $T_{k,\ell}$ does not embed in $H_{k,\ell,c}$, $\square$

Let us now quickly discuss an alternative example, which gives worse bounds than the ones given in Proposition 3.1.1, but might be interesting because of its different structure.

Example 3.1.5. *Let k, ℓ, c be as in Example 3.1.2. Let C be a complete graph of order $\frac{k}{2} + \frac{(c-1)(\ell+1)}{2}$. Let $G_{k,\ell,c}$ be obtained by taking C and the bipartite graph $H_1 = (A_1, B_1)$ from Example 3.1.2, and joining a new vertex x to all vertices from A_1 and to all vertices in C. Then $\delta(G_{k,\ell,c}) = \frac{k}{2} + \frac{(c-1)(\ell+1)}{2}$ and $\Delta(G_{k,\ell,c}) = \frac{3\ell-2}{2\ell}k + \frac{(c-3)(\ell+1)}{2} - 2$, and an analogue of Lemma 3.1.3 holds.*

Moreover, in the same way as in the proof of Proposition 3.1.1, we can show that if k is large enough in terms of (odd) $\ell \geqslant 3$ and γ, then

$$\delta(G_{k,\ell,c}) \geqslant (1 + \frac{1}{\ell} - \gamma)\frac{k}{2} \ \ \textit{and} \ \ \Delta(G_{k,\ell,c}) \geqslant \frac{3}{2}(1 - \frac{1}{\ell} - \gamma)k.$$

This example, as well as the examples underlying Propositions 3.1.4 and 3.1.1 illustrate that requiring a maximum degree of at least ck, for any $c < 2$, and a minimum degree of at least $\frac{k}{2}$ is not enough to guarantee that any graph obeying these conditions contains all k-edge tree as subgraphs. Nevertheless, we could not come up with any radically different examples, and it might be that graphs that look very much like the graph $H_{k,\ell,c}$ from Example 3.1.2 or the graph $G_{k,\ell,c}$ from Example 3.1.5 are the only obstructions for embedding all k-edge trees.

To finish this section, let us discuss about the values of α not covered in Proposition 3.1.1. For any $\alpha \in [0, \frac{1}{3})$ and $\gamma > 0$ small, we can construct examples of graphs with minimum degree at least $(1 + \alpha - \gamma)\frac{k}{2}$ and maximum degree at least $2(1 - g(\alpha) - \gamma)k$, where $g(\alpha)$ is a function which is bigger than α but reasonably close to it. In particular, $g(\alpha)$ satisfies $|\alpha - g(\alpha)| = O(\alpha^2)$, and, more explicitly, for any even $\ell \geqslant 3$ we obtain $g(\frac{1}{\ell}) = \frac{1}{\ell} + \frac{1}{\ell(\ell-2)}$. These examples are very similar to Example 3.1.2. The difference is that the small stars that make up the tree may have different sizes (more precisely, one star is smaller than the other ones). The host graph is the same, with slightly adjusted size of the sets A_i.

3.2 Finding a good cut vertex

In this section, we prove a series of results regarding how to find a cut vertex z in a tree T and a 2-colouring of the vertices of $T - z$ such that both colour classes have controlled size. This will be particularly useful when embedding trees into bipartite graphs.

We first prove an auxiliary lemma on partitioning sequences of integers. This lemma will be used in the proofs of both Lemma 3.2.3 and Proposition 3.2.5, and also in the proofs of Theorems 1.3.5 and 1.3.6.

Lemma 3.2.1. *Let $m, t \in \mathbb{N}_+$ and let $\{a_i\}_{i=1}^m$ be a sequence of positive integers such that $0 < a_i \leqslant \lceil \frac{t}{2} \rceil$, for each $i \in [m]$, and $\sum_{i=1}^m a_i \leqslant t$. Then*

1. *there is a partition $\{I_1, I_2, I_3\}$ of $[m]$ such that $\sum_{i \in I_3} a_i \leqslant \sum_{i \in I_2} a_i \leqslant \sum_{i \in I_1} a_i \leqslant \lceil \frac{t}{2} \rceil$; and*
2. *there is a partition $\{J_1, J_2\}$ of $[m]$ such that $\sum_{i \in J_2} a_i \leqslant \sum_{i \in J_1} a_i \leqslant \frac{2}{3} t$.*

Proof. We first pick a set $I_1 \subseteq [m]$ with $\sum_{i \in I_1} a_i \leqslant \lceil \frac{t}{2} \rceil$ that maximises the sum. From $[m] \setminus I_1$ we extract a second set I_2 with $\sum_{i \in I_2} a_i \leqslant \lceil \frac{t}{2} \rceil$ that maximises the sum. The choice of I_1 and I_2 ensures that for $I_3 := [m] \setminus (I_1 \cup I_2)$ it also holds that $\sum_{i \in I_3} a_i \leqslant \lceil \frac{t}{2} \rceil$, and that $\sum_{i \in I_3} a_i \leqslant \sum_{i \in I_2} a_i$. Therefore, the sets I_1, I_2, I_3 fulfil the conditions in (i). (Notice that I_3, and possibly also I_2, may be empty.)

For (ii) we proceed as follows. If $I_3 = \emptyset$ we just set $J_1 := I_1$ and $J_2 := I_2$, which clearly satisfies (ii). If $I_3 \neq \emptyset$ we define J_1 as one of the sets $I_2 \cup I_3$ and I_1, and J_2 as the other set, in a way that $\sum_{i \in J_2} a_i \leqslant \sum_{i \in J_1} a_i$. Observe that the second part of (i) implies that $\sum_{i \in I_2 \cup I_3} a_i \leqslant \frac{2}{3} t$. So again, (ii) is satisfied. □

Remark 3.2.2. *Observe that the set I_3 from Lemma 3.2.1 (i) has at most one element, because otherwise, due to the maximality of I_1 and I_2, there would exist $j, k \in I_3$ such that $a_j + \sum_{i \in I_1} a_i > \lceil \frac{t}{2} \rceil$ and $a_k + \sum_{i \in I_2} a_i > \lceil \frac{t}{2} \rceil$, a contradiction to the fact that $\sum_{i=1}^m a_i \leqslant t$.*

Lemma 3.2.1 tells us that after using Lemma 2.2.5 to cut a tree T at a vertex z, we can group the components of $T - z$ in such a way that the total size of each group is conveniently bounded. We would now like to say something about the balancedness of the resulting forest, and for this we resort to the concept of vertex colouring.

For a proper 2-colouring $c : V(G) \to \{0, 1\}$ of a graph G, with colours 0 and 1, we define

$$c_0 := \{v \in V(G) \,:\, c(v) = 0\} \quad \text{and} \quad c_1 := \{v \in V(G) \,:\, c(v) = 1\}.$$

For better readability, throughout all proofs we will stick to the convention that $|c_0| \geqslant |c_1|$ (but this will be restated in each proof).

Lemma 3.2.3. *Every tree T with t edges has a cut vertex z such that $T - z$ admits a proper 2-colouring $c : V(T - z) \to \{0, 1\}$ with $|c_0| \leqslant \frac{3t-1}{4}$ and $|c_1| \leqslant \frac{t}{2}$.*

Proof. We apply Lemma 2.2.5 to obtain a cut vertex z and a forest $T-z$ with components $\{T_i\}_{i=1}^m$ such that $|T_i| \leqslant \lceil \frac{t}{2} \rceil$, for every $i \in [m]$. We will now use Lemma 3.2.1 in order to group the components of $T-z$. Setting $a_i := |T_i|$, the lemma yields three sets I_1, I_2 and I_3 such that the forests $F_j := \bigcup_{i \in I_j} T_i$, with $j \in \{1,2,3\}$, cover at most $\lceil \frac{t}{2} \rceil$ vertices each. Also, the forest F_1 covers at least $\frac{t}{3}$ vertices.

For $j \in \{1,2,3\}$, consider any proper 2-colouring c^j of the forest F_j, with colour classes c_0^j and c_1^j, such that F_1 and F_2 each meet both colours (This is possible unless $|F_1|$ and/or $|F_2|$ is 1, and in that case we are done anyway). For each j, we assume that $|c_0^j| \geqslant |c_1^j|$.

We split the remainder of the proof into two cases.

Case 1: $|c_0^1| \geqslant \frac{3|F_1|-1}{4}$.

In this case, we define the colouring c by setting $c_0 := c_0^1 \cup c_1^2 \cup c_1^3$ and $c_1 := V(T-z) \setminus c_0 = c_1^1 \cup c_0^2 \cup c_0^3$. Then,

$$\begin{aligned}|c_0| = |c_0^1| + |c_1^2| + |c_1^3| &\leqslant |F_1| - 1 + \frac{|F_2|}{2} + \frac{|F_3|}{2} \\ &\leqslant \frac{t+1}{4} - 1 + \frac{t}{2} \\ &\leqslant \frac{3t-1}{4},\end{aligned}$$

where the second inequality follows from the equality $|F_1| - 1 + \frac{|F_2|}{2} + \frac{|F_3|}{2} = \frac{|F_1|}{2} - 1 + \frac{|T-z|}{2}$. Moreover,

$$|c_1| \leqslant t - |c_0^1| - |c_1^2| \leqslant t - \frac{3|F_1|-1}{4} - 1 \leqslant t - \frac{t-1}{4} - 1 \leqslant \frac{3t-1}{4},$$

where the penultimate inequality comes from the fact that $|F_1| \geqslant \frac{t}{3}$. Hence, $\max\{|c_0|, |c_1|\} \leqslant \frac{3t-1}{4}$; renaming the colour classes if necessary we get the desired result.

Case 2: $|c_0^1| < \frac{3|F_1|-1}{4}$.

In this case, we define the colouring c by setting $c_0 := c_0^1 \cup c_1^2 \cup c_0^3$ and $c_1 := V(T-z) \setminus c_0 = c_1^1 \cup c_0^2 \cup c_1^3$. Then,

$$\begin{aligned}|c_0| < \frac{3|F_1|-1}{4} + \frac{|F_2|}{2} + |F_3| &= \frac{|T-z|}{2} + \frac{|F_1| + 2|F_3| - 1}{4} \\ &\leqslant \frac{t}{2} + \frac{|F_1| + |F_2| + |F_3| - 1}{4} \\ &= \frac{3t-1}{4},\end{aligned}$$

and

$$|c_1| \leqslant \frac{|F_1|}{2} + |F_2| - 1 + \frac{|F_3|}{2} = \frac{t}{2} + \frac{|F_2|}{2} - 1 \leqslant \frac{3t-1}{4}.$$

Again, we obtain $\max\{|c_0|, |c_1|\} \leqslant \frac{3t-1}{4}$, swapping the colour classes if necessary. □

Let us remark that the bound $\frac{3t-1}{4}$, given by Lemma 3.2.3, is best possible if we insist that the cut vertex is as the given by Lemma 2.2.5. This is illustrated by the following example.

Let t be divisible by four and consider the tree obtained by identifying the central vertex of a star of order $\frac{t}{2}$ with an end vertex of a path of order $\frac{t}{2}+2$. Let z be the cut vertex provided by Lemma 2.2.5. Then z leaves exactly two components: a path of order $\frac{t}{2}$ and a star of order $\frac{t}{2}$. One of the colour classes of this forest necessarily contains $\frac{3t}{4}-1$ vertices.

Nevertheless, it is possible to cut the tree at a different cut vertex so that the resulting forest admits a significantly more balanced colouring than the one given by Lemma 3.2.3. This is the purpose of Proposition 3.2.5 below. Before we state this proposition, let us introduce some useful notation.

Definition 3.2.4 (Colouring imbalance)**.** *Given a graph G and a proper 2-colouring of its vertex set $c : V(G) \to \{0,1\}$, we define the imbalance of c as*

$$\sigma(c) := |c_0| - |c_1|.$$

For a tree T, we will use $\sigma(T)$ to denote the imbalance of its unique 2-colouring.

Proposition 3.2.5. *Let T be a tree with t edges. Then there exist $z \in V(T)$ and a proper 2-colouring $c : V(T-z) \to \{0,1\}$ of $T-z$, with $|c_1| \leqslant |c_0|$, such that $|c_0| \leqslant \frac{2t}{3}$ and $|c_1| \leqslant \frac{t}{2}$.*

Proof. We may assume that $t > 3$. Assume the proposition does not hold, that is, for every $z \in V(T)$ and every proper 2-colouring of $T-z$, the heavier colour class of $T-z$ contains more than $\frac{2t}{3}$ vertices.

Let $z_0 \in V(T)$ and $c : V(T-z_0) \to \{0,1\}$ as given by Lemma 3.2.3. By our assumption above, we know that c_0, the heavier colour class induced by c, contains between $\frac{2t}{3}$ and $\frac{3t-1}{4}$ vertices, while c_1, the lighter colour class, contains between $\frac{t}{4}$ and $\frac{t}{3}$ vertices.

Consider the set $\{T_i\}_{i\in I}$ of all components of $T-z_0$. Let $J \subseteq I$ be the set of all indices j such that T_j has more vertices in c_0 than in c_1. So clearly,

$$\sum_{i\in I\setminus J} \sigma(T_i) \leqslant 0 \tag{3.1}$$

and

$$\sum_{j\in J} \sigma(T_j) \geqslant |c_0| - |c_1| > \frac{t}{3}. \tag{3.2}$$

Moreover, we claim that

$$\text{for each } J' \subseteq J \text{ either } \sum_{j\in J'} \sigma(T_j) < \frac{t}{12} \text{ or } \sum_{j\in J'} \sigma(T_j) > \frac{t}{3}. \tag{3.3}$$

Indeed, if this were not true for some $J' \subseteq J$, we could invert the colours in all trees in $\{T_j\}_{j\in J'}$. This yields a colouring with both colour classes having at most $\frac{2t}{3}$ vertices, because c_0 would have lost at least $\frac{t}{12}$ vertices, and c_1 would have gained at most $\frac{t}{3}$ vertices. This contradicts our assumption, and thus proves (3.3).

We say that a family $J' \subseteq J$ is *small* if $\sum_{j\in J'} \sigma(T_j) < \frac{t}{12}$, and *large* otherwise (that is, by (3.3), if $\sum_{j\in J'} \sigma(T_j) > \frac{t}{3}$). Note that (3.2) implies that J is large.

Because of (3.3), for any partition $J = C \cup D$ we have that either C or D is large. So, taking a minimal large subset of J we see that

$$\text{there is a } j^* \in J \text{ such that } \{j^*\} \text{ is large.} \tag{3.4}$$

Notice that if $J \setminus \{j^*\}$ was large we could switch the colour classes in each of the associated trees and obtain a contradiction to the initial assumption. So,

$$J \setminus \{j^*\} \text{ is small,} \tag{3.5}$$

and therefore

$$\sum_{i\in I\setminus\{j^*\}} \sigma(T_i) \leqslant \sum_{i\in J\setminus\{j^*\}} \sigma(T_i) + \sum_{i\in I\setminus J} \sigma(T_i) < \frac{t}{12}. \tag{3.6}$$

Now, we apply Lemma 2.2.5 to obtain $z_1 \in V(T_{j^*})$ such that every component of $T_{j^*} - z_1$ covers at most $\lceil \frac{|T_{j^*}|-1}{2} \rceil$ vertices. Let T_{z_0} denote the component of $T - z_1$ that contains z_0 and let $\{C_\ell\}_{\ell\in L}$ denote the set of all other components of $T - z_1$. Further, let C_{z_0} denote the unique component of $T_{j^*} - z_1$ that is contained in T_{z_0}, if such a component exists. Observe that Lemma 2.2.5 allows us to assume that

$$|C_{z_0}| \leqslant \left\lfloor \frac{|T_{j^*}| - 1}{2} \right\rfloor \leqslant \left\lfloor \frac{\lceil \frac{t}{2} \rceil - 1}{2} \right\rfloor \leqslant \frac{t-1}{4}. \tag{3.7}$$

Next, we group the elements of $\{C_\ell\}_{\ell\in L}$ into two forests, F^A and F^B, satisfying

$$\max\{|F^A|, |F^B|\} \leqslant \frac{t+1}{3}, \tag{3.8}$$

which is possible by Lemma 3.2.1 (ii), and since $\max\{\frac{|T_{j^*}|-1}{2}, \frac{2}{3}|T_{j^*}|\} \leqslant \frac{t+1}{3}$.

For $i \in \{A, B\}$, consider the proper 2-colouring c^i induced by T_{j^*} on F^i. By symmetry, we may assume that

$$\sigma(c^A) \geqslant \sigma(c^B). \tag{3.9}$$

Observe that by (3.2), and by the choice of c^i, we have that

$$\frac{t}{3} < \sigma(T_{j^*}) \leqslant \sigma_{C_{z_0}\cup\{z_1\}} + \sigma(c^A) + \sigma(c^B),$$

where $\sigma_{C_{z_0}\cup\{z_1\}}$ denotes the imbalance that T_{j^*} induces on $C_{z_0} \cup \{z_1\}$. Note that $\sigma_{C_{z_0}\cup\{z_1\}} \leqslant \max\{|C_{z_0}|, 1\}$, Therefore,

$$\sigma(c^A) + \sigma(c^B) > \frac{t}{3} - \max\{|C_{z_0}|, 1\}. \tag{3.10}$$

Now we consider and separately treat two possible cases, according to the imbalance of the canonical colouring of T_{z_0}. For convenience, let $A(T_{z_0})$ denote the larger colour class of T_{z_0}

in this colouring, and let $B(T_{z_0})$ denote the smaller colour class.

Case 1: $\sigma(T_{z_0}) \leqslant \frac{t}{3}$.

In this case, we define a new colouring c' by setting $c'_0 := A(T_{z_0}) \cup c_1^A \cup c_0^B$ and $c'_1 := (T - z_1) \setminus c'_0 = B(T_{z_0}) \cup c_0^A \cup c_1^B$. Then, by (3.9),

$$|c'_0| = \frac{|T_{z_0}|}{2} + \frac{\sigma(T_{z_0})}{2} + \frac{|F^A|}{2} - \frac{\sigma(c^A)}{2} + \frac{|F^B|}{2} + \frac{\sigma(c^B)}{2} \leqslant \frac{t}{2} + \frac{\sigma(T_{z_0})}{2} \leqslant \frac{2t}{3},$$

and, moreover, by (3.8) we have

$$|c'_1| < \frac{|T_{z_0}|}{2} + \max\{|F^A| - 1, 1\} + \frac{|F^B|}{2} \leqslant \frac{2t}{3},$$

and hence after possibly swapping the colours we found a colouring as desired for the proposition (with z_1 in the role of z). This is a contradiction, since we assumed no such colouring exists.

Case 2: $\sigma(T_{z_0}) > \frac{t}{3}$.

This time we define c' by setting $c'_0 := A(T_{z_0}) \cup c_1^A \cup c_1^B$ and $c'_1 := B(T_{z_0}) \cup c_0^A \cup c_0^B$. Let $\sigma_{C_{z_0} \cup \{z_0\}}$ denote the imbalance that T_{z_0} induces on $C_{z_0} \cup \{z_0\}$ and note that by (3.7), we have

$$\sigma_{C_{z_0} \cup \{z_0\}} \leqslant \max\{|C_{z_0}|, 1\} \leqslant \frac{t-1}{4}.$$

Recalling (3.6) and (3.10), we obtain

$$\begin{aligned}
|c'_0| &= \frac{t}{2} + \frac{\sigma(T_{z_0}) - (\sigma(c^A) + \sigma(c^B))}{2} \\
&\leqslant \frac{t}{2} + \frac{\sum_{i \in I \setminus \{j^*\}} \sigma(T_i) + \sigma_{C_{z_0} \cup \{z_0\}} + \max\{|C_{z_0}|, 1\} - \frac{t}{3}}{2} \\
&\leqslant \frac{t}{2} + \frac{t}{24} + \max\{|C_{z_0}|, 1\} - \frac{t}{6} \\
&\leqslant \frac{2t}{3}.
\end{aligned}$$

Furthermore, by (3.8) we have

$$|c'_1| \leqslant \frac{|T_{z_0}|}{2} - \frac{\sigma(T_{z_0})}{2} + |c_0^A| + |c_0^B| \leqslant \frac{2t}{3},$$

and we thus again obtain a contradiction. $\square$

3.3 Embedding trees in robust components

In this section, we discuss the embedding of trees into a large *robust component* of some host graph, by which we mean we embed into graphs whose corresponding reduced graph has a large connected component. The arguments depend on whether the reduced graph is bipartite or not, and hence we deal with these situations separately.

The main results from this section are Propositions 3.3.1 and 3.3.9 and their corollaries. They will be used in the proof of Proposition 3.4.3, our main embedding result for robust components. Moreover, they will be one of the tools in the proof of our key embedding lemma (Lemma 3.5.3) on which most of our main results rely.

3.3.1 The bipartite case

As we mentioned in the introduction, any tree with k edges greedily embeds in any graph of minimum degree at least k. In a bipartite graph $H = (X, Y)$ the minimum degree condition can be relaxed: If the tree T has bipartition classes of sizes k_1 and k_2, then it is clearly enough to require the vertices from X to have degree at least k_1 and the vertices from Y to have degree at least k_2. In particular, if $\deg(x) \geqslant \lfloor \frac{k}{2} \rfloor$ for all $x \in X$, and $\deg(y) \geqslant k$ for all $y \in Y$, then each tree with k edges embeds in H.

If the tree we wish to embed has bounded degree, and the host graph has an (ε, η)-reduced graph which is bipartite and connected, for some ε and η, one can do even better: We will now show that in this case it is enough to require a minimum degree of roughly $\frac{k}{2}$ for the vertices in only one of the bipartition classes, as long as this class is not too small.

Proposition 3.3.1. *For all $\varepsilon \in (0, 10^{-8})$ and for all $d, M_0 \in \mathbb{N}$, there is k_0 such that for all $n, k \geqslant k_0$ the following holds. Let G be an n-vertex graph, with $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$ that satisfies $|\mathcal{R}| \leqslant M_0$, such that*

(i) $R = (\mathcal{A}, \mathcal{B})$ *is bipartite and connected;*

(ii) $\mathrm{diam}(\mathcal{R}) \leqslant d$*;*

(iii) $\deg(C) \geqslant (1 + 100\sqrt{\varepsilon})\frac{k}{2} \cdot \frac{|\mathcal{R}|}{n}$*, for all* $C \in \mathcal{A}$*; and*

(iv) $|\mathcal{A}| \geqslant (1 + 100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$.

Then G contains every tree T with k edges and $\Delta(T) \leqslant k^{\frac{1}{d}}$ as a subgraph.

Proof. Given ε, d and M_0 as in the Theorem, we set

$$k_0 := \left(\frac{8M_0^2}{\varepsilon^2} \right)^d .$$

Let G be a graph as in Proposition 3.3.1, and let $\bigcup \mathcal{A} = X_1 \cup \cdots \cup X_s$ and $\bigcup \mathcal{B} = Y_1 \cup \cdots \cup Y_t$ be the $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition of G corresponding to the reduced graph $\mathcal{R}$ (in particular $s + t \leqslant M_0$). Set $m := |X_i| = |Y_j|$ (for any i, j).

For each $i \in [s]$, we arbitrarily partition X_i into three sets $X_{i,S}$, $X_{i,L}$, $X_{i,C}$; and for each $j \in [t]$ we arbitrarily partition Y_j into three sets $Y_{j,S}$, $Y_{j,L}$, $Y_{j,C}$, such that

$$|X_{i,S}| = |X_{i,L}| = |Y_{j,S}| = |Y_{j,L}| = \lceil 10\sqrt{\varepsilon} m \rceil.$$

The letters S, L and C stand for seeds, links and clusters, respectively (sets $X_{i,C}$ and $Y_{j,C}$ contain the bulk of the clusters). We also call these subsets the L-, S- or C-slice of the corresponding cluster.

Note that, by Fact 2.3.1, for every (X_i, Y_j) with positive density, each of the pairs $(X_{i,K}, Y_{j,K'})$, with $K, K' \in \{S, L, C\}$, is $\frac{\sqrt{\varepsilon}}{5}$-regular with density greater than $4\sqrt{\varepsilon}$.

Root T at any vertex $r \in V(T)$. By Proposition 2.2.7, with parameters $\beta = \frac{\varepsilon}{s+t}$, we obtain a decomposition of (T, r) into a collection of pieces $\mathcal{P}$, each of order at most βk, and a family of seeds S of size at most $\frac{2}{\beta}$. Order the elements from $S \cup \mathcal{P}$ in a way that the first element is r, and the parent of each element is either an earlier seed or belongs to an earlier piece. (Note that the parent of a seed or piece is a vertex, so it either is a seed or belongs to a piece.)

Our plan is to embed the elements from $S \cup \mathcal{P}$ in this order. Seeds will go to S-slices of appropriate clusters $X_{i,S}$ or $Y_{j,S}$, with r going to cluster X_i if $r(T)$ belongs to the heavier bipartition class of T, and going to Y_j otherwise. Pieces from $\mathcal{P}$ will go into C-slices $(X_{i,C}, Y_{j,C})$ of appropriate pairs (X_i, Y_j), and into L-slices of other clusters.

More precisely, for each piece $P \in \mathcal{P}$ we will find a pair (X_i, Y_j) such that there is enough space left in $(X_{i,C}, Y_{j,C})$ to accommodate P. At this point, the parent of P is already embedded into some cluster Z, so we need to embed part of P into a path $ZZ_1Z_2Z_3 \dots Z_h$ that connects Z with the pair (X_i, Y_j). Because of the bounded degree of T, and since the diameter of G is also bounded, this path can be chosen short enough to ensure that the levels of P that are embedded into this path only contain relatively few vertices. So we can use the L-slices of the clusters Z_ℓ for these levels. The remaining levels of P will be embedded into the free space of $(X_{i,C}, Y_{j,C})$.

Let us make this sketch more precise. During the embedding procedure, we will write $X'_{i,C}$ and $Y'_{j,C}$ for the set of unoccupied vertices of $X_{i,C}$ and $Y_{j,C}$ respectively. We will say that a pair (X_i, Y_j) is *good* if $d(X_i, Y_j) > 0$ and $\min\{|X'_{i,C}|, |Y'_{j,C}|\} \geqslant 5\sqrt{\varepsilon} m$. Hence we will be able to apply Lemma 2.3.4 to any good pair and any piece belonging to $\mathcal{P}$.

The embedding $\phi : V(T) \to V(G)$ will be constructed iteratively, following the embedding order of $S \cup \mathcal{P}$ chosen above. Employing the strategy explained above, we make sure that at every step, the following conditions will be satisfied:

1. Each vertex is embedded into a neighbour of the image of its already embedded parent;
2. each $s \in S$ is embedded into the S-slice of some cluster;
3. for each $P \in \mathcal{P}$, the first (up to $d - 1$) levels are embedded into the L-slices of some clusters, and the rest goes into the C-slices; and

4. every $v \in V(T)$ is mapped into a vertex that is typical towards both the S-slice and the L-slice of some adjacent cluster.

Since the set S has constant size, and since we do not particularly care into which cluster a seed goes, as long as it goes to the S-slice, it is clearly possible to embed a seed s, when its time comes, satisfying conditions (1), (2) and (4).

So assume we are about to embed a piece $P \in \mathcal{P}$. The parent of the root $r(P)$ of P is already embedded into some vertex that is typical with respect to the L-slice of some cluster Z_1. In order to be able to embed P according to our plan, it suffices to ensure that

1. there exists some good pair (X_i, Y_j);
2. there is a path $Z_1 Z_2 Z_3 \ldots Z_h$ of length $h \leqslant d$ from Z_1 to X_i;
3. the union of the first $h-1$ levels of P is small enough to fit into the free space in the L-slices of $\{Z_1, Z_2, Z_3, \ldots, Z_{h-1}\}$.

If we can assure these properties, we can repeatedly apply Lemma 2.3.4 to embed the first levels of P into the L-slices of the clusters Z_ℓ, and the remaining levels of P into $(X'_{i,C}, Y'_{j,C})$ in a way that (1), (3) and (4) hold.

So, let us prove (1). We first note that there exists some cluster X_i such that $|\phi^{-1}(X_{i,C})| < |X_{i,C}| - 5\sqrt{\varepsilon}m$. Indeed, otherwise we have used at least

$$(1 - 25\sqrt{\varepsilon})|\bigcup \mathcal{A}| - 5\sqrt{\varepsilon}|\bigcup \mathcal{A}| \geqslant (1 - 30\sqrt{\varepsilon})(1 + 100\sqrt{\varepsilon})k > (1 + 2\sqrt{\varepsilon})k > k + 1$$

vertices from $\bigcup \mathcal{A}$ already, a contradiction, since $|T| = k + 1$.

Next, we claim there exists some cluster Y_j such that (X_i, Y_j) is good. If this was not the case, then we have used at least

$$(1 - 30\sqrt{\varepsilon})|N_G(X_i)| \geqslant (1 - 30\sqrt{\varepsilon})(1 + 100\sqrt{\varepsilon})\frac{k}{2} > (1 + 2\sqrt{\varepsilon})\frac{k}{2} > \frac{k+1}{2}$$

vertices of $\bigcup \mathcal{B}$ already, a contradiction, as we placed the root r of T in a way that guaranteed we would embed the smaller bipartition class of T into $\mathcal{B}$.

Observe that (1) implies (2), because of condition (ii) of Proposition 3.3.1. So it only remains to prove (3).

Using (3) for already embedded pieces P', and using the fact that, for any such piece P', the number of vertices in their first $d-1$ levels is bounded by $2(\Delta(T) - 1)^{d-2}$ (except if $\Delta(T) \leqslant 2$, in which case this number is bounded by $d - 1$), we have that the total number of occupied vertices in L-slices is at most

$$|S| \cdot \Delta(T) \cdot 2(\Delta(T) - 1)^{d-2} \leqslant \frac{4}{\beta} \cdot k^{\frac{d-1}{d}} \leqslant \frac{4M_0}{\varepsilon} \cdot k^{\frac{d-1}{d}} < \varepsilon \frac{k}{2M_0} \leqslant \varepsilon m$$

for $k \geqslant k_0$. In particular, each L-slice of a cluster Z_ℓ has at least $\lceil 9\sqrt{\varepsilon}m \rceil$ unused vertices. This is enough to ensure that the first $h-1$ levels of P fit into the L-slices of the clusters Z_1, Z_2, $Z_3, \ldots, Z_{h-1}$. This proves (3). $\square$

Remark 3.3.2. *It is easy to see that instead of conditions* (iii) *and* (iv) *from Proposition 3.3.1 we could use the weaker requirement that there is a set* $\mathcal{C}$ *of clusters in* $\mathcal{A}$ *such that* $\deg(C) \geqslant (1+100\sqrt{\varepsilon})\frac{k}{2} \cdot \frac{|\mathcal{R}|}{n}$*, for all* $C \in \mathcal{C}$*, and* $|\mathcal{C}| \geqslant (1+100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$.

Remark 3.3.3. *Observe that Proposition 3.3.1 remains true with the following additional conditions. Let* $U \subseteq V(G)$ *such that*

- $|U| + |T| \leqslant k+1$;
- $|U \cap V(\bigcup \mathcal{A})| + c_0(T) \leqslant k$*; and*
- $|U \cap V(\bigcup \mathcal{B})| + c_1(T) \leqslant \frac{k}{2}$,

where $c_0(T)$ *and* $c_1(T)$ *are the two colour classes of* T*. Then* T *can be embedded into* G *avoiding* U*, that is, we can embed* T *in such a way* $\phi(V(T)) \subseteq V(G) \setminus U$*.*

Moreover, observe that by repeatedly applying Proposition 3.3.1, together with Remark 3.3.3, we can actually embed a forest instead of a tree. We say that a forest F, with colour classes C_1 and C_2, is a (k_1, k_2, t)-*forest* if

1. $|C_i| \leqslant k_i$ for $i \in \{1,2\}$, and
2. $\Delta(F) \leqslant (k_1+k_2)^t$.

Corollary 3.3.4. *For all* $\varepsilon \in (0, 10^{-8})$ *and for all* $d, M_0 \in \mathbb{N}$ *there is* k_0 *such that for all* $n, k_1, k_2 \geqslant k_0$ *the following holds. Let* G *be an* n*-vertex graph having a* $(\varepsilon, 5\sqrt{\varepsilon})$*-reduced graph* $\mathcal{R}$ *that satisfies* $|\mathcal{R}| \leqslant M_0$*, such that*

(i) $\mathcal{R} = (\mathcal{A}, \mathcal{B})$ *is connected and bipartite;*

(ii) $\operatorname{diam}(\mathcal{R}) \leqslant d$*;*

(iii) $\deg(C) \geqslant (1+100\sqrt{\varepsilon})k_2 \cdot \frac{|\mathcal{R}|}{n}$*, for all* $C \in \mathcal{A}$*; and*

(iv) $|\mathcal{A}| \geqslant (1+100\sqrt{\varepsilon})k_1 \cdot \frac{|\mathcal{R}|}{n}$.

Then any $(k_1, k_2, \frac{1}{d})$*-forest* F*, with colour classes* C_1 *and* C_2*, can be embedded into* G*, with* C_1 *going to* $V(\bigcup \mathcal{A})$ *and* C_2 *going to* $V(\bigcup \mathcal{B})$*.*

Moreover, if F *has at most* $\frac{\varepsilon n}{|\mathcal{R}|}$ *roots, then the images of the roots going to* $V(\bigcup \mathcal{A})$ *can be mapped to any prescribed set of size at least* $2\varepsilon|\bigcup \mathcal{A}|$ *in* $\bigcup \mathcal{A}$*, and the images of the roots going to* $V(\bigcup \mathcal{B})$ *can be mapped to any prescribed set of size at least* $2\varepsilon|\bigcup \mathcal{B}|$ *in* $\bigcup \mathcal{B}$*.*

Remark 3.3.5. *An analogue of Remark 3.3.3 holds for the situation of Corollary 3.3.4.*

It is easy to see that we can bound the balancedness of trees whose maximum degree is bounded by some constant Δ. So, it comes as no surprise that for the class of all constant degree trees, it is possible to show the following improvement of Proposition 3.3.1.

Corollary 3.3.6. *For all $\varepsilon \in (0, 10^{-8})$, $d, M_0 \in \mathbb{N}$ and $\Delta \geqslant 2$ there is a k_0 such that for all $n, k \geqslant k_0$ the following holds. Let G be an n-vertex graph that has an $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$ that satisfies $|\mathcal{R}| \leqslant M_0$, such that*

1. *$\mathcal{R} = (\mathcal{A}, \mathcal{B})$ is connected and bipartite;*
2. *$\operatorname{diam}(\mathcal{R}) \leqslant d$;*
3. *$\deg(C) \geqslant (1 + 100\sqrt{\varepsilon})\frac{k}{2} \cdot \frac{|\mathcal{R}|}{n}$ for all $C \in \mathcal{A}$;*
4. *$|\mathcal{A}| \geqslant (1 + 100\sqrt{\varepsilon})\frac{(\Delta-1)}{\Delta}k \cdot \frac{|\mathcal{R}|}{n}$.*

Then G contains every tree T with k edges and $\Delta(T) \leqslant \Delta$ as a subgraph.

3.3.2 The nonbipartite case

In this section we treat tree embeddings into graphs with large nonbipartite components in the reduced graph. The proof of the corresponding proposition, Proposition 3.3.9 below, is very similar to the proof of Proposition 3.3.1. For convenience, we will now work with a matching in the reduced graph.

For a graph G with an (ε, η)-regular partition, we say that $\mathcal{M}$ is a cluster matching if it is a matching in the corresponding (ε, η)-reduced graph. We begin our treatment of the nonbipartite case by showing that we can always find a large cluster matching in graphs with large minimum degree that admit an (ε, η)-regular partition, for some $\varepsilon, \eta \in (0, 1)$. To do so, we first need the following result.

Lemma 3.3.7. *Let H be any graph. Then there exists an independent set I, a matching M, and a set of vertex disjoint triangles Γ so that $V(H) = I \cup V(M) \cup V(\Gamma)$. Moreover, there is a partition $V(M) = V_1 \cup V_2$ of $V(M)$ such that every edge of M has one vertex in V_1 and one vertex in V_2, and $N(x) \subseteq V_1$ for all $x \in I$.*

Proof. Let us choose a matching M and a family Γ of disjoint triangles, that are disjoint from M, maximising $|V(M)| + |V(\Gamma)|$. Then the set I consisting of all vertices not covered by $M \cup \Gamma$ is independent.

Consider a vertex $x \in I$. Note that because of our choice of M and Γ, we know that x is not adjacent to any vertex from any triangle from Γ. Also, note that for any edge uv in M, vertex x sees at most one of u, v. Finally, if x sees u, then no other vertex from I can see v. This proves the statement. □

Lemma 3.3.8. *Let $\varepsilon, \eta \in (0, 1)$, let $t, \ell \in \mathbb{N}$, and let G be a graph on $n \geqslant 2t + \ell$ vertices with $\delta(G) \geqslant t + \ell$ which has an (ε, η)-regular partition into ℓ parts. Then G has a subgraph G' with $|G'| \geqslant n - \ell$ that admits a $(5\varepsilon, \eta - \varepsilon)$-regular partition with 2ℓ parts whose corresponding reduced graph $\mathcal{R}$ contains a matching $\mathcal{M}$ and an independent family of clusters $\mathcal{I}$, disjoint from $\mathcal{M}$, such that* *(i)* $\bigcup V(\mathcal{M}) \cup V(\bigcup \mathcal{I}) = V(G')$*;*

(ii) $|\bigcup V(\mathcal{M})| \geqslant 2t$*; and*

(iii) there is a partition $V(\mathcal{M}) = \mathcal{V}_1 \cup \mathcal{V}_2$ *such that* $N_{\mathcal{R}}(\mathcal{I}) \subseteq \mathcal{V}_1$ *and every edge in* $\mathcal{M}$ *has one endpoint in* $\mathcal{V}_1$ *and one endpoint in* $\mathcal{V}_2$*.*

Proof. Let $\mathcal{R}$ be the reduced graph corresponding to the (ε, η)-regular partition of G. By applying Lemma 3.3.7 to $\mathcal{R}$, we obtain an independent set $\mathcal{I}'$, a matching $\mathcal{M}'$ and a set of disjoint triangles Γ, such that $V(\mathcal{R}) = \mathcal{I} \cup V(\mathcal{M}') \cup V(\Gamma)$. If Γ is empty, we are done by choosing $\mathcal{I} := \mathcal{I}'$ and $\mathcal{M} := \mathcal{M}'$. So suppose $\Gamma \neq \emptyset$.

We arbitrarily partition each cluster $X \in V(\mathcal{R})$ into $X^1 \cup X^2 \cup X^3$ so that $|X^1| = |X^2|$ and $|X^3| \leqslant 1$. Let $G' = G - V(\bigcup_{X \in V(\mathcal{R})} X^3)$. Thanks to Fact 2.3.1.2, the partition $V(G') = \bigcup_{X \in V(R)} X^1 \cup X^2$ is $(5\varepsilon, \eta - \varepsilon)$-regular and has 2ℓ atoms. We set

$$\mathcal{M} := \bigcup_{CD \in \mathcal{M}'} \{(C^1 D^1), (C^2 D^2)\} \cup \bigcup_{XYZ \in \Gamma} \{(X^1, Y^2), (Y^1, Z^2), (Z^1, X^2)\}$$

and

$$\mathcal{I} := \bigcup_{C \in \mathcal{I}'} \{C^1, C^2\}$$

Note that $\mathcal{I}$ and $\mathcal{M}$ inherit the properties of $\mathcal{I}'$ and $\mathcal{M}'$, respectively. Property (ii) follows from Property (iii) and the minimum degree of G. □

We will apply Lemma 3.3.8 to the reduced graph of a given host graph G. The lemma then says that, after modifying the reduced graph (cutting its clusters in half), one can find a cluster matching whose size depends on the minimum degree of G. In particular, given $\delta > 0$, if G has minimum degree at least $(1+\delta)\frac{k}{2}$, one can find a matching covering at least $(1+\delta)k$ vertices of G.

Now we are ready for Proposition 3.3.9 and its proof.

Proposition 3.3.9. *For all* $\varepsilon \in (0, 10^{-8})$ *and* $d, M_0 \in \mathbb{N}$ *there exists* k_0 *such that for all* $n, k \geqslant k_0$ *the following holds. Let* G *be an* n*-vertex graph that has an* $(\varepsilon, 5\sqrt{\varepsilon})$*-reduced graph* $\mathcal{R}$ *that satisfies* $|\mathcal{R}| \leqslant M_0$*, such that*

1. *$\mathcal{R}$ is connected and nonbipartite;*
2. $\mathrm{diam}(\mathcal{R}) \leqslant d$*; and*
3. *$\mathcal{R}$ has a matching $\mathcal{M}$ with* $|V(\mathcal{M})| \geqslant (1 + 100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$*.*

Then G *contains every tree* T *with* k *edges and* $\Delta(T) \leqslant k^{\frac{1}{3d+1}}$ *as a subgraph.*

Proof. Given ε, d and M_0 as in the Theorem, we set

$$k_0 := \left(\frac{8M_0^2}{\varepsilon^2}\right)^{3d+1}$$

Now, let G be a graph as in Proposition 3.3.9, let $V(G) = V_1 \cup \ldots \cup V_\ell$ be the $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition of G corresponding to the reduced graph $\mathcal{R}$ (in particular $\ell \leqslant M_0$). Set $m := |V_i|$ for any $i \in [\ell]$.

For each $i \in [\ell]$, we partition cluster V_i into sets $V_{i,S}$, $V_{i,L}$, $V_{i,C}$ in the same way as we did in Proposition 3.3.1. Also, consider the decomposition of T into $\mathcal{T}$ and S given by Proposition 2.2.7, with $\beta = \frac{\varepsilon}{\ell}$. We order $S \cup \mathcal{P}$ in the same way as in the proof of Proposition 3.3.1.

The embedding $\phi : V(T) \to V(G)$ will be constructed iteratively, following the order of $S \cup \mathcal{P}$. We make sure that at every step, the following conditions will be satisfied:

1. Each vertex is embedded into a neighbour of the image of its already embedded parent;
2. each $s \in S$ is embedded into the S-slice of some cluster;
3. for each $P \in \mathcal{P}$, the first $3d$ levels are embedded into the L-slices of some clusters, and the remaining levels go to the C-slices;
4. every $v \in V(T)$ is mapped to a vertex that is typical with respect to both the S-slice and the L-slice of some adjacent cluster; and
5. $\Big||\phi^{-1}(V_{i,C})| - |\phi^{-1}(V_{j,C})|\Big| \leqslant \epsilon m$ for each pair $(V_i, V_j) \in \mathcal{M}$.

We already know that it is no problem to embed a seed s, when its time comes, satisfying conditions (1), (2) and (4). So we mainly have to worry about (3) and (5).

Assume we are about to embed a piece $P \in \mathcal{P}$. The parent of the root $r(P)$ of P is already embedded into some vertex that is typical with respect to the L-slice of some cluster Z_1. In order to be able to embed P so that the above conditions are satisfied, it suffices to ensure that

1. there exists some good pair (V_i, V_j);
2. for either choice of $Z_{3d+1} \in \{V_i, V_j\}$ there is a walk $Z_1 Z_2 \ldots Z_{3d+1}$ in R;
3. the first $3d$ levels of P are small enough to fit into the free space in the L-slices of $\{Z_1, Z_2 \ldots, Z_{3d}\}$,

where a walk in a graph is a sequence $Z_1 Z_2 \ldots Z_h$ such that each Z_i is adjacent to Z_{i+1} for all $1 \leqslant i < h$.

Before we prove (1)–(3), let us explain why these conditions are enough to ensure we can embed T correctly. As before, we plan to repeatedly apply Lemma 2.3.4 in order to embed the first levels of P into the L-slices of the clusters $Z_1, Z_2, \ldots, Z_{3d}$, and the later levels into the C-slices of V_i, V_j, always avoiding all vertices used earlier.

Since our aim is to embed P in such a way that (5) is fulfilled, we take care to choose $Z_{3d+1} \in \{V_i, V_j\}$ in a way that the larger bipartition class of the tree P' obtained from P by deleting its first $3d$ levels goes to the less occupied slice from $V_{i,C}$, $V_{j,C}$. That is, assuming that $\left|\phi^{-1}(V_{i,C})\right| \leqslant \left|\phi^{-1}(V_{j,C})\right|$ (the other case is analogous), we proceed as follows. If the levels of P' that lie at even distance from the root of P in total contain more vertices than those lying at odd distance, we choose $Z_{3d+1} = V_j$. Otherwise, we choose $Z_{3d+1} = V_i$. We then embed P, making the first $3d$ levels go to L-slices, and embedding P' into $V_{i,C} \cup V_{j,C}$.

Let us now prove (1). Suppose there is no good pair in $\mathcal{M}$. This together with (5) implies that the number of embedded vertices is at least

$$\sum_{AB \in \mathcal{M}} (|A_{i,C}| - 6\sqrt{\varepsilon} m + |B_{i,C}| - 6\sqrt{\varepsilon} m) \geqslant (1 - 33\sqrt{\varepsilon})(1 + 100\sqrt{\varepsilon})k > k + 1,$$

a contradiction, since $|T| = k + 1$.

Next, we show (2). Assume we chose $Z_{3d+1} = V_i$ (the other case is analogous). Let $C = C_1 C_2 \dots C_p C_1$ be a minimal odd cycle in the reduced graph. Since C is minimally odd, the shortest path between two clusters in C is the shortest arc in the cycle, and hence $p \leqslant 2d+1$. Let $U := Z_1 U_1 \dots U_s C_1$ be a shortest path from Z_1 to C_1 and let $Q := C_{\lceil \frac{p}{2} \rceil} Q_1 \dots Q_t V_i$ be a shortest path from $C_{\lceil \frac{p}{2} \rceil}$ to V_i. As $\mathrm{diam}(\mathcal{R}) \leqslant d$, we have that $s + t + 2 \leqslant 2d$. So, by using the appropriate one of the two C_1–$C_{\lceil \frac{p}{2} \rceil}$ paths in C, we can extend $U \cup Q$ to an odd walk of length at most $2d + (d+1) = 3d + 1$, which connects Z_1 with V_i. By going back- and forwards on this walk, if necessary, we can obtain a walk of length exactly $3d+1$, which is as desired. So, condition (2) holds.

Finally, using the same reasoning as in Proposition 3.3.1 we can prove that the total number of occupied vertices in L-slices is at most

$$|S| \cdot \Delta(T) \cdot 2(\Delta(T) - 1)^{3d-1} \;\leqslant\; \frac{4}{\beta} \cdot k^{\frac{3d}{3d+1}} \;<\; \varepsilon m$$

for $k \geqslant k_0$. In particular, the L-slice of each cluster has at least $\lceil 9\sqrt{\varepsilon} m \rceil$ unused vertices and, therefore, we can embed each vertex of the first $3d$ levels of P into the L-slices of the clusters from the walk $Z_1 Z_2 \dots Z_{3d}$ without a problem. This proves (3). □

Remark 3.3.10. *If $d = 1$ we can actually embed trees with maximum degree bounded by ρk, where ρ is a sufficiently small constant, without modifying our proof significantly, because we can reach both V_i and V_j in one step from the image of the latest embedded seed.*

Remark 3.3.11. *Similarly to the bipartite case, we can add an extra hypothesis as in Remark 3.3.3. Consider an arbitrary set $U \subseteq V(G)$ such that $|U| + |T| \leqslant k + 1$ and such that U is reasonably balanced in $\mathcal{M}$, that is, $\big||U \cap C| - |U \cap D|\big| < \varepsilon |C|$ for all $CD \in \mathcal{M}$. Then T can be embedded into G avoiding U.*

Repeatedly applying Proposition 3.3.9, together with Remark 3.3.11, we can embed a forest instead of a tree.

Corollary 3.3.12. *Let $\varepsilon \in (0, 10^{-8})$ and let $d, M_0 \in \mathbb{N}$. There exists $k_0 \in \mathbb{N}$ such that for all $n, k \geqslant k_0$ the following holds. Let G be a n-vertex graph with an $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$ that satisfies $|\mathcal{R}| \leqslant M_0$. Suppose that*

1. $\mathcal{R}$ is connected and nonbipartite;

2. $\operatorname{diam}(\mathcal{R}) \leqslant d$*;*

3. $\mathcal{R}$ has a matching $\mathcal{M}$ with $|V(\mathcal{M})| \geqslant (1+100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$;

then any forest F on at most $k+1$ vertices that satisfies $\Delta(F) \leqslant k^{\frac{1}{3d+1}}$ is a subgraph of G. Moreover, if F has at most $\frac{\varepsilon n}{|\mathcal{R}|}$ roots, then the images of the roots can be mapped into any prescribed set of size at least $2\epsilon n$.

3.4 Improving the maximum degree bound

In the previous section, we proved that in graphs of minimum degree at least $(1+\delta)\frac{k}{2}$ having a large connected component, after applying regularity and performing the usual cleaning-up, all trees of maximum degree $k^{O(\frac{1}{d})}$ appear as subgraphs (see Propositions 3.3.1 and 3.3.9). The aim of the present section is to prove a similar statement as there, but with a significant weakening in the bound on the maximum degree of the tree. More precisely, the exponent in this bound will no longer depend on the diameter of the reduced graph.

We need a theorem from [45], which says that one can bound the diameter of any connected graph in terms of its number of vertices and its minimum degree.

Theorem 3.4.1 (Erdős, Pach, Pollach and Tuza [45])**.** *Let G be a connected graph on n vertices with minimum degree at least 2. Then*

$$\operatorname{diam}(G) \leqslant \left\lfloor \tfrac{3n}{\delta(G)+1} \right\rfloor - 1.$$

We also need the following lemma. Given a graph G and a vertex $v \in V(G)$, let $N_i(v)$ denote the i-th neighbourhood of v (i.e. the set of vertices of G at distance i from v).

Lemma 3.4.2. *Let $q \in \mathbb{N}$ and let G be a connected graph, and let $v \in V(G)$. Then*

$$\left| \bigcup_{i=0}^{3q+1} N_i(v) \right| \geqslant \min\{(q+1)(\delta(G)+1), |V(G)|\}.$$

Proof. If $N_i(v) = \emptyset$ for some $i \in [3q+1]$, then, as G is connected, $V(G) \subseteq \bigcup_{j=0}^{i-1} N_j(v)$ and thus $|\bigcup_{j=0}^{3q+1} N_j(v)| = |V(G)|$. Therefore, we assume that $N_i(v) \neq \emptyset$ for every $i \in [3q+1]$.

Now, for each $j \in [q]$, pick a vertex $v_{3j} \in N_{3j}(v)$. Observe that $N(v_{3j}) \subseteq N_{3j-1}(v) \cup N_{3j}(v) \cup N_{3j+1}(v)$, and hence,

$$|N_{3j-1}(v) \cup N_{3j}(v) \cup N_{3j+1}(v)| \geqslant \delta(G) + 1.$$

We also know that $|N_0(v) \cup N_1(v)| = |N(v)| + 1 \geqslant \delta(G) + 1$. This proves the statement. $\square$

The next result shows that we can make the exponent in Proposition 3.3.1 and Proposition 3.3.9 depending only on the minimum degree of G. In order to prove the result, we will first apply a strategy similar to the one used in Propositions 3.3.1 and 3.3.9. If this strategy fails, we will have found a good structure in the host graph and then, forgetting about the earlier attempt at an embedding of T, we make use of the structure to embed the tree in a different way.

Proposition 3.4.3. *For all $\alpha \in [\frac{1}{2}, 1)$, $\varepsilon \in (0, 10^{-8})$ and $M_0 \in \mathbb{N}$, there exists $k_0 \in \mathbb{N}$ such that for all $n, k \geqslant k_0$ the following holds.*

Let G be a n-vertex graph with $\delta(G) \geqslant (1 + 100\sqrt{\varepsilon})\alpha k$ that has a connected $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$ with $|\mathcal{R}| \leqslant M_0$. If

1. *$\mathcal{R} = (\mathcal{A}, \mathcal{B})$ is bipartite and such that $|\mathcal{A}| \geqslant (1 + 100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$; or*
2. *$\mathcal{R}$ is non-bipartite and $n \geqslant (1 + 100\sqrt{\varepsilon})k$;*

then G contains every k-edge tree of maximum degree at most $k^{\frac{1}{r}}$, where $r = 18\lceil \frac{2}{\alpha} \rceil - 5$.

Proof. Given α, we define

$$d_1 := 3\lceil \tfrac{2}{\alpha} \rceil - 2 \text{ and } d_2 := 2(d_1 + 1),$$

and observe that

$$r = 3d_2 + 1.$$

Given ε and M_0, let k_0 be the maximum of the outputs of Proposition 3.3.1 and Proposition 3.3.9, for input ε, d_2 and $2M_0$.

Let G be as in Proposition 3.4.3. Note that if $|V(G)| < (1+100\sqrt{\varepsilon})2k$, then Theorem 3.4.1 implies that $\mathrm{diam}(\mathcal{R}) \leqslant \lfloor \frac{6}{\alpha} \rfloor - 1 \leqslant d_2$. Therefore, we may apply either Proposition 3.3.1 or Proposition 3.3.9, together with Lemma 3.3.8, to conclude. Thus, from now on we will assume that

$$|V(G)| \geqslant (1 + 100\sqrt{\varepsilon})2k. \tag{3.11}$$

Let T be a tree with k edges and $\Delta(T) \leqslant k^{\frac{1}{r}}$, and let us root T at any vertex. We partition T using Proposition 2.2.7, with $\beta := \frac{\varepsilon n}{|\mathcal{R}|}$, obtaining a set S of seeds and a family $\mathcal{P}$ of pieces. We first try to emulate the embedding scheme used in the proof of Proposition 3.3.1.

Consider the regular partition associated to the reduced graph $\mathcal{R}$ of G, and divide each cluster X into three sets X_C, X_S, X_L, with $|X_S| = |X_L| = \lceil 10\sqrt{\varepsilon}|X| \rceil$. We are going to embed T in $|S|$ steps, letting ϕ denote the partial embedding defined so far.

At step j we consider a vertex $s_j \in S$ not embedded yet, but whose parent u_j is already embedded (except in the step $j = 1$, in which case we embed the root of T into any cluster of our choice). We know that $\phi(u_j)$ is typical towards the S-slice of some adjacent cluster Q. Embed s_j in Q_S, choosing $\phi(s_j)$ typical to U_L and to U_S, where U is any neighbour of Q.

Now, suppose there is a good pair (W, Z), that is, an edge WZ such that both clusters W and Z have free space of size at least $5\sqrt{\varepsilon}|W|$, and additionally, $\mathrm{dist}(U, W) \leqslant d_1$. Find a shortest path from U to W, say $X_0 X_1 \dots X_{t-1} X_t$, where $X_0 = U$ and $X_t = W$ and, further, $t \leqslant d_1$.

Consider a piece P adjacent to s_j that is not yet embedded. We map the root of P into the neighbourhood of $\phi(s_j)$ in $(X_0)_L$. We then embed the first t levels of P into the path $X_0 X_1 \dots X_{t-1} X_t$, mapping the vertices from the i-th level of P into unoccupied vertices from $(X_i)_L$ that are typical towards $(X_{i+1})_L$ and to $(X_{i+1})_S$, for each $i \in \{0, \dots, t-1\}$ respectively. Finish the embedding of P, by mapping the remaining levels into the unoccupied vertices of (W_C, Z_C). For this, we use Lemma 2.3.4, mapping the vertices from the t-th level of P into W_C and picking all the images typical towards the L-slice and the S-slice of some adjacent cluster. We repeat this procedure for every not yet embedded piece adjacent to s_j and then move on to the next seed.

If every step of this process is successful, then T is satisfactorily embedded into G. However, it might happen that the embedding cannot be completed, because at some step we could not find a good pair (W, Z) at close distance. In that case, consider the seed s^* where the process stopped and let C^* be the cluster to which s^* was assigned. Let us define $\mathcal{H}$ as the subgraph of $\mathcal{R}$ induced by all those clusters that lie at distance at most d_1 from C^*. Further, let $\mathcal{S}$ be the set of all those clusters $C \in V(\mathcal{H})$ that have free space of size at least $5\sqrt{\varepsilon}|C|$. Note that, since the embedding could not be finished,

$$\mathcal{S} \text{ is an independent set.} \tag{3.12}$$

By applying Lemma 3.4.2, with $q = \lceil \frac{2}{\alpha} \rceil - 1$, and since $\delta(\mathcal{R}) \geqslant (1 + 100\sqrt{\varepsilon})\alpha k \cdot \frac{|\mathcal{R}|}{n}$ and by (3.11), we deduce that

$$|V(\mathcal{H})| > (1 + 100\sqrt{\varepsilon})2k \cdot \frac{|\mathcal{R}|}{n}.$$

This is more than twice the space needed for embedding T. So, since we have embedded at most k vertices before we declared the embedding to have failed, we conclude that

$$|\mathcal{S}| \geqslant (1 + 200\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}. \tag{3.13}$$

Let us define $\mathcal{H}'$ as the subgraph of $\mathcal{R}$ induced by all clusters at distance at most $d_1 + 1$ from C^*. So, $V(\mathcal{H}')$ consists of $V(\mathcal{H})$ together with the neighbours of $\mathcal{H}$ in $\mathcal{R}$.

Forgetting about our previous attempt to embed T, we are now going to embed T with the help of our earlier propositions. We distinguish two cases, depending on whether $\mathcal{H}'$ is bipartite or not.

Case 1: $\mathcal{H}'$ is nonbipartite.

Let $\mathcal{M}$ be a matching in $\mathcal{H}$ covering a maximal number of clusters from $\mathcal{S}$. We claim that $|V(\mathcal{M})| \geqslant (1 + 100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$. Indeed, otherwise (3.13) implies that there is a cluster $X \in V(\mathcal{S}) \setminus V(\mathcal{M})$. By our choice of $\mathcal{M}$, and because of (3.12), we know that X sees at most one end vertex of each edge from $\mathcal{M}$, and no cluster outside $V(\mathcal{M})$. This contradicts the fact that $\deg_{\mathcal{H}'}(X) \geqslant (1+100\sqrt{\varepsilon})\frac{k}{2} \cdot \frac{|\mathcal{R}|}{n}$.

Hence, as $\operatorname{diam}(\mathcal{H}') \leqslant 2(d_1+1) = d_2$, we can apply Proposition 3.3.9 to $\mathcal{H}'$ and the subgraph of G induced by the clusters of $\mathcal{H}'$, and we are done.

Case 2: $\mathcal{H}'$ is bipartite.

Since $|V(\mathcal{H})| \geqslant (1+100\sqrt{\varepsilon})2k \cdot \frac{|\mathcal{R}|}{n}$, one of the bipartition classes of $\mathcal{H}'$, say $\mathcal{A}$, satisfies $|\mathcal{A} \cap \mathcal{H}| \geqslant (1+100\sqrt{\varepsilon})k \cdot \frac{|\mathcal{R}|}{n}$. Since $\deg_{\mathcal{H}'}(X) \geqslant (1+100\sqrt{\varepsilon})\frac{k}{2} \cdot \frac{|\mathcal{R}|}{n}$ for each $X \in V(\mathcal{H})$, we can apply Proposition 3.3.1, together with Remark 3.3.2, to obtain the embedding of T. □

3.5 The key embedding lemma

In the current section, we present and prove our key embedding lemma, namely Lemma 3.5.3. This lemma describes a series of configurations which, if they appear in a graph G, allow us to embed any bounded degree tree of the right size into G.

Before stating the lemma we need two simple definitions.

Definition 3.5.1 (θ-see)**.** *Let $\theta \in (0,1)$. A vertex x of a graph H θ-sees a set $U \subseteq V(H)$ if it has at least $\theta|U|$ neighbours in U. Furthermore, if $\mathcal{C}$ is a component of some reduced graph of $H-x$, we say that x θ-sees $\mathcal{C}$ if x has at least $\theta|\bigcup\mathcal{C}|$ neighbours in $V(\bigcup\mathcal{C})$.*

Definition 3.5.2 ((k,θ)-small and (k,θ)-large)**.** *Let $k \in \mathbb{N}$ and let $\theta \in (0,1)$. A nonbipartite graph G is said to be (k,θ)-small if $|V(G)| < (1+\theta)k$. A bipartite graph $H=(A,B)$ is said to be (k,θ)-small if $\max\{|A|,|B|\} < (1+\theta)k$. If a graph is not (k,θ)-small, we will say that it is (k,θ)-large.*

We are now ready for the key lemma (for an illustration of the situation described in the lemma, see Figure 3.2).

Lemma 3.5.3 (Key embedding lemma)**.** *For each $\alpha \in [\frac{1}{2},1)$, for each $\varepsilon \in (0,10^{-10})$ and for each $M_0 \in \mathbb{N}$ there is $n_0 \in \mathbb{N}$ such that for all $n,k \geqslant n_0$ the following holds.*

Let G be an n-vertex graph of minimum degree at least $(1+\sqrt[4]{\varepsilon})\alpha k$ and let T be a tree with k edges whose maximum degree is bounded by $k^{\frac{1}{r}}$, where $r = 18\lceil\frac{2}{\alpha}\rceil - 5$. Let $x \in V(G)$, and let $\mathcal{R}$ be an $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph of $G-x$, with $|\mathcal{R}| \leqslant M_0$, such that at least one of the following conditions (I)–(IV) holds:

1. *$\mathcal{R}$ has a $(k \cdot \frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large nonbipartite component; or*
2. *$\mathcal{R}$ has a $(k_1 \cdot \frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large bipartite component, where k_1 is the size of the larger bipartition class of T; or*

3. $\mathcal{R}$ *has a* $(\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$*-large bipartite component such that* x $\sqrt{\varepsilon}$*-sees both sides of the bipartition; or*

4. x $\sqrt{\varepsilon}$*-sees two components* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *of* $\mathcal{R}$ *in a way that one of the following holds:*

   (a) x *sends at least one edge to a third component* $\mathcal{C}_3$ *of* $\mathcal{R}$*;*

   (b) there is $i \in \{1,2\}$ *such that* $\mathcal{C}_i$ *is nonbipartite and* $(\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$*-large;*

   (c) there is $i \in \{1,2\}$ *such that* $\mathcal{C}_i$ *is bipartite and* x *sees both sides of the bipartition;*

   (d) there is $i \in \{1,2\}$ *such that* $\mathcal{C}_i$ *is bipartite with parts* $\mathcal{A}$ *and* $\mathcal{B}$*, such that* $\min\{|\mathcal{A}|, |\mathcal{B}|\} \geqslant (1+\sqrt[4]{\varepsilon})\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}$ *and* x *sees only one side of the bipartition;*

   (e) $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *are bipartite with parts* $\mathcal{A}_1, \mathcal{B}_1$ *and* $\mathcal{A}_2, \mathcal{B}_2$*, respectively, such that* $\min\{|\mathcal{A}_1|, |\mathcal{B}_2|\} \geqslant (1+\sqrt[4]{\varepsilon})\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}$ *and* x *does not see* $\mathcal{B}_1 \cup \mathcal{B}_2$*.*

Then T *embeds in* G*.*

Proof. Let k_0' be the maximum of the outputs k_0 of Proposition 3.4.3, Corollary 3.3.4 and Corollary 3.3.12, for inputs ε, $d = \frac{6}{\alpha}$ and $2M_0$, and choose $n_0 := k_0' + 1$ as the numerical output of Lemma 3.5.3.

Now assume we are given an n-vertex graph G with $x \in V(G)$, and let T be a k-edge tree as in Lemma 3.5.3. Let $\mathcal{R}$ be the $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph of $G - x$. An easy computation shows that

$$\delta(\mathcal{R}) \geqslant (1+\tfrac{1}{2}\sqrt[4]{\varepsilon})\alpha k\cdot\frac{|\mathcal{R}|}{n} \geqslant (1+100\sqrt{\varepsilon})\alpha k\cdot\frac{|\mathcal{R}|}{n}, \tag{3.14}$$

where the last inequality follows since $\varepsilon \leqslant 10^{-10}$. Furthermore, note that $\mathcal{R}$ must fulfill one of the conditions (I)–(IV) from Lemma 3.5.3. If $\mathcal{R}$ contains a $(k\cdot\frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large nonbipartite component or a $(k_1\cdot\frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large bipartite component, then we can conclude by Proposition 3.4.3.

So we can discard scenarios (1) and (2) from Lemma 3.5.3. Therefore, by Theorem 3.4.1, and by (3.14), we can assume that every connected component $\mathcal{C}$ of $\mathcal{R}$ satisfies

$$\operatorname{diam}(\mathcal{C}) \leqslant \frac{3|\mathcal{C}|}{\delta(\mathcal{C})+1} \leqslant \frac{3(1+\sqrt[4]{\varepsilon})2k\cdot\frac{|\mathcal{R}|}{n}}{(1+\frac{1}{2}\sqrt[4]{\varepsilon})\alpha k\cdot\frac{|\mathcal{R}|}{n}} \leqslant \frac{6}{\alpha}+1, \tag{3.15}$$

and thus

$$r \geqslant 3\cdot\operatorname{diam}(\mathcal{C})+1. \tag{3.16}$$

So, the maximum degree of T and the diameter of the components are in the right relation to each other, meaning that we could apply Corollaries 3.3.4 and 3.3.12 to each connected component of $\mathcal{R}$ (if the other conditions of these corollaries hold).

In order to embed T under scenarios (3) and (4), we use the results from Section 3.2.

Case 1 (scenario (3)**):** $\mathcal{R}$ has a $(\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large bipartite component $\mathcal{C}$ such that x $\sqrt{\varepsilon}$-sees both sides of the bipartition.

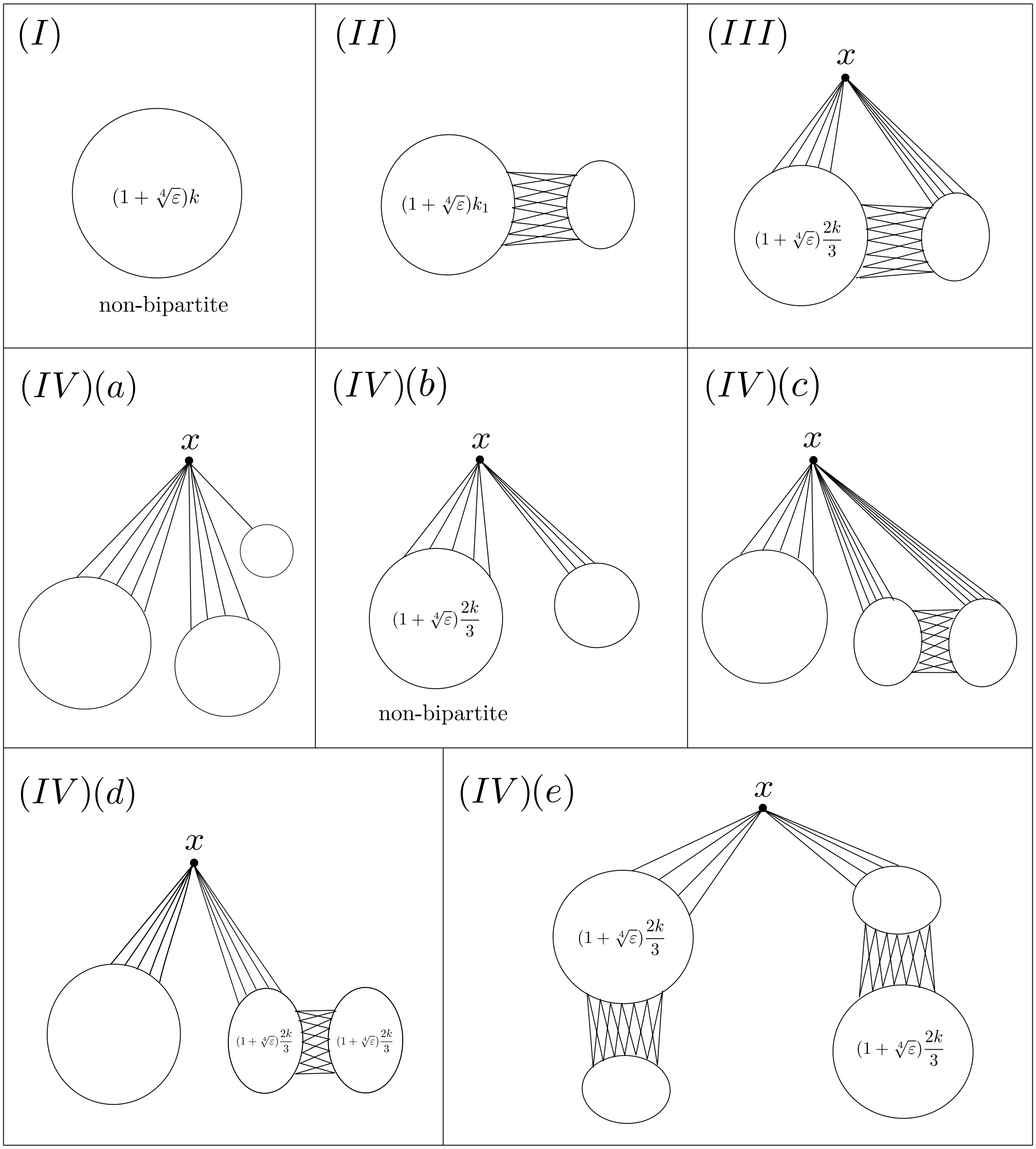

Figure 3.2: The scenarios described in Lemma 3.5.3

Applying Proposition 3.2.5 to T, we obtain a cut-vertex $z_0 \in V(T)$ and a proper 2-colouring $c : V(T - z_0) \to \{0, 1\}$ of $T - z_0$ such that

$$|c_1| \leqslant |c_0| \leqslant \frac{2k}{3} \quad \text{and} \quad |c_1| \leqslant \frac{k}{2}.$$

Let us note that, because of the bound on k_0, the number of components of $T - z_0$ is bounded by

$$\Delta(T) \leqslant k^{\frac{1}{r}} \leqslant \frac{\varepsilon k}{M_0} \leqslant \frac{\varepsilon n}{|\mathcal{R}|}. \tag{3.17}$$

Now, we map z_0 into x. Recalling (3.14), (3.15), (3.16) and the fact that $T - z_0$ is a $(\frac{2k}{3}, \frac{k}{2}, \frac{1}{r})$-forest we can apply Corollary 3.3.4 to embed $T - z_0$ into $\mathcal{C}$, and by (3.17) we may choose the images of the roots of $T - z_0$ as neighbours of x.

Case 2 (scenario (4)**):** x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{R}$.

Let $z_1 \in V(T)$ be the vertex given by Lemma 2.2.5 applied to T, with any leaf v. Let $\mathcal{T}$ be the set of connected components of $T - z_1$. Then $\mathcal{T}$ is a family of at most $\Delta(T)$ rooted trees whose roots are neighbours of z_1 in T, and $|V(T')| \leqslant \lceil \frac{k}{2} \rceil$ for every $T' \in \mathcal{T}$.

Apply Lemma 3.2.1 (i) to $\mathcal{T}$ to obtain a partition of $\mathcal{T}$ into three families of trees $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$, where $\mathcal{F}_3$ could be empty, such that

$$|V(\bigcup \mathcal{F}_3)| \leqslant |V(\bigcup \mathcal{F}_2)| \leqslant |V(\bigcup \mathcal{F}_1)| \leqslant \left\lceil \frac{k}{2} \right\rceil. \tag{3.18}$$

For later use, let us record here that

$$|\mathcal{F}_1| + |\mathcal{F}_2| + |\mathcal{F}_3| \leqslant \Delta(T) \leqslant \frac{\varepsilon n}{|\mathcal{R}|}. \tag{3.19}$$

Furthermore, due to Remark 3.2.2, we know that

$$|\mathcal{F}_3| \leqslant 1. \tag{3.20}$$

Similarly, applying Lemma 3.2.1 (ii) to $\mathcal{T}$ we obtain a partition of $\mathcal{T}$ into two families of trees $\mathcal{J}_1$ and $\mathcal{J}_2$ such that

$$|V(\bigcup \mathcal{J}_2)| \leqslant \frac{k}{2} \quad \text{and} \quad |V(\bigcup \mathcal{J}_2)| \leqslant |V(\bigcup \mathcal{J}_1)| \leqslant \frac{2k}{3}, \tag{3.21}$$

and again, we know that

$$|\mathcal{J}_1| + |\mathcal{J}_2| \leqslant \Delta(T) \leqslant \frac{\varepsilon n}{|\mathcal{R}|}. \tag{3.22}$$

We split the remainder of the proof into five cases, according to which of the conditions (4a), (4b), (4c), (4d) or (4e) holds. Depending on the case we will make use of partition $\{\mathcal{F}_i\}_{i=1,2}$ or $\{\mathcal{J}_i\}_{i=1,2,3}$.

Case 2a (scenario (4a)**):** x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$, $\mathcal{C}_2$ and sends at least one edge to a third component $\mathcal{C}_3$.

We embed z_1 into x, and then proceed to embed the roots of the trees from $\mathcal{F}_i$ into the neighbourhood of x in $\mathcal{C}_i$, for each $i \in \{1,2,3\}$. This is possible since by (3.20), there is at most one root to embed into $\mathcal{C}_3$. Furthermore, by (3.19), there are at most $\Delta(T) \leqslant \frac{\varepsilon n}{|\mathcal{R}|}$ roots to be embedded into $\mathcal{C}_i$, for $i \in \{1,2\}$. Finally, because of the minimum degree in G, and because of (3.18), we can greedily embed the remaining vertices of each forest $\mathcal{F}_i$ into $\mathcal{C}_i$.

Case 2b (scenario (4b)**):** x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$, and one of these components, say C_1, is nonbipartite and $(\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{n}, \sqrt[4]{\varepsilon})$-large.

We map z_1 into x, and then embed the roots of $\mathcal{J}_2$ into $\mathcal{C}_2$ (we know that x has enough neighbours in $\mathcal{C}_2$ because of (3.19)). We then embed the rest of $\bigcup \mathcal{J}_2$ greedily into $\mathcal{C}_2$.

For the trees from $\mathcal{J}_1$, we can make use of Corollary 3.3.12 and Lemma 3.3.8, whose conditions hold by (3.14), (3.15), (3.16) and (3.21), to map $\bigcup \mathcal{J}_1$ to $\mathcal{C}_1$.

Case 2c (scenario (4c)**):** x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$, one of these components, say $\mathcal{C}_1$, is bipartite, and x sees both sides $\mathcal{A}$, $\mathcal{B}$ of the bipartition.

First, we map z_1 into x and then embed $\bigcup \mathcal{F}_1$ greedily into $\mathcal{C}_2$ (embedding the roots into neighbours of x, as before). For the remaining forests, $\mathcal{F}_2$ and $\mathcal{F}_3$, observe that for any proper 2-colouring of $\bigcup \mathcal{F}_2$ and $\bigcup \mathcal{F}_3$, and for any $i \in \{2,3\}$, the larger colour class of $\bigcup \mathcal{F}_i$ and the smaller colour class of $\bigcup \mathcal{F}_{5-i}$ add up to at most

$$|\bigcup \mathcal{F}_i| + \frac{|\bigcup \mathcal{F}_{5-i}|}{2} \leqslant \frac{|\bigcup \mathcal{F}_1| + |\bigcup \mathcal{F}_2| + |\bigcup \mathcal{F}_3|}{2} = \frac{k}{2}. \tag{3.23}$$

Now, our aim is to embed the roots and all the even levels of $\bigcup \mathcal{F}_2$ into $\mathcal{A}$, while embedding the odd levels into $\mathcal{B}$. Moreover, we plan to embed $\bigcup \mathcal{F}_3$ in a way that its larger colour class goes to the same set as the smaller colour class of $\bigcup \mathcal{F}_2$.

As x $\sqrt{\varepsilon}$-sees $\mathcal{C}_1$, we may assume that x $\frac{\sqrt{\varepsilon}}{2}$-sees $\mathcal{A}$. Moreover, since x has at least one neighbour $b \in \bigcup B$, and since $\bigcup \mathcal{F}_3$ has only one root because of (3.20), we can choose whether we map the single root of $\bigcup \mathcal{F}_3$ into b, or into some neighbour of x in $\mathcal{A}$. We will make this choice according to our plan above (that is, it will depend on whether the even or the odd levels of $\bigcup \mathcal{F}_2$ contain more vertices).

We then greedily embed the rest of $\bigcup \mathcal{F}_3$ into $\mathcal{C}_1$. Now, we can make use of Corollary 3.3.4 together with Remark 3.3.5, whose conditions hold by (3.16) and by (3.23), to complete the embedding of $\bigcup \mathcal{F}_2$ into $\mathcal{C}_1$, while avoiding the image of $\bigcup \mathcal{F}_3$.

Case 2d (scenario (4d)**):** x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$, one of them is bipartite with parts $\mathcal{A}$ and $\mathcal{B}$, such that $\min\{|\mathcal{A}|, |\mathcal{B}|\} \geqslant (1+\sqrt[4]{\varepsilon})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{n}$ and x sees only one side of the bipartition.

Let us assume that $\mathcal{C}_1$ is the bipartite component with parts $\mathcal{A}$ and $\mathcal{B}$ containing at least $(1+\sqrt[4]{\varepsilon})\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}$ clusters each and that x only sees the set $\mathcal{A}$. We map z_1 into x and then embed $\bigcup\mathcal{J}_2$ greedily into $\mathcal{C}_2$ (embedding the roots into neighbours of x, as before). Note that there are few roots of trees in $\mathcal{J}_1\cup\mathcal{J}_2$, because of (3.22). Since $\mathcal{J}_1$ is a $(\frac{2k}{3},\frac{k}{2},\frac{1}{r})$-forest, we may apply Corollary 3.3.4 so that we can embed $\bigcup\mathcal{J}_1$ into $\mathcal{C}_1$ in a way that the images of its roots are neighbours of x. This works because of (3.21).

Case 2e (scenario (4e)**):** x $\sqrt{\varepsilon}$-sees two bipartite components $\mathcal{C}_1$ and $\mathcal{C}_2$, with parts $\mathcal{A}_1,\mathcal{B}_1$ and $\mathcal{A}_2,\mathcal{B}_2$ respectively, such that $\min\{|\mathcal{A}_1|,|\mathcal{B}_2|\}\geqslant(1+\sqrt[4]{\varepsilon})\frac{2k}{3}\cdot\frac{|\mathcal{R}|}{n}$ and x sees only $\mathcal{A}_1$ and $\mathcal{A}_2$.

We map z_1 into x, note that x $\sqrt{\varepsilon}$-sees $\mathcal{A}_1$ and $\mathcal{A}_2$. Consider the colouring φ that T induces in $\bigcup\mathcal{J}_1$. If the roots of the trees in $\mathcal{J}_1$ are contained in the heavier colour class of φ, then we embed $\bigcup\mathcal{J}_1$ into $\mathcal{C}_1$, otherwise we embed $\bigcup\mathcal{J}_1$ into $\mathcal{C}_2$. In any case, and since $\mathcal{J}_1$ is a $(\frac{2k}{3},\frac{k}{2},\frac{1}{r})$-forest, we may use Corollary 3.3.4 to embed $\bigcup\mathcal{J}_1$ (taking care of mapping the roots into neighbours of x). Finally, we greedily embed $\bigcup\mathcal{J}_2$ into the remaining component.

This completes the proof of Lemma 3.5.3. □

3.6 Embedding trees with degree conditions

In this section we prove Theorems 1.3.5, 1.3.6 and 1.3.3. All of them will be proved using Lemma 3.5.3, which, fortunately, makes all these proofs quite straightforward.

We begin by proving the approximate version of the $2k-\frac{k}{2}$ conjecture (Theorem 1.3.5) in Section 3.6.1. Then, we show the approximate version of $\frac{2}{3}$–conjecture (Theorem 1.3.3) in Section 3.6.2. In Section 3.6.3, we show Theorem 1.3.6 (our extension of Theorem 1.3.5 to constant degree trees).

3.6.1 An approximate version of the $2k-\frac{k}{2}$ conjecture

Proof of Theorem 1.3.5. Given $\delta\in(0,1)$, we set

$$\varepsilon:=\frac{\delta^4}{10^{10}},\ \text{and}\ \alpha:=\frac{1}{2}.$$

Let N_0, M_0 be given by Lemma 2.3.2, with input ε, $\eta:=5\sqrt{\varepsilon}$ and $m_0:=\frac{1}{\varepsilon}$, and let n_0' be given by Lemma 3.5.3, with input α,ε and M_0. We choose $n_0:=(1-\varepsilon)^{-1}\max\{n_0',N_0\}+1$ as the numerical output of the theorem.

Let G be an n-vertex graph as in Theorem 1.3.5, where $n \geqslant k \geqslant \delta n$ and $n \geqslant n_0$, and let $x \in V(G)$ be a vertex of degree at least $2(1+\delta)k$. Let T be a k-edge tree with maximum degree at most $k^{\frac{1}{r}}$, where $r = 67 = 18 \cdot 4 - 5$.

We apply Lemma 2.3.2 to $G - x$ so that we get a subgraph $G' \subseteq G - x$, with $|G'| \geqslant (1-\varepsilon)(n-1)$, that admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition. Moreover, the minimum degree in G' is at least

$$\delta(G') \geqslant (1+\delta)\frac{k}{2} - (\varepsilon + 5\sqrt{\varepsilon})(n-1) - 1 \geqslant (1+\sqrt[4]{\varepsilon})\frac{k}{2}. \tag{3.24}$$

Let $\mathcal{R}$ be the corresponding $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph of G'. Our aim is to show that $\mathcal{R}$ fulfills at least one of the conditions (1)–(4) from Lemma 3.5.3, for inputs α, ε and M_0. We will assume that

$$\text{all the components of } \mathcal{R} \text{ are } (k \cdot \tfrac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})\text{-small}, \tag{3.25}$$

as otherwise either we have (1) or (2) from Lemma 3.5.3, and we are done.

Since G' misses less than εn vertices from G, we have that

$$\deg_G(x, G') \geqslant 2(1+\tfrac{\delta}{2})k \geqslant 2(1+100\sqrt[4]{\varepsilon})k. \tag{3.26}$$

Suppose that x does not $\sqrt{\varepsilon}$-see any component of $\mathcal{R}$. Since $\delta n \leqslant k$ and because of (3.26), we have that

$$2\delta n \leqslant 2(1+\tfrac{\delta}{2})k \leqslant \deg_G(x, G') = \sum_{\mathcal{C}} \deg_G(x, V(\bigcup \mathcal{C})) \leqslant \sqrt{\varepsilon} n, \tag{$\heartsuit$}$$

a contradiction. Therefore, there is some component $\mathcal{C}_1$ of $\mathcal{R}$ receiving more than $\sqrt{\varepsilon}|\bigcup \mathcal{C}_1|$ edges from x.

By (3.25), x can have at most $2(1+\sqrt[4]{\varepsilon})k$ neighbours in $\bigcup \mathcal{C}_1$. So by (3.26), there are more than $\sqrt[4]{\varepsilon}k$ neighbours of x outside $\bigcup \mathcal{C}_1$. Following the same reasoning as in ($\heartsuit$), there must be a second component $\mathcal{C}_2$ receiving at least $\sqrt{\varepsilon}|\bigcup \mathcal{C}_2|$ edges from x. We can assume that x has no neighbours outside $\bigcup \mathcal{C}_1 \cup \mathcal{C}_2$, as otherwise condition (4a) from Lemma 3.5.3 holds.

By (3.26) and by symmetry, we can assume that

$$\deg_G(x, V(\bigcup \mathcal{C}_1)) \geqslant (1+\tfrac{\delta}{2})k.$$

In particular, we can again employ (3.25) to see that $\mathcal{C}_1$ is bipartite, and moreover x has to see both classes of the bipartition. Therefore, condition (4c) from Lemma 3.5.3 holds and the proof is finished. $\square$

3.6.2 An approximate version of the $\frac{2}{3}$-conjecture

Proof of Theorem 1.3.3. Given $\delta \in (0,1)$, we set

$$\varepsilon := \frac{\delta^4}{10^{10}}, \text{ and } \alpha := \frac{2}{3}.$$

Let N_0, M_0 be given by Lemma 2.3.2, with input ε, $\eta := 5\sqrt{\varepsilon}$ and $m_0 := \frac{1}{\varepsilon}$, and let n_0' be given by Lemma 3.5.3, with input α, ε and M_0. We choose $n_0 := (1-\varepsilon)^{-1}\max\{n_0', N_0\} + 1$ as the numerical output of the theorem.

Let G be an n-vertex graph as in Theorem 1.3.3, where $n \geqslant k \geqslant \delta n$ and $n \geqslant n_0$, and let $x \in V(G)$ be a vertex of degree at least $(1+\delta)k$. Let T be a k-edge tree with maximum degree at most $k^{\frac{1}{r}}$, where $r = 49 = 18 \cdot 3 - 5$.

We apply Lemma 2.3.2 to $G - x$ so that we get a subgraph $G' \subseteq G - x$, with $|G'| \geqslant (1-\varepsilon)(n-1)$, that admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition. Let $\mathcal{R}$ be the corresponding $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph of G', we will assume that every component of $\mathcal{R}$ is $(k \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-small. An easy computation shows that

$$\delta(G') \geqslant \left(1 + \tfrac{\delta}{2}\right)\frac{2k}{3} \geqslant (1 + 100\sqrt[4]{\varepsilon})\frac{2k}{3}, \tag{3.27}$$

because of the minimum degree in G. Also, note that $\deg_G(x, G') \geqslant (1 + \frac{\delta}{2})k$. Following the same reasoning as in (♡), and because of the degree of x, there is some component $\mathcal{C}_1$ of $\mathcal{R}$ such that x $\sqrt{\varepsilon}$-sees $\mathcal{C}_1$.

First, assume that x has more than $(1 + 2\sqrt[4]{\varepsilon})k$ neighbours in $\bigcup \mathcal{C}_1$. Since $\mathcal{C}_1$ is small, $\mathcal{C}_1$ must be bipartite and x must see at least a $\sqrt{\varepsilon}$-portion of both sides of the bipartition, namely $\mathcal{A}$ and $\mathcal{B}$. Then, by (3.27) we have $\max\{|\mathcal{A}|, |\mathcal{B}|\} \geqslant (1 + \sqrt[4]{\varepsilon})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}$ and, therefore, G' satisfies condition (3) from Lemma 3.5.3.

Now, we may assume that x has less than $(1 + 2\sqrt[4]{\varepsilon})k$ neighbours in $\bigcup \mathcal{C}_1$. As in (♡), we can calculate that there is a second component $\mathcal{C}_2$ containing at least $\sqrt{\varepsilon}|\bigcup \mathcal{C}_2|$ neighbours of x. We can assume that x does not send edges to any other component, otherwise we are in case (4a) from Lemma 3.5.3, and are done.

Also, by symmetry we can assume that $\deg_G(x, V(\bigcup \mathcal{C}_1)) \geqslant (1 + \frac{\delta}{2})\frac{k}{2}$. Following the same reasoning as before we conclude that $|\mathcal{C}_1| \geqslant (1 + \frac{\delta}{2})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}$. In particular, if $\mathcal{C}_1$ is nonbipartite, then G' satisfies condition (4b) from Lemma 3.5.3 and we are done.

So we may suppose that $\mathcal{C}_1$ is bipartite. If x sees both sides of the bipartition, condition (4c) from Lemma 3.5.3 holds, so let us assume this is not the case. The minimum degree tells us that one of the sides of the bipartition of $\mathcal{C}_1$ has size at least $(1 + \frac{\delta}{2})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}$ clusters, and we can argue similarly for the other side of the bipartition. This means that G' satisfies condition (4d) from Lemma 3.5.3, which completes the proof. □

3.6.3 Embedding trees with maximum degree bounded by a constant

Proof of Theorem 1.3.6. Given $\delta \in (0, 1)$ and $\Delta \geqslant 2$, we set

$$\varepsilon := \frac{\delta^4}{10^{10}} \text{ and } \alpha := \frac{1}{2}.$$

Let N_0, M_0 be given by Lemma 2.3.2, with input $\varepsilon, \eta := 5\sqrt{\varepsilon}$ and $m_0 := \frac{1}{\varepsilon}$, and let n_0' be given by Lemma 3.5.3, with input α, ε and M_0. We choose $n_0 := (1-\varepsilon)^{-1}\max\{n_0', N_0\} + 1$ as the numerical output of the theorem.

Let G be an n-vertex graph as in Theorem 1.3.6, where $n \geqslant k \geqslant \delta n$ and $n \geqslant n_0$, and let $x \in V(G)$ be a vertex of degree at least $2(\frac{\Delta-1}{\Delta} + \delta)k$. Let T be a k-edge tree with maximum degree at most Δ.

We apply Lemma 2.3.2 to $G - x$ and we obtain a subgraph G', with $|G'| \geqslant (1-\varepsilon)(n-1)$, that admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition. Let $\mathcal{R}$ be the corresponding $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph.

Observe that each k-edge tree T with maximum degree at most Δ will satisfy

$$k_1 \leqslant \frac{\Delta-1}{\Delta}k, \tag{3.28}$$

where k_1 is the size of the larger bipartition class of T. We can discard scenarios (1) and (2) and therefore assume that

$$\text{all nonbipartite components of } \mathcal{R} \text{ are } (k \cdot \tfrac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})\text{-small}, \tag{3.29}$$

and, by (3.28),

$$\text{all bipartite components of } \mathcal{R} \text{ are } (\tfrac{\Delta-1}{\Delta}k \cdot \tfrac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})\text{-small}. \tag{3.30}$$

As we removed only few vertices from G, it is clear that x has at least $2(\frac{\Delta-1}{\Delta} + \frac{\delta}{2})k$ neighbours in G'. This, together with (3.29) and (3.30), implies that there are components $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{R}$ such that

$$\deg_G(x, V(\bigcup \mathcal{C}_i)) \geqslant \sqrt{\varepsilon}|\bigcup \mathcal{C}_i|, \quad \text{for } i \in \{1, 2\}.$$

Moreover, we may assume that x does not see any other components, otherwise G' satisfies condition (4a) from Lemma 3.5.3 and we are done. First, suppose that $\Delta = 2$, that is, T is a path of length k. In this case, we choose a cut vertex z of T that splits T into two paths of length $\frac{k}{2}$ and then we embed z into x. After that, we can greedily embed each component of $T - z$ into $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

Now, suppose that $\Delta \geqslant 3$. By symmetry, we may assume that

$$\deg_G(x, V(\bigcup \mathcal{C}_1)) \geqslant (\tfrac{\Delta-1}{\Delta} + \tfrac{\delta}{2})k. \tag{3.31}$$

If $\mathcal{C}_1$ is nonbipartite, G' satisfies condition (4b) from Lemma 3.5.3 as $\Delta \geqslant 3$. If $\mathcal{C}_1$ is bipartite with parts $\mathcal{A}$ and $\mathcal{B}$, we can employ (3.30) together with (3.31) to conclude that G' satisfies condition (4c) from Lemma 3.5.3. This concludes the proof. $\square$

3.7 An approximate version of the intermediate range conjecture

In this section, we prove an approximate version of the *intermediate range conjecture* (Theorem 1.3.8). The proof is based on a structural result for graphs with minimum degree above

$\frac{k}{2}$ and maximum degree above $\frac{4k}{3}$ avoiding some tree with k edges and bounded degree. Let us start with the definition of the *extremal graphs*.

Definition 3.7.1 ((ε, x)-extremal)**.** *Let $\varepsilon > 0$ and let $k \in \mathbb{N}$. Given a graph G and a vertex $x \in V(G)$, we say that G is* (ε, x)-extremal *if for every $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$ of $G - x$ the following conditions hold:*

(i) every component of R is $(k \cdot \frac{|\mathcal{R}|}{|G|}, \sqrt[4]{\varepsilon})$-small;

(ii) x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{R}$ and x does not see any other component of $\mathcal{R}$;

and furthermore, assuming that $\deg(x, \bigcup V(\mathcal{C}_1)) \geqslant \deg(x, \bigcup V(\mathcal{C}_2))$,

(iii) $\mathcal{C}_1$ is bipartite and $(\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G|}, \sqrt[4]{\varepsilon})$-large, with x only seeing the larger side of $\mathcal{C}_1$;

(iv) if $\mathcal{C}_2$ is non-bipartite, then $\mathcal{C}_2$ is $(\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G|}, \sqrt[4]{\varepsilon})$-small, and if $\mathcal{C}_2$ is bipartite, then x sees only one side of the bipartition.

Now we will prove that a graph of minimum degree above $\frac{k}{2}$ and maximum degree above $\frac{4k}{3}$ either contains every tree with k edges and bounded degree or is (ε, x) extremal for each vertex x of high degree.

Theorem 3.7.2. *For all $\delta \in (0, 1)$ there is $n_0 \in \mathbb{N}$ such that for all $k, n \geqslant n_0$ with $n \geqslant k \geqslant \delta n$, the following holds for every n-vertex graph G with $\delta(G) \geqslant (1+\delta)\frac{k}{2}$ and $\Delta(G) \geqslant (1+\delta)\frac{4k}{3}$. If T is a tree with k edges such that $\Delta(T) \leqslant k^{\frac{1}{67}}$, then either*

(a) T embeds in G; or

*(b) G is $(\frac{\delta^4}{10^{10}}, x)$-*extremal *for every $x \in V(G)$ of degree at least $(1 + \delta)\frac{4k}{3}$.*

Proof. Given $\delta \in (0, 1)$, we set
$$\varepsilon := \frac{\delta^4}{10^{10}}. \tag{3.32}$$

Let N_0, M_0 be given by Lemma 2.3.2, with input ε, $\eta := 5\sqrt{\varepsilon}$ and $m_0 := \frac{1}{\varepsilon}$, and let n_0' be given by Lemma 3.5.3, with input ε and M_0. We choose
$$n_0 := (1 - \varepsilon)^{-1} \max\{n_0', N_0\} + 1$$
as the numerical output of Theorem 3.7.2.

Let G and T be given as in Theorem 3.7.2. Consider an arbitrary vertex $x \in V(G)$ with $\deg(x) \geqslant (1 + \delta)\frac{4}{3}k$, and apply Lemma 2.3.2 to $G - x$. We obtain a subgraph $G' \subseteq G - x$ which admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition of $G - x$, with corresponding $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph $\mathcal{R}$. Note that
$$\delta(G') \geqslant (1 + \tfrac{\delta}{2})\frac{k}{2} \geqslant (1 + 100\sqrt[4]{\varepsilon})\frac{k}{2}.$$

If $\mathcal{R}$ has a $(k\cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-large component, we are either in scenario (1) or (2) from Lemma 3.5.3, and we can embed T. So let us assume this is not the case. In particular, we can assume that condition (i) of Definition 3.7.1 holds.

Since G' misses less than $\varepsilon n + 1$ vertices from G, we have that

$$\deg_G(x, G') \geqslant (1 + \tfrac{\delta}{2})\frac{4}{3}k \geqslant (1 + 100\sqrt[4]{\varepsilon})\frac{4}{3}k. \tag{3.33}$$

It is clear that x has to $\sqrt{\varepsilon}$-see at least one component $\mathcal{C}_1$ of $\mathcal{R}$. Indeed, otherwise, we would have that

$$\frac{4}{3}\delta n \leqslant \frac{4}{3}k \leqslant \deg_G(x, G') = \sum_{\mathcal{C}} \deg_G(x, \bigcup V(\mathcal{C})) \leqslant \sqrt{\varepsilon} n, \tag{3.34}$$

where the sum is over all components $\mathcal{C}$ of $\mathcal{R}$, and this contradicts (3.32). Suppose that x sees only one component. Since $\mathcal{C}_1$ is $(k \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-small and $\deg_G(x, \bigcup V(C)) \geqslant (1 + \frac{\delta}{2})\frac{4k}{3}$, it follows that $\mathcal{C}_1$ is bipartite and thence the largest bipartition class of $\mathcal{C}_1$ has size at least $(1 + \frac{\delta}{2})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}$ and x $\sqrt{\varepsilon}$-sees both bipartition classes. Therefore we are in scenario (3) from Lemma 3.5.3 and thus we can embed T.

Suppose from now that x sends edges outside of $\mathcal{C}_1$. Since $\mathcal{C}_1$ is $(k \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-small, it follows that

$$\deg_G(x, G' \setminus \bigcup V(\mathcal{C}_1)) \geqslant (1 + 50\sqrt[4]{\varepsilon})\frac{k}{3}. \tag{3.35}$$

We claim that x $\sqrt{\varepsilon}$-sees at least two components of $\mathcal{R}$. Indeed, since $k \geqslant \delta n$ and from (3.35) we have

$$\frac{\delta n}{3} \leqslant (1 + 50\sqrt[4]{\varepsilon})\frac{k}{3} \leqslant \sum_{\mathcal{C} \neq \mathcal{C}_1} \deg_G(x, \bigcup V(\mathcal{C})) \leqslant \sqrt{\varepsilon} n,$$

which contradicts (3.32).

If x sends at least one edge to a third component, then we are in scenario (4a) from Lemma 3.5.3 and thus T can be embedded. Therefore, we know that x actually $\sqrt{\varepsilon}$-sees exactly two components, which we will call $\mathcal{C}_1$ and $\mathcal{C}_2$ (In particular, we know that condition (ii) of Definition 3.7.1 holds). By symmetry, we may assume that $\deg(x, \bigcup V(\mathcal{C}_1)) \geqslant \deg(x, \bigcup V(\mathcal{C}_2))$ and thus, by (3.33),

$$\deg(x, \bigcup V(\mathcal{C}_1)) \geqslant (1 + 100\sqrt[4]{\varepsilon})\frac{2k}{3}. \tag{3.36}$$

Thus, if $\mathcal{C}_1$ is non-bipartite we are in scenario (4b) from Lemma 3.5.3, and therefore, we can assume $\mathcal{C}_1 = (\mathcal{A}_1, \mathcal{B}_1)$ is bipartite. Also, x only sees one side of the bipartition, say $\mathcal{A}_1$, since otherwise we are in scenario (4c). Moreover, by (3.36), and since we may assume we are not in scenario (4d), we know that

$$|\mathcal{A}_1| \geqslant (1 + 100\sqrt[4]{\varepsilon})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|} \quad \text{and} \quad |\mathcal{B}_1| \leqslant (1 + \sqrt[4]{\varepsilon})\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}. \tag{3.37}$$

So, condition (iii) of Definition 3.7.1 holds.

Furthermore, if $\mathcal{C}_2$ is non-bipartite, then it is $(\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-small, as otherwise we are in case (4b). If $\mathcal{C}_2$ is bipartite, then x can only see one side of the bipartition, since otherwise we are in scenario (4c). Therefore, $\mathcal{C}_2$ satisfies condition (iv) of Definition 3.7.1, implying that G is $(\frac{\delta^4}{10^{10}}, x)$-extremal. $\square$

Now we are ready for the proof of Theorem 1.3.8.

Proof of Theorem 1.3.8. Given $\delta \in (0,1)$, we set

$$\varepsilon := \frac{\delta^4}{10^{10}}$$

and apply Lemma 2.3.2, with inputs ε, $\eta = 5\sqrt{\varepsilon}$ and $m_0 := \frac{1}{\varepsilon}$, to obtain numbers n_0 and M_0. Next, apply Corollary 3.3.4, with input ε and further inputs $d := 3$ and M_0 to obtain a number k_0'. Choose k_0 as the larger of n_0, k_0' and the output of Theorem 3.7.2.

Now, let $k, n \in \mathbb{N}$, let $\alpha \in [0, \frac{1}{3})$, let T be a tree and let G be a graph as in Theorem 1.3.8. Let x be an arbitrary vertex of maximum degree in G. Note that

$$\deg_G(x) \geqslant 2(1+\delta)(1-\alpha)k \geqslant (1+\delta)\tfrac{4k}{3}.$$

We apply the regularity lemma (Lemma 2.3.2) to $G - x$ to obtain a subgraph $G' \subseteq G - x$ which admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition with a corresponding reduced graph $\mathcal{R}$. Moreover, since G' misses only few vertices from G, we know that

$$\deg_G(x, G') \geqslant 2(1+\tfrac{\delta}{2})(1-\alpha)k \tag{3.38}$$

and

$$\delta(G') \geqslant (1+\tfrac{\delta}{2})(1+\alpha)\frac{k}{2}, \tag{3.39}$$

and thus

$$\delta(\mathcal{R}) \geqslant (1+\tfrac{\delta}{2})(1+\alpha)\frac{k}{2} \cdot \frac{|\mathcal{R}|}{|G'|}. \tag{3.40}$$

Apply Theorem 3.7.2 to T and G. This either yields an embedding of T, which would be as desired, or tells us that G is an (ε, x)-extremal graph. We assume the latter from now on.

So, we know that x $\sqrt{\varepsilon}$-sees two components $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{R}$, where $\mathcal{C}_1 = (\mathcal{A}, \mathcal{B})$ is bipartite, say with $|\mathcal{A}| \geqslant |\mathcal{B}|$. Moreover, x does not see any other component of $\mathcal{R}$. Furthermore,

(A) $\mathcal{C}_i$ is $(k \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-small, for $i \in \{1,2\}$; and

(B) $\mathcal{C}_1$ is $(\frac{2k}{3} \cdot \frac{|\mathcal{R}|}{|G'|}, \sqrt[4]{\varepsilon})$-large, and x does not see $\mathcal{B}$.

By (3.38), and since we assume that x sends more edges to $\bigcup V(\mathcal{C}_1)$ than to $\bigcup V(\mathcal{C}_2)$, we know that

$$\deg_G(x, \textstyle\bigcup V(\mathcal{C}_1)) \geqslant (1+\tfrac{\delta}{2})(1-\alpha)k, \tag{3.41}$$

and thus, by (B),

$$|\mathcal{C}_1| \geqslant |\mathcal{A}| \geqslant (1+\tfrac{\delta}{2})(1-\alpha)k \cdot \frac{|\mathcal{R}|}{|G'|}, \tag{3.42}$$

since x has at least that many neighbours in $\mathcal{A}$, because of inequality (3.41).

Also, note that because of (A) and because of the bound (3.40), we know that any pair of clusters from the same bipartition class of $\mathcal{C}_1$ has a common neighbour. Therefore,

$$\text{the diameter of } \mathcal{C}_1 \text{ is bounded by 3.} \tag{3.43}$$

Let us now turn to the tree T. We apply Lemma 2.2.5 to find a cut vertex z of T such that every component of $T - z$ has size at most $\lceil \frac{k}{2} \rceil$. Let $\mathcal{F}$ denote the set of all components of $T - z$. Then

$$\text{each component of } \mathcal{F} \text{ has size at most } \left\lceil \frac{t}{2} \right\rceil. \tag{3.44}$$

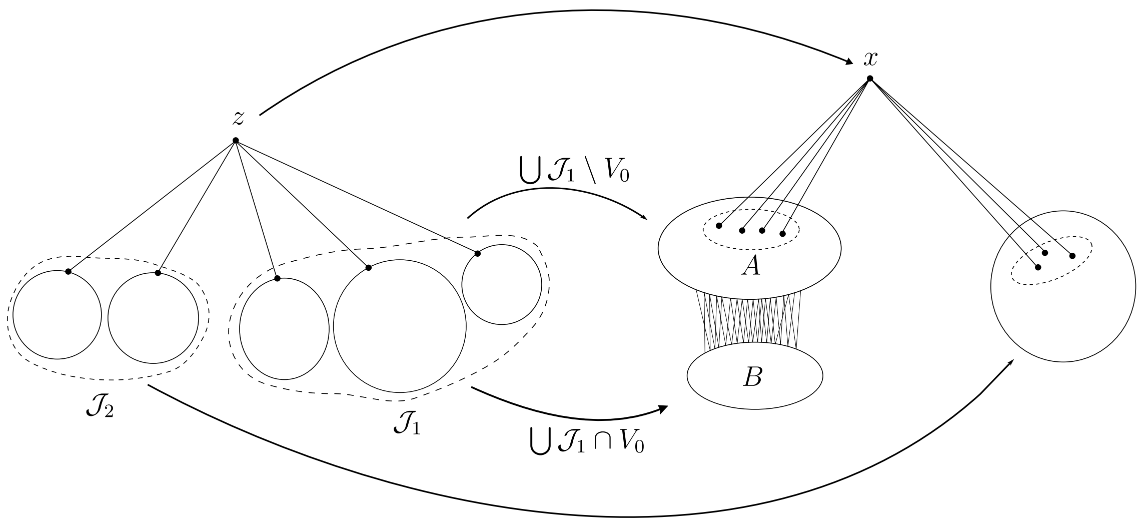

Figure 3.3: Embedding if (3.45) does not to hold.

Let V_0 denote the set of all vertices of $T - z$ that lie at even distance to z. We claim that if we cannot embed T, then

$$|V_0| \geqslant (1+\alpha)\frac{k}{2}. \tag{3.45}$$

Indeed, suppose otherwise. Then we can apply Lemma 3.2.1 to obtain a partition of $\mathcal{F}$ into two sets $\mathcal{J}_1$ and $\mathcal{J}_2$ such that

$$|\bigcup \mathcal{J}_1| \leqslant \frac{2}{3}k \quad \text{and} \quad |\bigcup \mathcal{J}_2| \leqslant \frac{k}{2}.$$

We embed z into x. Our plan is to use Corollary 3.3.4 with reduced host graph $\mathcal{C}_1$, and with

$$k_1 + k_2 := |\bigcup \mathcal{J}_1| \leqslant \frac{2}{3}k$$

where $k_1 := |\bigcup \mathcal{J}_1 \setminus V_0|$ and $k_2 := |\bigcup \mathcal{J}_1 \cap V_0|$ are the sizes of the two bipartition classes of $\bigcup \mathcal{J}_1$. Since we assumed (3.45) does not to hold, we have

$$k_2 \leqslant |V_0| \leqslant (1+\alpha)\frac{k}{2}. \tag{3.46}$$

We now embed $\bigcup \mathcal{J}_1$ into $\mathcal{C}_1$, with the roots of $\mathcal{J}_1$ embedded in the neighbourhood of x. Observe that condition (iii) of Corollary 3.3.4 holds because of (3.40) and (3.46), and condition (iv) holds because of (3.42). Moreover, the neighbourhood of x is large enough to accommodate the roots of the trees from $\mathcal{J}_1$ because of (3.41) and the bound on $\Delta(T)$. In order to see condition (ii) of Corollary 3.3.4, it suffices to recall (ii).

Also, because of (3.39), and since x also $\sqrt{\varepsilon}$-sees the component $\mathcal{C}_2$, we can embed the trees from $\mathcal{J}_2$ into $\mathcal{C}_2$. We do this by first mapping the roots of the trees from $\mathcal{J}_2$ into the neighbourhood of x in $\mathcal{C}_2$. Then, since the minimum degree of G' is larger than $|\bigcup \mathcal{J}_2|$ we may complete the embedding of $\bigcup \mathcal{J}_2$ greedily. In this way, we have embedded all of T, as desired.

So, from now we can and will assume that (3.45) holds. We split the remainder of the proof into two complementary cases, which will be solved in different ways. Our two cases are defined according to whether or not there is a tree $F^* \in \mathcal{F}$ such that $|V(F^*) \cap V_0| > \alpha k$. Let us first treat the case where such a tree F^* does not exist.

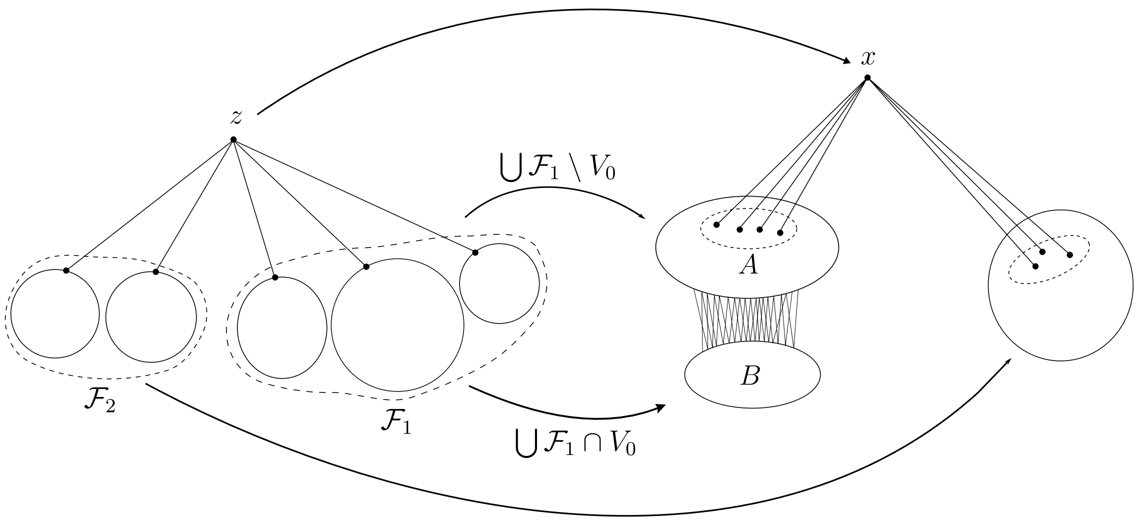

Figure 3.4: Embedding in Case 1.

Case 1: $|V(F) \cap V_0| \leqslant \alpha k$ for each $F \in \mathcal{F}$.

In this case, we proceed as follows. First, we embed z into x. We take an inclusion-maximal subset $\mathcal{F}_1$ of $\mathcal{F}$ such that

$$|\bigcup \mathcal{F}_1 \cap V_0| \leqslant (1+\alpha)\frac{k}{2} \tag{3.47}$$

holds. Then, because of the maximality of $\mathcal{F}_1$ and our assumption on $|V(F) \cap V_0|$ for the trees $F \in \mathcal{F}$, we know that

$$|\bigcup \mathcal{F}_1 \cap V_0| \geqslant (1-\alpha)\frac{k}{2}. \tag{3.48}$$

Hence, the trees from $\mathcal{F}_1$ can be embedded into $\mathcal{C}_1$, by using Corollary 3.3.4 as before, with $k_1 + k_2 := |\bigcup \mathcal{F}_1|$ where $k_1 := |\bigcup \mathcal{F}_1 \setminus V_0|$ and $k_2 := |\bigcup \mathcal{F}_1 \cap V_0|$. Indeed, inequalities (3.47) and (3.40) ensure that condition (iii) of the lemma holds. Furthermore, because of (3.42) and (3.48), we know that

$$k_1 = |\bigcup \mathcal{F}_1 \setminus V_0| \leqslant (1+\alpha)\frac{k}{2} \leqslant \frac{1}{1+\frac{\delta}{2}}|\bigcup V(\mathcal{A})|,$$

and hence, it is clear that also condition (iv) of Corollary 3.3.4 holds.

Condition (ii) of Corollary 3.3.4 holds because of (ii). Finally, inequality (3.41) ensures we can embed $\mathcal{F}_1$ in $\mathcal{C}_1$ in such a way the roots of $\mathcal{F}_1$ are embedded into the neighbourhood of x in $\mathcal{C}_1$.

Now, the trees from $\mathcal{F}_2 := \mathcal{F} \setminus \mathcal{F}_1$ can be embedded into $\mathcal{C}_2$. First, embed the neighbours of z into the neighbourhood of x in $\mathcal{C}_2$. Then, observe that (3.48) implies that

$$|\bigcup \mathcal{F}_2| \leqslant (1+\alpha)\frac{k}{2} \leqslant \delta(G').$$

Therefore, we can embed the remainder of the trees from $\mathcal{F}_2$ into $\mathcal{C}_2$ in a greedy fashion.

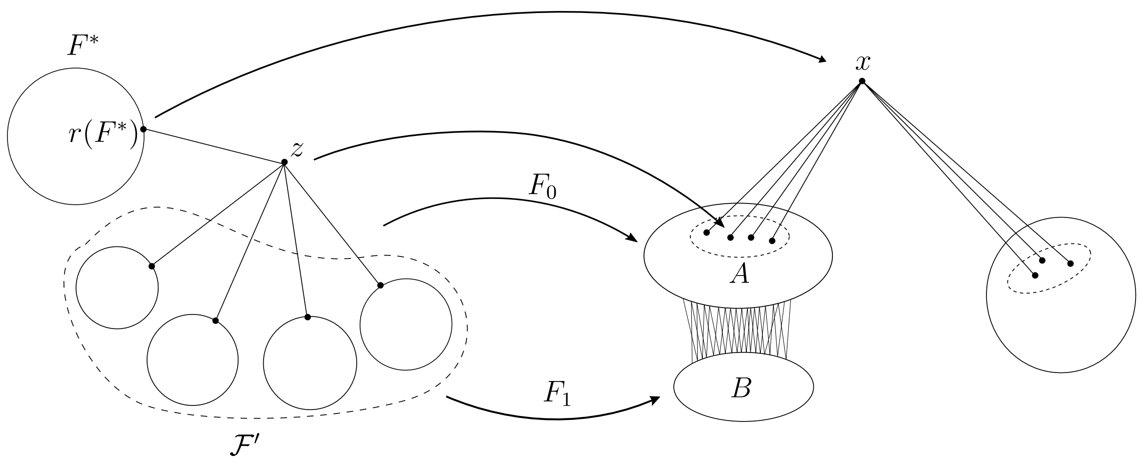

Figure 3.5: Embedding in Case 2.

Case 2: There is a tree $F^* \in \mathcal{F}$ such that $|V(F^*) \cap V_0| > \alpha k$.

In this case, let us set $\mathcal{F}' := \mathcal{F} \setminus \{F^*\}$ and note that

$$|\bigcup \mathcal{F}' \cap V_0| \leqslant (1-\alpha)k. \tag{3.49}$$

Our plan is to embed z into a neighbour of x in $\mathcal{A}$, and embed all trees from $\mathcal{F}'$ into $\mathcal{C}_1$. We then complete the embedding by mapping the root of F^* to x, and the rest of F^* to $\mathcal{C}_2$.

For the embedding of $\{z\} \cup \bigcup \mathcal{F}'$, we will use Corollary 3.3.4 as before, but this time the roles of $\mathcal{A}$ and $\mathcal{B}$ will be reversed. That is, all of

$$F_0 := (\{z\} \cup \bigcup \mathcal{F}') \cap V_0$$

is destined to go to $\mathcal{A}$, while all of

$$F_1 := (\{z\} \cup \bigcup \mathcal{F}') \setminus V_0$$

is destined to go to $\mathcal{B}$.

We choose $k_1 + k_2 := |\bigcup \mathcal{F}'| + 1$ where k_1 and k_2 are the sizes of the bipartition classes of $\{z\} \cup \bigcup \mathcal{F}'$, that is, we set $k_1 := |F_0|$ and $k_2 := |F_1|$. Because of (3.45), there are at most $(1-\alpha)\frac{k}{2}$ vertices in $T - z$ lying at odd distance from z. In particular, $k_2 \leqslant (1-\alpha)\frac{k}{2}$. So, by (3.40), we know that condition (iii) of Corollary 3.3.4 holds (and condition (i) is obviously true).

Now, condition (iv) of Corollary 3.3.4 is ensured by inequality (3.49) together with (3.42). Observe that condition (ii) of Corollary 3.3.4 holds because of (ii). Therefore, we can embed all of $\{z\} \cup \bigcup \mathcal{F}'$ with the help of Corollary 3.3.4. Furthermore, we can make sure that z is embedded into a neighbour of x.

It remains to embed the tree F^*. We embed its root $r(F^*)$ into x, and embed all the neighbours of $r(F^*)$ into arbitrary neighbours of x in $\mathcal{C}_2$. We then embed the rest of F^* greedily inyo $\mathcal{C}_2$. Note that this is possible, since by (3.44), we know that

$$|F^* - r(F^*)| \leqslant \left\lceil \frac{k}{2} \right\rceil - 1,$$

and so, our bound (3.39) guarantees that the minimum degree in $\bigcup \mathcal{C}_2$ is large enough to embed the remainder of F^* greedily into $\mathcal{C}_2$. □

Chapter 4

On the Erdős–Sós conjecture for bounded degree trees

Based on joint work with Guido Besomi and Maya Stein [18]

In this chapter we present a proof of the Erdős–Sós conjecture for trees with maximum degree bounded by a given constant and dense host graphs (Theorem 1.1.2). As an application, in Section 4.5 we present a new upper bound on the multicolour Ramsey number for bounded degree trees.

Our proof of Theorem 1.1.2 relies on a stability analysis of the structure of dense graphs with average degree above $k-1$ avoiding some tree with k edges and bounded maximum degree. Namely, we will prove that if a graph G, satisfying the conditions of Theorem 1.1.2, does not contain some tree T with k edges and bounded maximum degree, then G looks like a union of *extremal graphs*. In that case, we may use a single edge of G to connect two of those extremal graphs to embed T there.

In order to prove this structural result we use the regularity method. Let G be a graph with n vertices and $d(G) > k-1$, where $n > k \geqslant \delta n$, and let us further assume the size of G is considerable larger than k. We apply the regularity lemma to G to obtain a regular partition. We know that the corresponding reduced graph $\mathcal{R}$ roughly preserves the average degree of G. We first prove an approximate version of Theorem 1.1.2 using our embedding lemma (Lemma 3.5.3). This approximate result turns to imply that $\mathcal{R}$ has average degree roughly $k \cdot \frac{|\mathcal{R}|}{n}$, and then we can prove that each connected component of $\mathcal{R}$ has roughly the same average degree. Let $\mathcal{C}$ be a connected component of $\mathcal{R}$. If $\mathcal{C}$ is large enough, we can show that if $\mathcal{C}$ is either bipartite or contains a useful matching structure, then we can embed any given k-edge tree T with bounded degree into $\mathcal{C}$ using the tools from Section 3.3. Otherwise, the reduced graph is a union of graphs corresponding to the description given in Section 1.1, that is, graphs which are almost complete and of size roughly k or balanced almost complete bipartite graphs of size roughly $2k$.

If, on the other hand, the order of G is very close to k or if the host graph is close to being a bipartite graph of order $2k$, then a different approach is needed. To take care of these cases, we prove Theorem 4.0.1. This result might be of independent interest, as it greatly improves the main result from [54] for bounded degree trees. Note that given a graph G with $d(G) > k-1$, a standard argument[1] shows that G has a subgraph of minimum degree $\delta(G) \geqslant \frac{k}{2}$ that preserves the average degree. So, since in the Erdős-Sós conjecture and all our theorems, we are looking for subgraphs, we may always assume that in addition to the average degree condition, G satisfies a minimum degree condition. (In particular, this is assumed in Theorem 4.0.1.)

Before stating Theorem 4.0.1 we need the following definition. Given $\beta > 0$, we say that a graph H is β-bipartite if there is a partition $V(H) = A \cup B$ such that $e(A)+e(B) \leqslant \beta e(H)$.

Theorem 4.0.1. *For each $k, \Delta \in \mathbb{N}$ and each graph G with $d(G) > k-1$ and $\delta(G) \geqslant \frac{k}{2}$ the following holds.*

(a) If $k \geqslant 10^6$ and $|G| \leqslant (1+10^{-11})k$ then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \frac{\sqrt{k}}{1000}$.

(b) If $k \geqslant 8\Delta^2$ and $G = (A, B)$ is $\frac{1}{50\Delta^2}$-bipartite with $|A|, |B| \leqslant (1+\frac{1}{25\Delta^2})k$ then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.

This chapter is organized as follows. In Section 4.1 we proved some tools needed to prove that if a graph with average degree greater than $k-1$ contains no copy of some k-edge tree with bounded degree, then the its degree must be concentrated around its mean. In Section 4.3 we proved that average degree a bit lower than k is enough to ensure the containment every k-edge tree with bounded degree if the host graph is sufficiently "nice". We put everything together in Section 4.4 to prove Theorem 1.1.2. Finally, in Section 4.5 we show a consequence of Theorem 1.1.2 in Ramsey theory for trees.

4.1 Tools

In this section, we collect some of the tools that will allow us to analyse the structure of graphs avoiding some tree with bounded degree. We first prove an approximate version of the Erdős–Sós conjecture for trees of bounded degree and dense host graphs.

Lemma 4.1.1. *For all $\Delta \geqslant 2$ and $\delta, \theta \in (0,1)$, there is $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ and $k \in \mathbb{N}$ with $n > k \geqslant \delta n$ the following holds. Let G be an graph on n vertices such that $d(G) \geqslant (1+\theta)k$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.*

Proof. Let $\varepsilon > 0$ be a sufficiently small constant so that $\varepsilon \ll \theta, \delta$. Let N_0, M_0 be the numerical outputs of the regularity lemma (Lemma 2.3.2) with parameters ε and $m_0 = \frac{1}{\varepsilon}$. Let G be a graph on $n \geqslant N_0$ vertices such that $d(G) \geqslant (1+\theta)k$, where $n > k \geqslant \delta n$. Moreover, we may assume that $\delta(G) \geqslant (1+\theta)\frac{k}{2}$. By the regularity lemma, there exists a subgraph $G' \subseteq G$, with

[1] We iteratively remove from G vertices of degree less than $\frac{k}{2}$. This will not affect the average degree, and result in the desired minimum degree, unless we end up removing all vertices. However, that cannot happen, as then $|E(G)| < \frac{k}{2} \cdot n \leqslant d(G) \cdot \frac{n}{2}$, a contradiction.

$|G'| \geqslant (1-\varepsilon)n$, such that G' admits an $(\varepsilon, 5\sqrt{\varepsilon})$-regular partition $V(G') = V_0 \cup V_1 \cup \cdots \cup V_\ell$, where $m_0 \leqslant \ell \leqslant M_0$. Moreover, by the choice of ε we have

$$\delta(G') \geqslant \left(1+\frac{\theta}{2}\right)\frac{k}{2} \text{ and } d(G') \geqslant \left(1+\frac{\theta}{2}\right)k.$$

Let $\mathcal{R}$ be the corresponding $(\varepsilon, 5\sqrt{\varepsilon})$-reduced graph. Then, by Fact 2.3.3, we have

$$\delta(\mathcal{R}) \geqslant \left(1+\frac{\theta}{2}\right)\frac{k}{2}\cdot\frac{|\mathcal{R}|}{n} \text{ and } d(\mathcal{R}) \geqslant \left(1+\frac{\theta}{2}\right)k\cdot\frac{|\mathcal{R}|}{n}.$$

Let $\mathcal{C}_1, \dots, \mathcal{C}_t$ be the collection of connected components of $\mathcal{R}$. By averaging, there exists some component $\mathcal{C}_i$ such that

$$\delta(\mathcal{C}_i) \geqslant \left(1+\frac{\theta}{2}\right)\frac{k}{2}\cdot\frac{|\mathcal{R}|}{n} \text{ and } d(\mathcal{C}_i) \geqslant \left(1+\frac{\theta}{2}\right)k\cdot\frac{|\mathcal{R}|}{n},$$

and thus $\mathcal{C}_i$ is large enough in order to use either Proposition 3.3.1 or 3.3.9 to conclude. □

Now we prove two results regarding the concentration of a given function around its mean value. Given $N \in \mathbb{N}$ and a function $f : [N] \to \mathbb{R}$, we write

$$\|f\|_\infty = \max_{n\in[N]} |f(n)|$$

for the infinity norm of f. If μ is a probability measure on $[N]$ then

$$\mathbb{E}_\mu[f] = \sum_{n\in[N]} f(n)\mu(n)$$

denotes the expectation of f under μ, and if μ is the uniform probability we write

$$\mathbb{E}_{n\in[N]} f(n) = \frac{1}{N}\sum_{n\in[N]} f(n).$$

Lemma 4.1.2. *Let $N \in \mathbb{N}$, $t \in \mathbb{R}$ and $\varepsilon \in (0,1)$. Let μ be a probability measure on $[N]$ and let $f : [N] \to \mathbb{R}_+$ satisfying $\sqrt{\varepsilon}\|f\|_\infty < t \leqslant \mathbb{E}_\mu(f)$. Then at least one of the following holds*

(i) $\mu(\{n : f(n) > (1+\sqrt{\varepsilon})t\}) \geqslant \varepsilon$, *or*

(ii) $\mu(\{n : f(n) > (1-\sqrt[4]{\varepsilon})t\}) \geqslant 1-\sqrt[4]{\varepsilon}$.

Proof. Let A be the set of all $n \in [N]$ with $f(n) > (1+\sqrt{\varepsilon})t$ and set $B := [N] \setminus A$. Suppose that (i) does not hold. Then $\mu(A) \leqslant \varepsilon$, and therefore,

$$\sum_{n\in B} \mu(n)f(n) = \mathbb{E}_\mu(f) - \sum_{n\in A}\mu(n)f(n) \geqslant t - \mu(A)\|f\|_\infty \geqslant (1-\sqrt{\varepsilon})t. \tag{4.1}$$

Let B_1 be the set of all $n \in B$ such that $f(n) < (1+\sqrt{\varepsilon}-2\sqrt[4]{\varepsilon})t$, and set $B_2 := B \setminus B_1$. From (4.1) and the definition of B we deduce that

$$(1-\sqrt{\varepsilon})t \leqslant (1+\sqrt{\varepsilon})t\cdot\mu(B) - 2\sqrt[4]{\varepsilon}t\cdot\mu(B_1) \leqslant (1+\sqrt{\varepsilon})t - 2\sqrt[4]{\varepsilon}t\cdot\mu(B_1),$$

and hence, $\mu(B_1) \leqslant \sqrt[4]{\varepsilon}$. Therefore, $\mu(A\cup B_2) \geqslant 1-\mu(B_1) \geqslant 1-\sqrt[4]{\varepsilon}$, which implies (ii). □

In the proof of Theorem 1.1.2, we will use Lemma 4.1.2 with $f(x) = \deg_G(x)$ for some graph G. A useful corollary of Lemma 4.1.2 is the case when we have an even better upper bound on the $\|\cdot\|_\infty$ norm of the function f and μ is the uniform measure.

Lemma 4.1.3. *Let $N \in \mathbb{N}$, and let $\varepsilon \in (0, \frac{1}{2})$. Let $f : [N] \to \mathbb{R}_+$ be a function and let $t > 0$ such that $t \leqslant \mathbb{E}_{n\in[N]} f(n)$ and $\|f\|_\infty \leqslant (1+\varepsilon)t$. Then $f(n) \geqslant (1-\sqrt{\varepsilon})t$ for every n in a set of size at least $(1-\sqrt{\varepsilon})N$.*

4.2 Small host graph

In this section we prove Theorem 4.0.1, which will follow directly from Propositions 4.2.1 and 4.2.3. We first deal in Proposition 4.2.1 the case when the host graph is almost bipartite. Recall that H is β-bipartite if at least a $(1-\beta)$-fraction of its edges lie between A and B.

Proposition 4.2.1. *Let $k, \Delta \in \mathbb{N}$ such that $k \geqslant 8\Delta^2$. Let $G = (A, B)$ be a $\frac{1}{50\Delta^2}$-bipartite graph, with $|A|, |B| \leqslant (1 + \frac{1}{25\Delta^2})k$, $d(G) > k-1$ and $\delta(G) \geqslant \frac{k}{2}$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.*

Proof. Set $\varepsilon := \frac{1}{25\Delta^2}$ and write $n = |V(G)|$. Then, $n \leqslant (1+\varepsilon)2k$. Since G is $\frac{\varepsilon}{2}$-bipartite, we know that $e(A, B) \geqslant (1-\varepsilon)\frac{kn}{2}$. Suppose that $|B| \geqslant \frac{n}{2} \geqslant |A|$. Then

$$\frac{1}{|A|}\sum_{a\in A} \deg(a, B) > \frac{(1-\varepsilon)kn}{2|A|} \geqslant (1-\varepsilon)k, \tag{4.2}$$

and thus $|B| \geqslant (1-\varepsilon)k$. Furthermore, since $n = |A| + |B|$, we have

$$|A||B| \geqslant e(A, B) \geqslant (1-\varepsilon)\frac{kn}{2} \geqslant (1-\varepsilon)k\sqrt{|A||B|},$$

and thus, the fact that $|B| \leqslant (1+\varepsilon)k$ implies that $|A| \geqslant \frac{(1-\varepsilon)^2}{1+\varepsilon}k \geqslant (1-3\varepsilon)k$. Now we can give a lower bound for the average degree from B to A by using the first inequality from (4.2) and the fact that $n = |A| + |B|$ to calculate

$$\frac{1}{|B|}\sum_{b\in B} \deg(b, A) > (1-\varepsilon)\frac{k}{2}\left(1 + \frac{|A|}{|B|}\right) \geqslant \frac{1-\varepsilon}{2}\left(1 + \frac{1-3\varepsilon}{1+\varepsilon}\right)k \geqslant (1-4\varepsilon)k. \tag{4.3}$$

Using Lemma 4.1.3 with $f_A(a) = \deg(a, B)$ for $a \in A$, $t_A = (1-\varepsilon)k$ and $\varepsilon_A = 4\varepsilon$, and with $f_B(b) = \deg(b, A)$ for $b \in B$, $t_B = (1-3\varepsilon)k$ and $\varepsilon_B = 9\varepsilon$, we see that all but at most $2\sqrt{\varepsilon}|A|$ vertices from A have degree at least $(1-2\sqrt{\varepsilon})k$ to B, and all but at most $3\sqrt{\varepsilon}|B|$ vertices from B have degree at least $(1-3\sqrt{\varepsilon})k$ to A. Let A_0 and B_0 be the set of vertices of low degree in A and B respectively, and let H be the bipartite graph induced by $A' = A \setminus A_0$ and $B' = B \setminus B_0$. Then the minimum degree of H is at least $(1-5\sqrt{\varepsilon})k$. Now, given a tree $T \in \mathcal{T}(k, \Delta)$, if $V(T) = C \cup D$ is its natural bipartition, Fact 2.2.2 implies that

$$\max\{|C|, |D|\} \leqslant \left(1 - \frac{1}{\Delta}\right)k \leqslant (1-5\sqrt{\varepsilon})k,$$

and therefore, by Lemma 2.2.4, we can embed T in H. $\square$

Now we turn to the non-bipartite case. In this case we can embed trees with maximum degree at most $\varepsilon\sqrt{k}$, for some small constant ε. As a first step, we will embed a small but linear size subtree $T^* \subseteq T$ trying to fill up as many low degree vertices of G as possible. We can then use the following result to embed the leftover vertices from $T - T^*$.

Lemma 4.2.2 (Lemma 4.4 from [61]). *Let $0 < \nu < \frac{1}{200}$, let $k \in \mathbb{N}$ and let H be a $k+1$-vertex graph with $\delta(H) \geqslant (1-2\nu)k$, and let $v \in V(H)$ be a vertex of degree k. If (T, r) is a rooted tree with at most k edges such that every vertex is adjacent to at most $\nu k/2$ leaves, then T can be embedded in H and any vertex in $H - v$ can be chosen as the image of r.*

Proposition 4.2.3. *Let $k \geqslant 10^6$ and let G be a graph on $n \leqslant (1+10^{-11})k$ vertices such that $d(G) > k-1$ and $\delta(G) \geqslant \frac{k}{2}$. Then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \frac{\sqrt{k}}{1000}$.*

Proof. Given G and k, set $\varepsilon := 10^{-11}$ and note that necessarily, $n > k$. Moreover, for the complement $\bar{G}$ of G, we have that $d(\bar{G}) < n - k$. Thus,

$$2e(\bar{G}) < n(n-k) \leqslant (1+\varepsilon)k \cdot \varepsilon k \leqslant 2\varepsilon k^2. \tag{4.4}$$

Let X be the set of all vertices of G having degree at most $\lfloor (1-\sqrt{\varepsilon})k \rfloor$ in G, and let Y be the set of all vertices of G having degree at least k in G. Since $\deg(v) \leqslant k-1$ for all $v \notin Y$, we have that

$$\sum_{v \in V(G) \setminus (X \cup Y)} \deg(v) \leqslant (k-1)|V(G) \setminus (X \cup Y)|$$

and thus, since $d(G) > k-1$ and hence $\sum_{v \in V(G)} \deg(v) > (k-1)|V(G)|$, we obtain

$$(k-1)|X \cup Y| \;<\; \sum_{v \in X \cup Y} \deg(v) \;\leqslant\; |X|(1-\sqrt{\varepsilon})k + |Y|(1+\varepsilon)k.$$

Therefore,

$$|X| < 2\sqrt{\varepsilon}|Y| < 3\sqrt{\varepsilon}k. \tag{4.5}$$

For each $v \in Y$ set $X_v := N(v) \cap X$. Let $v^\star \in Y$ be a vertex that minimises $|X_v|$ among all $v \in Y$. So,

$$\text{for each } v \in Y, \quad \deg(v, X) \geqslant |X_{v^\star}|. \tag{4.6}$$

Let $T \in \mathcal{T}(k, \frac{\sqrt{k}}{1000})$. Now if $X_{v^\star} = \emptyset$, then the graph induced by $v^\star$ and a k-subset of $N(v^\star)$ satisfies the conditions of Lemma 4.2.2, with $\nu := \sqrt{\varepsilon}$, and thus we can embed T. So, we will from now on assume that $X_{v^\star} \neq \emptyset$.

We use Lemma 2.2.6, with $\gamma := 168\sqrt{\varepsilon}$, to obtain a subtree (T^*, t^*) such that

$$84\sqrt{\varepsilon}k \leqslant |T^*| \leqslant 168\sqrt{\varepsilon}k \tag{4.7}$$

and such that every component of $T - T^*$ is adjacent to t^*. We will now embed T^* in a way that at least $|X_{v^\star}|$ vertices from X will be used. Then, we embed the rest of T into $G - X$ with the help of Lemma 2.2.1. Before we start, we quickly prove two claims that will be helpful for the embedding of T^*.

First, using (4.5) and the fact that $\delta(G) \geqslant \frac{k}{2}$, the following claim is easy to see.

Claim 4.2.4. *For every $x, x' \in V(G)$, there are more than $2^{-4}k$ internally disjoint paths of length at most 3 connecting x and x'.*

Second, we will see now that a useful subset of Y can be 'reserved' for later use.

Claim 4.2.5. *There is a subset $Y' \subseteq Y \setminus \{v^\star\}$ of size at most $\lfloor 5\sqrt{\varepsilon}k \rfloor$ such that all but at most $\lfloor 2\varepsilon k \rfloor$ vertices in $G - X$ have at least $|X|$ neighbours in Y'.*

To see this, suppose first that $|Y| \geqslant \lfloor 5\sqrt{\varepsilon}k \rfloor + 1$ and take any subset $Y' \subseteq Y \setminus \{v^\star\}$ of size $\lfloor 5\sqrt{\varepsilon}k \rfloor$. Since every vertex v in $G-X$ has degree at least $\lceil (1-\sqrt{\varepsilon})k \rceil$ and since $n \leqslant (1+\varepsilon)k$, we know that v has at least $\lceil 3\sqrt{\varepsilon}k \rceil \geqslant |X|$ neighbours in Y', and we are done.

Assume now that $|Y| \leqslant \lfloor 5\sqrt{\varepsilon}k \rfloor$ and let us write Z for the set of vertices in $G - X$ having less than $|X|$ neighbours in $Y \setminus \{v^\star\}$. Then one has the estimates

$$e(Y \setminus \{v^\star\}, G) = \sum_{y \in Y \setminus \{v^\star\}} \deg(y) \geqslant (|Y| - 1)k,$$

and

$$e(Y \setminus \{v^\star\}, G) = \sum_{z \in Z} \deg(z, Y \setminus \{v^\star\}) + \sum_{z \notin Z} \deg(z, Y \setminus \{v^\star\}) \leqslant |Z||X| + (n - |Z|)(|Y| - 1).$$

Therefore, as $|X| < 2\sqrt{\varepsilon}|Y|$ by (4.5), and since by assumption $n \leqslant (1+\varepsilon)k$, we have $|Z| < 2\varepsilon k$ and we can take $Y' = Y \setminus \{v^\star\}$. This finishes the proof of Claim 4.2.5.

By applying Lemma 2.2.1, with $\ell = 3$, we deduce that T^* has either $|T^*|/12$ bare paths, each of length 3, or it has at least $|T^*|/12$ leaves. The embedding of T^* splits into two cases depending on the structure of T^*.

Case 1: T^* has a set $\mathcal{B}$ of $|T^*|/12$ vertex disjoint bare paths, each of length 3.

We embed T^* vertex by vertex in a pseudo-greedy fashion always avoiding $v^\star$. We start by embedding t^* arbitrarily into any vertex of degree at least $(1 - \sqrt{\varepsilon})k$ of $G - v^\star$. Now suppose we are about to embed a vertex u' whose parent u has already been embedded into a vertex $\phi(u)$. If u' is not the starting point of a path from $\mathcal{B}$ or if all of $X_{v^\star}$ is already used, we embed u' greedily. Now assume that u' is the starting point of some $B \in \mathcal{B}$ and there is at least one unused vertex $x \in X_{v^\star}$. By Claim 4.2.4 and since $|T^*| < 2^{-4}k$, vertices x and $\phi(u)$ are connected by a path P of length at most 3 that uses only unoccupied vertices. Embed B (including u) into P, and if $|B| > |P|$, choose its last vertices greedily. Since by (4.5) and (4.7),

$$|X| \leqslant 3\sqrt{\varepsilon}k < \frac{|T^*|}{12} = |\mathcal{B}|,$$

we know that after embedding T^* every vertex in $X_{v^\star}$ is used.

Case 2: T^* has at least $|T^*|/12$ leaves.

In this case, we cannot ensure that every vertex in $X_{v^\star}$ is used for the embedding of T^*, however, we can still guarantee that at least $|X_{v^\star}|$ vertices from X are used.

Because of our bound on the maximum degree of T, we can find a set $U^* \subseteq V(T^*) \setminus \{t^*\}$ of parents of leaves such that the number of leaves pending from U^* is at least $6\sqrt{\varepsilon}k$, which by (4.5) is greater than $2|X|$. We then take an independent set $U \subseteq U^*$ such that for the set L of leaves pending from U we have $|L| \geqslant |X|$, and such that $|U| \leqslant |X|$.

Starting from t^* we embed T^*, following its natural order but leaving out the vertices from L. All vertices are embedded greedily into $G - Y'$, except vertices from U and their parents which are embedded in a different way. Assume $v \in V(T^*)$ is a parent of some vertex in U. Since T^* is small, because of (4.5), because of our assumption on the minimum degree of G, and because of Claim 4.2.5, we may embed v into a vertex having at least $|X|$ neighbours in Y'. After this, we embed the children of v in U into unoccupied vertices of Y'. Other children of v are embedded greedily. At the end of this process we have embedded all of $T^* - L$. If we have used at least $|X_{v^\star}|$ vertices from X, we complete the embedding of T^* greedily, so let us assume we have used less than $|X_{v^\star}|$ vertices from X. We embed the leaves pending from U one by one into vertices from X until we use $|X_{v^\star}|$ vertices, which is possible since U was embedded into Y' and because of (4.6). After this point, we simply embed the leftover leaves of T^* greedily but always avoiding $v^\star$.

This finishes the case distinction. Set $T' := T - (T^* - t^*)$. Denoting by ϕ the embedding we note that

$$|N(v^\star) \setminus (\phi(T^*) \cup X_{v^\star})| \geqslant k - |\phi(T^*)| - |X_{v^\star}| + |\phi(T^*) \cap X_{v^\star}| + |\phi(T^*) \setminus N(v^\star)| \geqslant |T'| - 2.$$

Therefore, the graph H induced by $v^\star$, $\phi(t^*)$ and any $(|T'|-2)-$subset of $|N(v^\star) \setminus (\phi(T^*) \cup X_{v^\star})|$ has order $|T'|$ and we may complete the embedding of (T', t^*) by using Lemma 4.2.2 for H, with $\nu := 86\sqrt{\varepsilon}$, fixing the image of t^* as $\phi(t^*)$. $\square$

4.3 Using the regularity method

In this section, we will use the embedding tools from Section 3.3 to show that average degree slightly below k is enough to ensure the containment of every k-edge tree with bounded degree, provided the host graph has a regular partition and is considerable larger than the tree.

Lemma 4.3.1. *For all $\Delta \geqslant 2$, $M_0 \in \mathbb{N}$, $\delta, \varepsilon, \eta \in (0,1)$ with $\varepsilon \ll \eta \leqslant \frac{\delta^2}{10^4}$ there is $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0$, $n \in \mathbb{N}$ with $\delta^{-1}k \geqslant n \geqslant k$ the following holds.*
Let G be an n-vertex graph with an (ε, η)-regular partition and corresponding reduced graph $\mathcal{R}$, with $|\mathcal{R}| \leqslant M_0$, which is connected and bipartite with parts $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| \geqslant |\mathcal{B}|$. If

(i) $d(G) \geqslant (1 - 3\sqrt{\eta})k$;

(ii) $\delta(G) \geqslant (1 - 3\sqrt{\eta})\frac{k}{2}$; *and*

(iii) $|\bigcup \mathcal{A}| \geqslant (1 + \delta)k$,

then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.

Proof. Given Δ, M_0, ε and η, we choose k_0 as the output of Proposition 3.3.1. Given G as in Lemma 4.3.1, we suppose for contradiction that G contains no copy of some tree T with k edges such that $\Delta(T) \leqslant \Delta$. Set

$$t = \frac{|\mathcal{R}|}{n}$$

and let $|\bigcup \mathcal{A}| = a$ and $|\bigcup \mathcal{B}| = b$. We claim that

$$b \geqslant \left(1 + \frac{\delta}{4}\right)\frac{k}{2}. \tag{4.8}$$

Indeed, otherwise can use (i) to calculate that

$$(1 - 3\sqrt{\eta})kn \leqslant 2e(G) \leqslant 2ab \leqslant \left(1 + \frac{\delta}{4}\right)ka \leqslant \left(1 + \frac{\delta}{4}\right)k \cdot \left(1 - \frac{\delta}{4}\right)n \leqslant \left(1 - \frac{\delta^2}{16}\right)kn$$

where the second to last inequality follows from the fact that because of (ii) we have $a = n - b \leqslant n - (1 - 3\sqrt{\eta})\frac{k}{2} \leqslant (1 - \frac{\delta}{4})n$. But this is a contradiction to our assumptions on η and δ. This proves (4.8), and so, we also know that

$$|\mathcal{A}| \geqslant |\mathcal{B}| \geqslant \left(1 + \frac{\delta}{4}\right)\frac{k}{2}t. \tag{4.9}$$

Now we turn to the tree T. Let A and B denote its colour classes, and assume $|A| \geqslant |B|$. Moreover, we may assume that

$$(1 - 4\sqrt{\eta})\frac{k}{2} < |B| \leqslant \frac{k+1}{2} \quad \text{and} \quad \frac{k+1}{2} \leqslant |A| \leqslant (1 + 4\sqrt{\eta})\frac{k}{2}. \tag{4.10}$$

as otherwise, since $\varepsilon \ll \eta$ we have $\delta(G) \geqslant (1 + 100\sqrt{\varepsilon})|B|$ and so, by (iii), we can use Proposition 3.3.1 to embed T.

Let $\mathcal{V}_A \subseteq \mathcal{A}$ and $\mathcal{V}_B \subseteq \mathcal{B}$ be the sets of all clusters of degree at least $(1 + \sqrt{\eta})\frac{k}{2}t$. We claim that

$$|\mathcal{V}_A| + |\mathcal{V}_B| \geqslant (1 + \sqrt{\eta})kt. \tag{4.11}$$

Suppose this is not the case. Then Fact 2.3.3 (i), condition (i), and (4.9) imply that

$$\begin{aligned}
(1 - 3\sqrt{\eta})kt|\mathcal{R}| &\leqslant 2e(\mathcal{R}) \\
&\leqslant |\mathcal{V}_A||\mathcal{B}| + |\mathcal{V}_B||\mathcal{A}| + (1 + \sqrt{\eta})\frac{k}{2}t\Big(|\mathcal{R}| - |\mathcal{V}_A| - |\mathcal{V}_B|\Big) \\
&= (1 + \sqrt{\eta})\frac{k}{2}t|\mathcal{R}| + |\mathcal{V}_A|\Big(|\mathcal{B}| - (1 + \sqrt{\eta})\frac{k}{2}t\Big) + |\mathcal{V}_B|\Big(|\mathcal{A}| - (1 + \sqrt{\eta})\frac{k}{2}t\Big) \\
&< (1 + \sqrt{\eta})\frac{k}{2}t|\mathcal{R}| + (1 + \sqrt{\eta})kt \cdot \left(\frac{\delta}{8} - \sqrt{\eta}\right)\frac{k}{2}t.
\end{aligned}$$

Therefore, and since $n \geqslant k$, we have

$$\frac{1}{2}t \cdot k \leqslant (1 - 7\sqrt{\eta})tn = (1 - 7\sqrt{\eta})|\mathcal{R}| < (1 + \sqrt{\eta})\left(\frac{\delta}{4} - \sqrt{\eta}\right)kt \leqslant \frac{3}{2} \cdot \frac{\delta}{4}kt,$$

a contradiction. So, assuming that $|\mathcal{V}_A| \geqslant (1+\sqrt{\eta})\frac{k}{2}t$, by Proposition 3.3.4 (see Remark 3.3.2) we can embed T into G, with A going to clusters in $\mathcal{V}_A$ and B going to clusters in $\mathcal{B}$. $\square$

Now we turn to the case when the reduced graph is connected, non-bipartite and large. We first derive some useful information about the structure of the reduced graph of G if it is connected and non-bipartite, and G contains no copy of some k-edge tree with bounded degree.

Lemma 4.3.2. *For all $\Delta \geqslant 2$, $M_0 \in \mathbb{N}$, $\delta, \varepsilon, \eta \in (0,1)$ with $\varepsilon \ll \eta \leqslant \frac{\delta^4}{10^8}$, there is $k_0 \in \mathbb{N}$ such that for all $k, n \geqslant k_0$ with $\delta^{-1}k \geqslant n \geqslant (1+\delta)k$ the following holds. Let G be an n-vertex graph that admits an (ε, η)-regular partition into M_0 parts, and assume the corresponding (ε, η)-reduced graph is connected and non-bipartite. If furthermore,*

(i) $d(G) \geqslant (1 - 3\sqrt{\eta})k$; and

(ii) $\delta(G) \geqslant (1 - 3\sqrt{\eta})\frac{k}{2}$,

and G contains no copy of some tree T with k edges such that $\Delta(T) \leqslant \Delta$, then G has a subgraph $G' \subseteq G$ of size $|G'| \geqslant |G| - M_0$ such that there is a partition $V(G') = I \cup V_1 \cup V_2$ with

(a) $|V_i| = (1 \pm 3\sqrt{\eta})\frac{k}{2}$ for $i \in \{1, 2\}$;

(b) I is an independent set in G' and there are no edges between I and V_2 in G';

(c) $\deg_{G'}(x) \geqslant (1 - 5\sqrt[4]{\eta})n$ for at least $(1 - 4\sqrt[4]{\eta})|V_1|$ vertices $x \in V_1$;

(d) $\deg_{G'}(y) \geqslant (1 - 3\sqrt[8]{\eta})k$ for at least $(1 - 2\sqrt[8]{\eta})|V_2|$ vertices $y \in V_2$.

Proof. Let $k_0 \geqslant \frac{M_0}{\varepsilon}$ be at least as large as the output of Proposition 3.3.9 for $\frac{\varepsilon}{5}$ and Δ. Applying Lemma 3.3.8 to G, with $\ell = M_0$ and $t = (1 - 3\sqrt{\eta})\frac{k}{2}$, we find a subgraph G' of size $|G'| \geqslant n - M_0$ that admits an $(5\varepsilon, \frac{\eta}{2})$-regular partition. Moreover, the corresponding reduced graph $\mathcal{R}$ contains a matching $\mathcal{M}$ and a disjoint independent set $\mathcal{I}$ such that $V(\mathcal{R}) = \mathcal{I} \cup V(\mathcal{M}) = \mathcal{I} \cup \mathcal{V}_1 \cup \mathcal{V}_2$ and $N_{\mathcal{R}}(\mathcal{I}) \subseteq \mathcal{V}_1$.

Letting $I = \bigcup \mathcal{I}$ and $V_i = \bigcup \mathcal{V}_i$ for $i \in \{1, 2\}$ we have (b). Furthermore, because of Proposition 3.3.9 we know that $|\mathcal{V}_i| \leqslant (1+\eta)\frac{k}{2}\frac{|\mathcal{R}|}{n}$ and thus $|V_i| \leqslant (1+\eta)\frac{k}{2}$ for $i \in \{1, 2\}$. Therefore, and because of condition (ii) we have (a).

In order to see (c) and (d), we do the following. For any subset $A \subseteq V(G')$ let d_A denote the average degree in G' of the vertices in A. By (b), we have $d_I \leqslant |V_1| \leqslant (1+\eta)\frac{k}{2}$. By condition (i) and since $\deg_{G'}(x) \geqslant \deg_G(x) - M_0$ for every $x \in V(G')$, we have

$$\begin{aligned}(1 - 4\sqrt{\eta})kn \;\leqslant\; 2e(G') &= |I|d_I + |V_1|d_{V_1} + |V_2|d_{V_2} \\ &\leqslant (1+\eta)\tfrac{k}{2}(|I| + d_{V_1} + d_{V_2}) \\ &\leqslant (1+\eta)\tfrac{k}{2}(n - (|V_1| + |V_2|) + d_{V_1} + d_{V_2}) \\ &\leqslant (1+\eta)\tfrac{k}{2}(n - (1 - 3\sqrt{\eta})k + d_{V_1} + d_{V_2}),\end{aligned}$$

and therefore,

$$d_{V_1} + d_{V_2} \geqslant (1 - 8\sqrt{\eta})n + (1 - 3\sqrt{\eta})k. \tag{4.12}$$

Because of (b), we have $d_{V_2} \leqslant |V_1| + |V_2| \leqslant (1+\eta)k$. Thus (4.12) implies that $d_{V_1} \geqslant (1-12\sqrt{\eta})n$. Since $d_{V_1} \leqslant n$ and since $n \leqslant \delta^{-1}k$, inequality (4.12) also implies that $d_{V_2} \geqslant (1-\sqrt[4]{\eta})k$. Apply Lemma 4.1.3 to $f_1(v) = \deg_{G'}(v)$ for $v \in V_1$, with parameters $t_1 = (1-12\sqrt{\eta})n$, and $\varepsilon_1 = 16\sqrt{\eta}$, and to $f_2(v) = \deg_{G'}(v)$ for $v \in V_2$, with $t_2 = (1-\sqrt[4]{\eta})k$ and $\varepsilon_2 = 4\sqrt[4]{\eta}$, to obtain (c) and (d). □

The next lemma finishes the analysis of the non-bipartite case.

Lemma 4.3.3. *For all $\Delta \geqslant 2, M_0 \in \mathbb{N}$, $\delta, \varepsilon, \eta \in (0,1)$ with $\varepsilon \ll \eta \leqslant \frac{\delta^8}{10^{80}}$, there is $k_0 \in \mathbb{N}$ such that for all $k, n \geqslant k_0$ with $\delta^{-1}k \geqslant n \geqslant (1+\delta)k$ the following holds. Let G be an n-vertex graph that admits an (ε,η)-regular partition into at most M_0 parts and assume the corresponding reduced graph is connected and non-bipartite. If*

(i) $d(G) \geqslant (1-3\sqrt{\eta})k$; and

(ii) $\delta(G) \geqslant (1-3\sqrt{\eta})\frac{k}{2}$,

then G contains a copy of every tree T with k edges such that $\Delta(T) \leqslant \Delta$.

Proof. Let k_0 be the output of Lemma 4.3.2 and let G be given. Let T be a tree with k edges and $\Delta(T) \leqslant \Delta$ and suppose we cannot embed T into G. Then by Lemma 4.3.2 we may find a subgraph $G' \subseteq G$ and a partition $V(G') = I \cup V_1 \cup V_2$ fulfilling the properties of Lemma 4.3.2.

Let $U_1 \subseteq V_1$ be the set of all vertices $x \in V_1$ with $\deg_{G'}(x) \geqslant (1-5\sqrt[4]{\eta})n$, and let $U_2 \subseteq V_2$ be the set of all vertices $x \in V_2$ with $\deg_{G'}(x) \geqslant (1-3\sqrt[8]{\eta})k$. In particular, because of Lemma 4.3.2 (a), we have that

$$\text{each vertex } x \in U_1 \text{ has at least } (1-\sqrt{\eta})|I| \text{ neighbours in } I. \tag{4.13}$$

Also, note that $|U_1| \geqslant (1-4\sqrt[4]{\eta})|V_1| \geqslant |V_1| - \sqrt[8]{\eta}k$ and $|U_2| \geqslant (1-2\sqrt[8]{\eta})|V_2|$, by Lemma 4.3.2 (a), (c) and (d). Let H be the graph induced by U_1 and U_2. Note that because of Lemma 4.3.2 (b) and (d), we know that the vertices from U_2 have minimum degree at least $(1-6\sqrt[8]{\eta})k$ in H, and because of Lemma 4.3.2 (a) and (d), the vertices from U_1 have minimum degree at least $(1-9\sqrt[4]{\eta})n - 2\sqrt[8]{\eta}k - |I| \geqslant (1-3\sqrt[8]{\eta})k$ in H. Hence,

$$\delta(H) \geqslant (1-6\sqrt[8]{\eta})k. \tag{4.14}$$

So, by Lemma 2.2.3 every tree with at most $(1-6\sqrt[8]{\eta})k$ edges can be embedded greedily into H. Let (T^*, t^*) be the subtree given by Lemma 2.2.6 for $\gamma = \frac{1}{2}$, so that $\frac{k}{4} \leqslant |T^\star| \leqslant \frac{k}{2}$ and every component of $T - T^*$ is adjacent to t^*. We apply Lemma 2.2.1 to T^*, with $\ell = 3$, which splits the proofs into two cases.

Case 1: $T^\star$ has a set $\mathcal{B}$ of $|T^\star|/12$ vertex disjoint bare paths, each of length 3.

Note that each vertex from H has at least Δ neighbours in U_1, because of Lemma 4.3.2 (a) and our bound from (4.14), which will be tacitly used in what follows.

We embed $t^\star$ into any vertex from H. The rest of $T^\star$ will be embedded in DFS order into H. We will use the following strategy until we have occupied $\lceil\frac{\delta}{100}k\rceil$ vertices from I. For each path $P \in \mathcal{B}$, we proceed as follows. We embed the first vertex v_1 of the path P into a vertex $u_1 \in U_1$, and then find another vertex $u_3 \in U_1$ which has a common neighbour u_2 with u_1 in I. Note that the vertex u_3 exists because of (4.13). We then embed the middle vertex v_2 of P into $u_2 \in I$, and the end point v_3 into $u_3 \in U_1$. The remaining vertices of $T^\star$ are embedded greedily into H.

Case 2: $T^\star$ has $|T^\star|/12$ leaves.

In this case, the embedding of $T^\star$ follows a similar strategy. We embed $t^\star$ into any vertex from H and the rest will be embedded in DFS order. We take care to embed all parents of leaves into U_1 and all leaves into I, until we have used $\lceil\frac{\delta}{100}k\rceil$ vertices from I. The remaining vertices of $T^\star$ are embedded greedily into H.

Now, let m be the number of vertices we have embedded so far into H, and let $H' \subseteq H$ contain all unused vertices of H. By our embedding strategy, we have that $m \leqslant |T^\star| - \frac{\delta}{100}k$. Therefore, and by (4.14),

$$\delta(H') \geqslant (1 - 6\sqrt[8]{\eta})k - m \geqslant (1 - 6\sqrt[8]{\eta})k + \tfrac{\delta}{100}k - |T^\star| \geqslant (1 + \tfrac{\delta}{200})k - |T^\star|,$$

and so we can finish the embedding of T by embedding $T - T^*$ greedily into H'. □

4.4 Proof of the Erdős–Sós conjecture for trees with bounded degree and dense host graph

In this section we finally prove Theorem 1.1.2 with the help of the results from the previous sections.

Proof of Theorem 1.1.2. Given Δ and δ, we set $\nu = \min\{\frac{\delta^2}{2^{10}}, \frac{1}{10^{11}}, \frac{1}{25\Delta^2}\}$ and we fix parameters ε, η, θ such that

$$0 < \varepsilon \ll \eta \ll \theta \leqslant \frac{\nu^8}{10^{80}}.$$

Let k_0 be the maximum of $\frac{3}{\varepsilon}$ and the outputs of Lemma 2.3.2, Lemma 4.3.1, Lemma 4.3.3 and Lemma 4.1.1 (with ν playing the role of δ, and $m_0 = \lceil\frac{1}{\varepsilon}\rceil$). Set $n_0 = \lceil\delta^{-1}k_0\rceil$.

By Proposition 4.2.3 we may assume that $|G| \geqslant (1+\nu)k$ and if G is ν-bipartite, Proposition 4.2.1 allows us to assume that the larger bipartition class of G has at least $(1+\nu)k$ vertices. Now the regularity lemma (Lemma 2.3.2) provides us with a subgraph G' with $|G'| \geqslant (1-\varepsilon)n$ that has an (ε, η)-regular partition. Let $\mathcal{R}$ be the corresponding reduced graph and let $\mathcal{U}_1, \dots, \mathcal{U}_\ell$ be the connected components of $\mathcal{R}$. Then, since we may assume that $\delta(G) \geqslant \frac{k}{2}$ (see the footnote in the Introduction), we have

$$\deg_{G'}(x) \geqslant (1 - 2\sqrt{\eta}) \deg_G(x) \geqslant (1 - 2\sqrt{\eta})\frac{k}{2} \quad \text{for all } x \in V(G'),$$

and therefore

$$\frac{\ell k}{4} \leqslant (1-2\sqrt{\eta})\frac{k}{2}\ell \leqslant \sum_{i\in[\ell]} |\bigcup \mathcal{U}_i| \leqslant n \leqslant \delta^{-1}k,$$

implying that

$$\ell \leqslant 4\delta^{-1}. \tag{4.15}$$

We set $U_i' = \bigcup \mathcal{U}_i$ for each $i \in [\ell]$.

Claim 4.4.1. *Suppose G' contains no copy of some k-edge tree T with $\Delta(T) \leqslant \Delta$, then*

(i) $d(G'[U_i']) = (1 \pm \frac{\nu}{2})k$ and $\delta(G'[U_i']) \geqslant (1-\frac{\nu}{2})\frac{k}{2}$ for all $i \in [\ell]$; and

(ii) for each $i \in [\ell]$ either

(a) $G'[U_i']$ is non-bipartite and $|U_i| = (1 \pm \frac{\nu}{2})k$, or

(b) $G'[U_i']$ is bipartite with $V(U_i) = A_i \cup B_i$ such that $|A_i|, |B_i| = (1 \pm \frac{\nu}{2})k$.

In order to see this claim, observe that since T cannot be embedded into G', Lemma 4.1.1 implies that $d(G'[U_i']) < (1+\theta)k$ for each $i \in [\ell]$. Note that

$$\sum_{i=1}^{\ell} \frac{|U_i'|}{n} d(G'[U_i']) = d(G') \geqslant (1-3\sqrt{\eta})k.$$

Set $t = (1-3\sqrt{\eta})k$. Applying Lemma 4.1.2 with $N = \ell$, $\mu(i) = |U_i'|/n$ and $f(i) = d(G'[U_i'])$, and with $\sqrt{\theta}$ in the role of ε, we see that the set $I = \{i \in [\ell] : d(G'[U_i']) < (1-2\sqrt{\theta})t\}$ satisfies

$$\frac{t|I|}{2n} \leqslant \mu(I) \leqslant 2\sqrt{\theta}$$

(where for the first inequality we use that $|U_i| > \frac{t}{2}$ for each i). Thus, $|I| \leqslant 8\delta^{-1}\sqrt{\theta} < 1$. In other words, $I = \emptyset$, and therefore, for each $i \in [\ell]$ we have

$$d(G'[U_i']) \geqslant (1-2\sqrt{\theta})(1-3\sqrt{\eta})k \geqslant (1-3\sqrt{\theta})k. \tag{4.16}$$

This, together with the minimum degree in G', proves (i). In order to see (ii), we use (4.16) and Lemmas 4.3.1 and 4.3.3. This proves Claim 4.4.1.

Now we distribute the vertices from $G - G'$ into the sets U_i'. We successively assign each leftover vertex to the set U_i' it sends most edges to (or to any one of these sets, if there is more than one). Then for each $i \in [\ell]$ and all $x \in U_i$ we have

$$\deg(x, U_i') \geqslant \frac{k}{2\ell} \geqslant \frac{\delta}{8}k,$$

where we used (4.15) for the second inequality. Since we add at most $\varepsilon n \ll \nu k$ vertices to each set, we end up with a partition $V(G) = U_1 \cup \ldots \cup U_\ell$ satisfying, for each $i \in [\ell]$,

(I) $d(G[U_i]) = (1 \pm \nu)k$ and $\delta(G[U_i]) \geqslant \frac{\delta}{8}k$;

(II) $\deg(x, U_i) < (1-\nu)\frac{k}{2}$ for less than νk vertices $x \in U_i$; and

(III) either $G[U_i]$ is non-bipartite and $|U_i| = (1 \pm \nu)k$, or $G[U_i]$ is ν-bipartite with $U_i = A_i \cup B_i$ such that $|A_i|, |B_i| = (1 \pm \nu)k$.

For each $i \in [\ell]$, we use Lemma 4.1.3 for $f(x) = \deg(x, U_i)$, with 2ν playing the role of ε, to deduce that

$$\deg(x, U_i) \geqslant (1 - \sqrt{2\nu})k \text{ for at least } (1 - \sqrt{2\nu})|U_i| \text{ vertices from } U_i. \tag{4.17}$$

Now we embed T using this structural information of G. We apply Lemma 2.2.6 to T, with $\gamma = \frac{1}{2}$, to obtain a subtree (T, t^*) with $\frac{k}{4} \leqslant |T^*| \leqslant \frac{k}{2}$ such that every component of $T - T^*$ is adjacent to t^*. Moreover, since $\Delta(T) \leqslant \Delta$ there is a component T' of $T - T^*$ with $\frac{k}{2\Delta} \leqslant |T'| \leqslant \frac{3k}{4}$.

Note that if there are no edges between different sets U_i, then an averaging argument shows that there is $i^\star \in [\ell]$ such that $d(G[U_{i^\star}]) \geqslant d(G) > k-1$. But then, because of (III) and because of Theorem 4.0.1, we are done. Thus, we may assume that there is an edge $u_i u_j$ with $u_i \in U_i$ and $u_j \in U_j$. We map t^* into u_i and map the root of T' into u_j. Note that by (I), we have

$$\delta(G[U_i]) \geqslant \frac{\delta}{8}k \geqslant 4\sqrt{\nu}k \geqslant \sqrt{2\nu}|U_i| + \Delta \tag{4.18}$$

and that (III), together with our choice of ν ensures that $\sqrt{2\nu}|U_i| \leqslant \frac{k}{2\Delta}$. So, we may finish the proof by using Lemma 2.2.3 and Lemma 2.2.4 to embed $T - T'$ into U_i and T' into U_j, which we can do because of (4.17) and (4.18).

4.5 Multicolour Ramsey number of bounded degree trees

To finish this chapter, let us briefly mention a consequence of the Erdős–Sós conjecture in Ramsey theory. Given an integer $\ell \geqslant 2$ and a graph H, the ℓ-colour Ramsey number $r_\ell(H)$ of H is the smallest $n \in \mathbb{N}$ such that every ℓ-colouring of the edges of K_n yields a monochromatic copy of H. In general, determining the Ramsey number of a graph is a very difficult problem and have received considerable attention over more than 70 years. We do not aim to describe Ramsey theory for graphs here, however, we recommend the survey of Conlon, Fox, and Sudakov [36] for recent developments.

Regarding Ramsey numbers of trees, Erdős and Graham conjectured [44] in 1973 that every tree T with k edges satisfies

$$r_\ell(T) = \ell k + O(1), \tag{4.19}$$

and they established the lower bound $r_\ell(T) > \ell(k-1)+1$ for large enough ℓ satisfying $\ell \equiv 1 \mod k$. Moreover, Erdős and Graham also observed that the upper bound in (4.19) would follow from the Erdős–Sós conjecture. Indeed, already for $n \geqslant \ell(k-1)+2$ note that the most popular colour in any ℓ-colouring of K_n has at least $\frac{1}{\ell}\binom{n}{2}$ edges and thus average degree at least $\frac{n-1}{\ell} > k-1$. So the Erdős–Sós conjecture would imply that the most popular colour contains a copy of every tree with k edges. Therefore, from Theorem 1.1.2 we deduce the following result.

Corollary 4.5.1. *Let $\ell, \Delta \geqslant 2$ be two integers. Then there exists $k_0 \in \mathbb{N}$ such that for every $k \geqslant k_0$ the following holds. For every tree T with k edges and $\Delta(T) \leqslant \Delta$ we have $r_\ell(T) \leqslant \ell(k-1)+2$.*

We remark that in Corollary 4.5.1 one can actually find a copy of every k-edge tree with bounded degree in the same colour, at the same time. Regarding the lower bound, we observe that the construction of Erdős and Graham works for fixed k and large ℓ depending on k, while Corollary 4.5.1 works for large k depending on ℓ. Therefore, a construction showing the lower bound is still missing.

□

Chapter 5

Global resilience of trees in sparse random graphs

Based on joint work with Pedro Araújo and Luiz Moreira [10].

The study of random analogues of classical results in extremal combinatorics has been an active area of research in the last decades with remarkable results (see [35] for a survey). One particular line of research is looking for the containment of large graphs in sparse random graphs. For instance, studying the threshold for the containment of a perfect matching, a Hamilton cycle, or a spanning tree have been very popular problems in the last 20 years. Moreover, it has been also studied how *resilient* is $G(n,p)$ with respect to some property that it typically possesses. For instance, a prototype problem is to determine how many edges can be removed from $G(n,p)$ so that the resulting subgraph still contains a Hamilton cycle. We recommend the survey of Böttcher [27] for a general overview of this area. In this chapter, we will make some progress towards this program showing a sparse random analogue of the Erdős–Sós conjecture (Theorem 1.4.3). Actually, we will prove an even stronger statement by replacing the random graph $G(n,p)$ with a *random-like* graph. To make this statement precise, let us give some definitions.

Definition 5.0.1 (Uniform graph)**.** *Let $\eta, p \in (0,1)$. We say that a graph G on n vertices is (η,p)-uniform, if for every pair of disjoint sets $A, B \subseteq V(G)$ such that $|A|, |B| \geqslant \eta n$ we have*

$$(1-\eta)p|A||B| \leqslant e_G(A,B) \leqslant (1+\eta)p|A||B| \tag{5.1}$$

and

$$(1-\eta)p\binom{|A|}{2} \leqslant e_G(A) \leqslant (1+\eta)p\binom{|A|}{2}. \tag{5.2}$$

Furthermore, we say that G is (η,p)-upper-uniform if (possibly) only the upper bounds in (5.1) *and* (5.2) *hold for all $A, B \subseteq V(G)$ as above.*

It is not hard to prove that, with high probability, the random graph $G(n,p)$ is (η,p)-uniform provided that pn is large enough (see Lemma 5.1.4). The main result of this chapter states that one can replace $G(n,p)$ in Theorem 1.4.3 by a (η,p)-uniform graph. Namely, we will prove the following result.

Theorem 5.0.2. *Let $\delta, \varrho \in (0,1)$ and $\Delta \geqslant 2$. There are positive constants n_0, η_0 and C such that for all $0 < \eta \leqslant \eta_0$ and $n \geqslant n_0$ the following holds. Let $p \in [0,1]$ with $p \geqslant \frac{C}{n}$ and let G be a (η, p)-uniform graph on n vertices. If $G' \subseteq G$ is a subgraph such that $e(G') \geqslant (\varrho + \delta)\, e(G)$, then G' contains a copy of every tree T with ϱn edges such that $\Delta(T) \leqslant \Delta$.*

The proof of Theorem 5.0.2 is based on the *sparse regularity method* combined with tree embedding results in bipartite *expander graphs*. Let us give a rough outline of the proof here.

Let G be an (η, p)-uniform graph and let $G' \subseteq G$ be a subgraph of G such that $e(G') \geqslant (\varrho + \delta)e(G)$. We may apply the *sparse regularity lemma* to G' to obtain a regular partition of $V(G')$. We will work on the reduced graph $\mathcal{R}'$ of G' in order to find a "good" structure. Let k be the number of vertices of $\mathcal{R}'$. As in the standard regularity lemma, one can prove that $\mathcal{R}'$ inherits the edge density of G', but scaled by p, so that the average degree of $\mathcal{R}'$ satisfies $d(\mathcal{R}') \geqslant (\varrho + \frac{\delta}{3})k$. Using the large average degree we find a large *matching structure* (see Section 5.4) which will allow us to embed any given bounded degree tree. The matching structure consists of cluster X, a matching $\mathcal{M}$, and a bipartite graph $\mathcal{H} = (\mathcal{Y}, \mathcal{Z})$, such that $N(X) = V(\mathcal{M}) \cup \mathcal{Y}$ and $\mathcal{Y}$ has large minimum degree in $\mathcal{H}$ (see Figure 5.1).

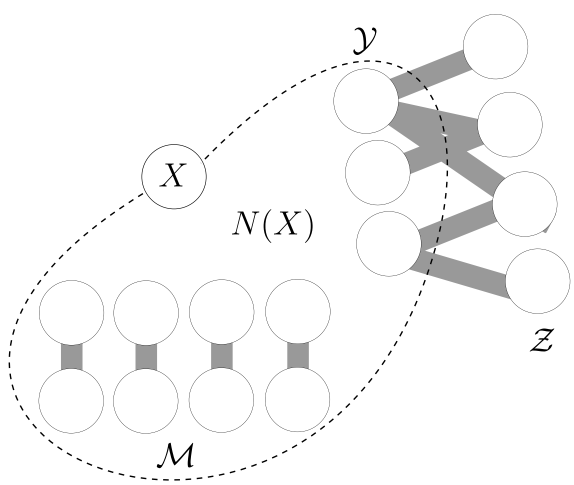

Figure 5.1: Matching structure.

Let T be a tree with ϱn edges such that $\Delta(T) \leqslant \Delta$. Our goal is to embed T using the matching structure. To do so, we first cut the tree into very small subtrees and then locate every such subtree into some edge of the reduced graph. If $\mathcal{M}$ is large enough, then we will locate each subtree into an edge of the matching, using both clusters of the edge in a balanced way. Otherwise, we will first locate subtrees into edges from $\mathcal{H}$, until a large proportion of $\mathcal{Y} \cup \mathcal{Z}$ is used. The leftover subtrees can be located into $\mathcal{M}$, always using both clusters from each edge in a balanced way. In any case, once we have located the subtrees, we will use an embedding technique due to Balogh, Csaba and Samotij [13] in order to embed each of this subtrees into the regular pair that was assigned to this subtree. The role of X here is to connect the embedding, meaning that X will be used in order to go from one edge to another in $\mathcal{M} \cup \mathcal{H}$.

This chapter is organized as follows. In Section 5.1 we introduce the regularity lemma for sparse graphs and in Section 5.2 we present the embedding techniques that will be used to embed tees into regular pairs. In Section 5.3 we state a result that allow us to partition a bounded degree tree into smaller subtrees in such a way that each subtree is connected to few other subtrees. In Section 5.4 we find the matching structure and in Section 5.5 we put everything together in order to prove Theorem 5.0.2. Finally, in Section 5.6 we discuss some applications of Theorem 1.4.3 in Ramsey theory.

5.1 Szemerédi's regularity lemma for sparse graphs

It is well known that the Szemerédi's regulariyt lemma does not work in sparse graphs, and the reason is that the typical cleaning procedure might delete all the edges if the graph is not dense enough. In this section, we introduce a sparse variant of the regularity lemma which works for graphs with even a linear number of edges. Let us start with some definitions.

Let G be a graph and let $p \in (0,1)$. Given two disjoint sets $A, B \subseteq V(G)$, we define the p-density of the pair (A,B) by

$$d_p(A,B) = \frac{e(A,B)}{p|A||B|}.$$

Given $\varepsilon > 0$, we say that the pair (A,B) is (ε,p)-regular if for all $A' \subseteq A$ and $B' \subseteq B$, with $|A'| \geqslant \varepsilon|A|$ and $|B'| \geqslant \varepsilon|B|$, we have

$$|d_p(A',B') - d_p(A,B)| \leqslant \varepsilon.$$

Now we state some standard results regarding properties of regular pairs (we refer to the survey [52] for the proofs).

Lemma 5.1.1. *Given $\alpha > \varepsilon > 0$, let G be a graph and let $A, B \subseteq V(G)$ be disjoint sets such that (A,B) is (ε,p)-regular with $d_p(A,B) = d > 0$. Then the following properties hold.*

1. *Let $A' \subseteq A$ with $|A'| \geqslant \alpha|A|$ and $B' \subseteq B$ with $|B'| \geqslant \alpha|B|$. Then the pair (A',B') is $(\varepsilon/\alpha, p)$-regular with p-density at least $d - \varepsilon$.*

2. *There are at most $\varepsilon|A|$ vertices in A with less than $(d-\varepsilon)p|B|$ neighbours in B.*

A partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$ is said to be (ε,p)-regular if

(a) $|V_0| \leqslant \varepsilon|V(G)|$,

(b) $|V_i| = |V_j|$ for all $i,j \in [k]$, and

(c) all but at most εk^2 pairs (V_i,V_j) are (ε,p)-regular.

We may now state a sparse version of Szemerédi's regularity lemma, due to Kohayakawa and Rödl [70, 71] .

Theorem 5.1.2 (Sparse Regularity Lemma). *Given $\varepsilon > 0$ and $k_0 \in \mathbb{N}$, there are $\eta > 0$ and $K_0 \geqslant k_0$ such that the following holds. Let G be an η-upper-uniform graph on $n \geqslant k_0$ vertices and let $p \in (0,1)$, then G admits an (ε, p)-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$ with $k_0 \leqslant k \leqslant K_0$.*

Let G be a graph that admits an (ε, p)-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$. Let $d \in (0,1)$. The (ε, p, d)-reduced graph $\mathcal{R}$, with respect to this (ε, p)-regular partition of G, is the graph with vertex set $V(R) = \{V_i : i \in [k]\}$, called clusters, such that V_iV_j is an edge if and only if (V_i, V_j) is an (ε, p)-regular pair with $d_p(V_i, V_j) \geqslant d$. The following proposition establishes that the edge density of R is roughly the same as in G, but scaled by p.

Proposition 5.1.3. *Let $\varepsilon, \eta, p, d \in (0,1)$ and let $k \in \mathbb{N}$ such that $k \geqslant 1/\varepsilon$. Let G be an (η, p)-upper uniform graph on n vertices that admits an (ε, p)-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$, and let $\mathcal{R}$ be the (ε, p, d)-reduced graph of G with respect to this partition. Then*

$$e(\mathcal{R}) \geqslant \frac{e(G)}{(1+\eta)p}\left(\frac{k}{n}\right)^2 - (6\varepsilon + d)k^2.$$

Proof. Let $G^{\mathcal{R}}$ be the subgraph of G obtained by deleting all the edges within clusters, between irregular pairs, and between regular pairs with density less than d. Since $k \geqslant 1/\varepsilon$, for $i \in [k]$ we have

$$|V_i| \leqslant \frac{n}{k} \leqslant \varepsilon n.$$

Let us choose η so that $|V_i| \geqslant n/(k+1) \geqslant \eta n$ for all $i \in [k]$. The (η, p)-upper uniformity of G implies that $G^{\mathcal{R}}$ misses at most

- $k \cdot (1+\eta)p|V_0|\frac{n}{k} \leqslant 2\varepsilon p n^2$ between V_0 and $V(G) \setminus V_0$;
- $(k+1) \cdot (1+\eta)p\binom{\varepsilon n}{2} \leqslant 2\varepsilon p n^2$ edges within the clusters;
- $\varepsilon k^2 \cdot (1+\eta)p(\frac{n}{k})^2 \leqslant 2\varepsilon p n^2$ edges between irregular pairs;
- $\binom{k}{2} \cdot dp(\frac{n}{k})^2 \leqslant dpn^2$ edges between regular pairs with density below d.

Thus we have $e(G^{\mathcal{R}}) \geqslant e(G) - (6\varepsilon + d)pn^2$. Note that the (V_i, V_j) induces a non-empty graph in $G^{\mathcal{R}}$ if and only if V_iV_j is an edge in $\mathcal{R}$. Therefore, we have

$$e(G) - (6\varepsilon + d)pn^2 \leqslant e(G^{\mathcal{R}}) \leqslant (1+\eta)p\frac{n^2}{k^2}e(\mathcal{R}),$$

which implies the desired bound. □

Finally, let us remark that Theorem 5.1.2 works for very sparse random graphs due to the following lemma.

Lemma 5.1.4 (Lemma 4 from [13]). *Let $\eta > 0$ and let $pn > \frac{8}{\eta^4(1-\eta)}$. Then, with high probability, the random graph $G(n,p)$ is (η, p) uniform.*

5.2 Tree embeddings in bipartite expander graphs

In this section we show how to embed trees in (ε, p)-regular pairs. One of the main difficulties while working with the sparse regularity lemma is that the vertex-by-vertex embedding strategy does not work in general. This is because the neighbourhood of typical vertex in a sparse regular pair is too small in order to use the *regularity inheritance*. That is, if (A, B) is an (ε, p)-regular pair of density p and $p = o(1)$, then the degree of a typical vertex in A is roughly $dp|B| = o(|B|)$, which is much smaller that what one needs in order to use a vertex-by-vertex embedding strategy. To deal with this problem, we will make use of the expansion properties of regular pairs and tree embedding results in expander graphs.

Roughly speaking, we say that a graph is *expander* if every set of vertices has a large outer neighbourhood. The neighbour expansion notion is particularly useful while embedding graphs such as paths, trees, and cycles. For instance, Friedman and Pippenger [51] proved that graphs satisfying some expansion condition contain all small trees size of bounded maximum degree. This result was improved by Haxell [62] allowing the embedding of larger trees in expander graphs. In our context, we want to use expansion conditions to embed trees in regular pairs. To do so, we will use an embedding result due to Balogh, Csaba, and Samotij [13] which extends the result of Friedman and Pippenger to the bipartite setting.

Definition 5.2.1 (Bipartite (q, d)-expander)**.** *Let $H = (V_1, V_2)$ be a bipartite graph such that $|V_1| \leqslant |V_2|$. Let $d \geqslant 2$ and let q be a positive integer such that $q < |V_1|$. We say that H is a bipartite (q, d)-expander if the following holds.*

1. *For every subset $X \subseteq V_i$ of size at most q we have $|N(X)| \geqslant d|X|$ for $i \in \{1, 2\}$.*
2. *For every subset $X \subseteq V_i$ of size at least q we have $|N(X)| \geqslant |V_{3-i}| - q$ for $i \in \{1, 2\}$.*

Lemma 5.2.2 (Corollary 12 from [13])**.** *Let $d \geqslant 2$ and let $H = (V_1, V_2)$ be a bipartite graph with $|V_1| \leqslant |V_2|$. Suppose that H is a bipartite $(q, d+1)$-expander with $0 < q < \frac{|V_1|}{(2d+1)}$. Then H contains a copy of every tree T with $\Delta(T) \leqslant d$ and bipartition classes A_1, A_2 with $|A_1| \leqslant |V_1| - (2d+1)m$ and $|A_2| \leqslant |V_2| - (2d+1)m$, respectively. Furthermore, for each $u \in A_i$ and $v \in V_i$ there exists an embedding $\varphi : V(T) \to H$ such that $\varphi(u) = v$.*

We cannot use Lemma 5.2.2 directly in (ε, p)-regular pairs, since regular pairs are not bipartite expanders. Indeed, it might be that a regular pair has vertices with too low degree or even isolated vertices. However, one can prove that any large subgraph of an (ε, p)-regular pairs contains an almost spanning subgraph which is a bipartite expander.

Lemma 5.2.3 (Lemma 19 from [13])**.** *Let (A, B) be an (ε, p)-regular pair such that $d_p(A, B) > \varepsilon$. Suppose that $|A| = |B| = m$ and let $A' \subseteq A$ and $B' \subseteq B$ be sets of size at least $(4D+6)\varepsilon m$. Then there are subsets $A'' \subseteq A'$ and $B'' \subseteq B'$ such that*

(a) $|A' \setminus A''| \leqslant \varepsilon m$ and $|B' \setminus B''| \leqslant \varepsilon m$, and

(b) the subgraph induced by (A'', B'') is a bipartite $(\varepsilon m, 2D + 2)$-expander.

5.3 Cutting trees with bounded maximum degree

Now we show how to cut a given tree T into a constant number of tiny rooted subtrees. The main difference with the results of Section 3.2 is that we need to control the number of neighbours that each subtree has and we also need that the root of each of these subtrees is at even distance from the root of T. To do so, we will modify the following result due to Balogh, Csaba, and Samotij [13].

Lemma 5.3.1 (Lemma 15 from [13])**.** *Let $\Delta \geqslant 2$ and let (T, r) be a rooted tree with $\Delta(T) \leqslant \Delta$. If $|T| \geqslant \beta^{-1}$, then there exists a family of $t \leqslant 4\beta^{-1}$ disjoint rooted subtrees $(T_i, r_i)_{i\in[t]}$ such that $V(T) = V(T_1) \cup \cdots \cup V(T_t)$ and for each $i \in [t]$ we have*

1. *$|T_i| \leqslant \Delta^2\beta|T|$,*
2. *T_i is connected (by an edge) to at most Δ^3 other subtrees, and*
3. *T_i is rooted at r_i and all the children of r_i belong to T_i.*

Given a tree T, let $(T_i, r_i)_{i\in[t]}$ be the family given by Lemma 5.3.1. We may define an auxiliary graph T_Π (see figure 5.2), called *cluster tree*, with vertex set $V(T_\Pi) = [t]$ and edge set

$$E(T_\Pi) = \{ij \mid\ T_i \text{ and } T_j \text{ are adjacent in } T\}.$$

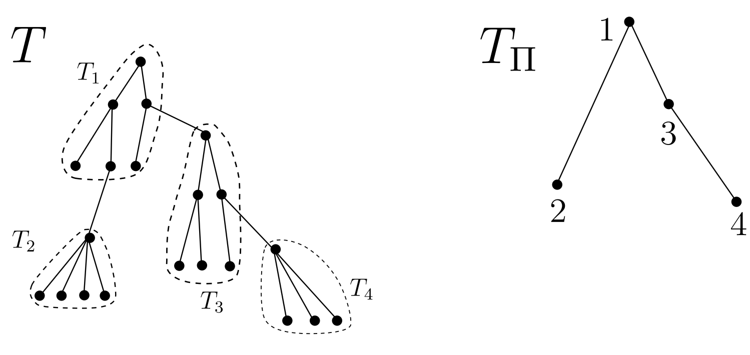

Figure 5.2: Cluster tree.

Note that Lemma 5.3.1 implies that $|T_\Pi| \leqslant 4\beta^{-1}$, which is the best one could hope by imposing property (1). Moreover, property (2) implies that $\Delta(T_\Pi) \leqslant \Delta^3$, which will play a crucial role in our embedding strategy. We only need to refine the partition given by Lemma 5.3.1 in order to impose that the root of each subtree is at even distance from the root of T, which is a stronger property than (3).

Proposition 5.3.2. *Let $\Delta \geqslant 2$ and let (T, r) be a rooted tree with $\Delta(T) \leqslant \Delta$. If $|T| \geqslant \beta^{-1}$, then there exists a family of $t \leqslant 4\beta^{-1}\Delta$ disjoint rooted subtrees $(T_i, r_i)_{i\in[t]}$ such that $V(T) = V(T_1) \cup \cdots \cup V(T_t)$ and for each $i \in [t]$ we have*

1. $|T_i| \leqslant \Delta^4\beta|T|$,

2. T_i *is rooted at* r_i *and the distance from* r_i *to* r *is even,*

3. all the children of r_i *belong to* T_i, *and*

4. the corresponding cluster tree has maximum degree at most Δ^4.

Proof. Starting with the partition given by Lemma 5.3.1, we will refine this partition as we run a breadth first search (BFS) on (T, r). Suppose that in this search we have reached a vertex v, which is the root of a subtree in the current partition, such that v and every root before v in the search are at even distance from each other in the current partition.

If there is a root u of some subtree in the current partition, which is at odd distance from v and such that the subtree rooted at v is adjacent to u, then we may update the partition by splitting the tree rooted at u (each neighbour of u is now the root of a subtree) and adding u to the subtree rooted at v. We repeat this process for every such u. Note that after these splittings, the root of each tree that is adjacent to the tree rooted at v is at even distance from all the previous roots. Moreover, a subtree of the original partition can only be split by this process when the BFS reaches its parent. Since each subtree has only one parent, they are split at most once into at most Δ new subtrees and therefore the final partition has at most $4\Delta\beta^{-1}$ new subtrees. For the same reason, the maximum degree of the cluster tree cannot go higher than Δ^4, since the original T_Π had maximum degree at most Δ^3.

Finally, the size of each subtree grows by at most Δ^3 if the roots of its children are added. Since the update only moves forward in the BFS order, at the end of the process each subtree has size at most $\Delta^2\beta|T| + \Delta^3 \leqslant \Delta^4\beta|T|$. $\square$

5.4 Matching structure in the reduced graph

In this section we prove that if H is an (η, p)-upper-uniform graph with $2e(H) \geqslant (\varrho + \frac{\delta}{2})pn^2$, then H has an (ε, p, d)-reduced graph $\mathcal{R}$ containing a cluster X of large degree such that its neighbourhood can be partitioned as $N(X) = V(\mathcal{M}) \cup \mathcal{Y}$, where $\mathcal{M}$ is a matching and $\mathcal{Y}$ is an independent set. Moreover, denoting by $\mathcal{H}$ the bipartite graph induced by $\mathcal{Y}$ and $\mathcal{Z} := N(\mathcal{Y}) \setminus (X \cup N(X))$, then either $\mathcal{M}$ is large enough or every cluster in $\mathcal{Y}$ has large degree in $\mathcal{H}$.

Proposition 5.4.1. *Let* $\varepsilon, \delta, \varrho \in (0, 1)$ *and let* $d = \frac{\delta}{100}$. *There exist* $n_0, k, K_0 \in \mathbb{N}$ *and* $n_0 > 0$ *such that* $1/\varepsilon \leqslant k \leqslant K_0$ *and that for all* $0 < \eta \leqslant \eta_0$, $p \in (0, 1)$ *and* $n \geqslant n_0$ *the following holds. Every* (η, p)*-upper uniform graph* H *on* n *vertices with* $2e(H) \geqslant (\varrho+\delta/2)pn^2$ *admits an* (ε, p)*-regular partition with* k *parts such that its* (ε, p, d)*-reduced graph* $\mathcal{R}$ *satisfies the following. There exist* $X \in V(\mathcal{R})$, *a matching* $\mathcal{M}$ *and a bipartite induced subgraph* $\mathcal{H} = \mathcal{R}[\mathcal{Y}, \mathcal{Z}]$ *such that*

(a) $N(X) = V(\mathcal{M}) \cup \mathcal{Y}$ *and* $V(\mathcal{M}) \cap \mathcal{Y} = \emptyset$;

(b) $|V(\mathcal{M})| + |\mathcal{Y}| \geqslant \left(\varrho + \frac{\delta}{3}\right) k$; *and*

(c) *for all $Y \in \mathcal{Y}$ we have*
$$|N_{\mathcal{H}}(Y)| \geqslant \left(\varrho + \frac{\delta}{4}\right)\frac{k}{2} - \frac{|V(\mathcal{M})|}{2}.$$

Proof. Given $\varepsilon' = \min\{\frac{\varepsilon}{5}, \frac{\delta}{1000}\}$ and $k_0 = \frac{1}{\varepsilon'}$, let η_0, n_0' and K_0' be the outputs of the Sparse Regularity Lemma (Theorem 5.1.2) with parameters ε' and k_0. Setting $n_0 = n_0'$ and $\eta_0 = \min\{\eta_0', \frac{\delta}{1000}\}$, let H be an (η, p)-upper uniform graph on $n \geqslant n_0$ vertices and $0 < \eta \leqslant \eta_0$. Then H admits an (ε', p)-regular partition $V(H) = V_0' \cup V_1' \cup \cdots \cup V_\ell'$, with $\frac{1}{\varepsilon'} \leqslant \ell \leqslant K_0$, and let us denote by $\mathcal{R}'$ the $(\varepsilon', p, 2d)$-reduced graph of H with respect to this regular partition. By Proposition 5.1.3 and the bound on $e(H)$ we have

$$e(\mathcal{R}') \geqslant (1+\eta)^{-1}\left(\varrho + \frac{\delta}{2}\right)\frac{\ell^2}{2} - (6\varepsilon' + 2d)\ell^2 \geqslant \left(\varrho + \frac{\delta}{3}\right)\frac{\ell^2}{2}. \tag{5.3}$$

Note that (5.3) implies that the average degree of $\mathcal{R}'$ is at least $(\varrho + \frac{\delta}{3})\ell$. Thus, by successively removing vertices of low degree, we may find a subgraph $\mathcal{R}_0 \subseteq \mathcal{R}'$ such that

$$d(\mathcal{R}_0) \geqslant \left(\varrho + \frac{\delta}{3}\right)\ell \quad \text{and} \quad \delta(\mathcal{R}_0) \geqslant \left(\varrho + \frac{\delta}{3}\right)\frac{\ell}{2}.$$

In particular, this implies that there exists a cluster $X' \in V(\mathcal{R}_0)$ with degree at least $(\varrho + \frac{\delta}{3})\ell$ in $\mathcal{R}_0$. Applying Lemma 3.3.7 to $N_{\mathcal{R}_0}(X')$, we find an independent set $\mathcal{I}$, a matching $\mathcal{M}'$ and a collection of triangles Γ that partition $N_{\mathcal{R}_0}(X') = \mathcal{I} \cup V(\mathcal{M}') \cup V(\Gamma)$, and moreover, by writing $V(\mathcal{M}') = \mathcal{V}_1 \cup \mathcal{V}_2$ we have that $N_{\mathcal{R}_0}(\mathcal{I}) \subseteq \mathcal{V}_1$. Note that the minimum degree on $\mathcal{R}_0$ implies that for all $Y \in \mathcal{I}$ we have

$$|N_{\mathcal{R}_0}(Y) \setminus (X' \cup N_{\mathcal{R}_0}(X))| \geqslant \left(\varrho + \frac{\delta}{3}\right)\frac{\ell}{2} - 1 - \frac{|V(\mathcal{M})|}{2} \geqslant \left(\varrho + \frac{\delta}{4}\right)\frac{\ell}{2} - \frac{|V(\mathcal{M})|}{2}. \tag{5.4}$$

If there are no triangles in this decomposition, then we would finish the proof by setting $\mathcal{M} = \mathcal{M}'$ and $\mathcal{H}$ as the bipartite graph induced by $\mathcal{I}$ and $N_{\mathcal{R}'}(\mathcal{I}) \setminus (X \cup N_{\mathcal{R}'}(X))$. If is not the case, for each $i \in [\ell]$ we may arbitrarily partition $V_i = V_{i,0} \cup V_{i,1} \cup V_{i,2}$ so that $|V_{0,i}| \leqslant 1$ and $|V_{i,1}| = |V_{i,2}|$. Noting that $|V_{i,1}| = |V_{i,2}| \geqslant |V_i|/3$ for every $i \in [\ell]$, because of Lemma 5.1.1, for each $V_iV_j \in E(\mathcal{R}')$ and $a, b \in \{1, 2\}$ the pair $(V_{i,a}, V_{j,b})$ is (ε, p)-regular with density at least d. Moreover, by setting $V_0 = V_0' \cup V_{1,0} \cup \cdots \cup V_{\ell,0}$ we conclude that $V(H) = V_0 \cup V_{1,2} \cup V_{2,2} \cup \cdots \cup V_{\ell,1} \cup V_{\ell,2}$ is an (ε, p)-regular partition with $2\ell + 1$ parts. Let $\mathcal{R}$ be the (ε, p, d)-reduced graph of H with respect to this partition, and let $k = 2\ell$ be the number of vertices of $\mathcal{R}$ (note that $\mathcal{R}$ is a blow-up of $\mathcal{R}'$). Let X be one of the clusters coming from X', and $\mathcal{Y}$ be the set of all the $V_{i,a}$ such that $V_i' \in \mathcal{I}$ and $a \in \{1, 2\}$. Now note that each triangle in Γ can be decomposed as three disjoint edges in $\mathcal{R}$. Then we set

$$\mathcal{M} = \bigcup_{V_iV_j \in \mathcal{M}'} \{V_{i,1}V_{j,1}, V_{i,2}V_{j,2}\} \cup \bigcup_{V_aV_bV_c \in \Gamma} \{V_{a,1}V_{b,1}, V_{b,2}V_{c,1}, V_{c,2}V_{a,2}\}$$

and $\mathcal{Z} = N_{\mathcal{R}}(\mathcal{Y}) \setminus (X \cup N_{\mathcal{R}}(X))$. Letting $\mathcal{H}$ as the bipartite graph induced by $\mathcal{Y}$ and $\mathcal{Z}$, it is clear that X, $\mathcal{M}$ and $\mathcal{H}$ satisfy (a) and (b), (c) follows from (5.4).

□

5.5 Proof of Theorem 5.0.2

As we mentioned in the sketch of the proof, the idea is to use the structure given by Proposition 5.4.1 to embed any tree with ϱn edges and bounded maximum degree. To do so, we first need to cut the tree into a family $(T_i, r_i)_{i\in[t]}$ of tiny subtrees such that the root of all the subtrees are in the same colour class (see Proposition 5.3.2). The main idea of the proof is to first assign each T_i to some edge of $\mathcal{M} \cup \mathcal{H}$ so that each of those regular pairs has enough space to embed all the trees assigned to it. After this, we use Lemma 5.2.3 to clean each regular pair that is used, and thus each subtree T_i is assigned to a pair $(Y_{i,1}, Y_{i,2})$ which induces a bipartite expander graph that connects well with a large subset of X (see Claim 5.5.1). We embed the subtrees one-by-one following a BFS order in the cluster tree T_Π, using Lemma 5.2.2 to map each subtree into the bipartite expander graph assigned to it.

Now we are ready to prove Theorem 5.0.2.

Proof of Theorem 5.0.2. Let n_0', K_0 and η_0 be the outputs of Proposition 5.4.1 with inputs δ, ϱ and $\varepsilon = 2^{-28}\Delta^{-6}\delta^4$. We set

$$\beta = \frac{\delta^2}{2^{12}k\Delta^4} \quad \text{and} \quad C_0 = \frac{2^{17}10^2\Delta^5 K_0^2}{\delta^3}, \tag{5.5}$$

and let $n_0 = \max\{n_0', \beta^{-1}\}$ and $n \geqslant n_0$. Given $p \in (0,1)$ such that $pn \geqslant C_0$ and $0 < \eta \leqslant \eta_0$, let G be an (η, p)-uniform graph on n vertices and let $G' \subseteq G$ be a subgraph with

$$2e(G') \geqslant (\varrho+\delta)2e(G) \geqslant (1-\eta)(\varrho+\delta)pn^2 \geqslant \left(\varrho + \frac{\delta}{2}\right)pn^2.$$

Since G' is (η,p)-upper uniform, by Proposition 5.4.1 we may find an (ε,p)-regular partition $V(G') = V_0 \cup V_1 \cup \cdots \cup V_k$, with $\frac{1}{\varepsilon} \leqslant k \leqslant K_0$, such that the $(\varepsilon, p, \frac{\delta}{100})$-reduced graph $\mathcal{R}$, with respect to this partition, contains a cluster X, a matching $\mathcal{M}$ and a bipartite subgraph $\mathcal{H} = (\mathcal{Y}, \mathcal{Z})$ satisfying the conclusions of Proposition 5.4.1.

Let T be a tree with ϱn edges such that $\Delta(T) \leqslant \Delta$. We consider the bipartition of T that assigns colour 1 to the smaller partition class of T and colour 2 to the larger one, and then we choose an arbitrary vertex r in colour 1 as the root of T. We apply Proposition 5.3.2 to (T, r), with parameters β and Δ, obtaining a family $(T_i, r_i)_{i\in[t]}$ of $t \leqslant 4\beta^{-1}\Delta$ rooted trees such that $|T_i| \leqslant \Delta^4\beta\varrho n$ for all $i \in [t]$. Furthermore, each root r_i is at even distance from r and therefore every root has colour 1. For $i \in [t]$, let us write $T_{i,j}$ for the set of vertices of T_i having colour $j \in \{1,2\}$.

Let m denote the size of the clusters in the regular partition and observe that $m \geqslant (1-\varepsilon)\frac{n}{k}$. The heart of the proof is the following claim.

Claim 5.5.1. *For each $i \in [t]$, there are sets $(Y_{i,1}, Y_{i,2})$ and $W_i \subseteq X$ such that the following holds.*

(1) *There is an edge $V_{i,1}V_{i,2} \in \mathcal{M} \cup E(\mathcal{H})$ such that $Y_{i,1} \subseteq V_{i,1}$ and $Y_{i,2} \subseteq V_{i,2}$. Moreover, if $V_{i,1}V_{i,2} \in E(\mathcal{H})$ then $V_{i,2} \in \mathcal{Y}$.*

(2) *For* $\ell \neq i$ *and* $j, j' \in \{1,2\}$, $Y_{i,j} \cap Y_{\ell,j'} = \emptyset$.

(3) *For* $j \in \{1,2\}$, $|Y_{i,j}| \geqslant |T_{i,j}| + 13\Delta\varepsilon m$.

(4) $G'[Y_{i,1}, Y_{i,2}]$ *is a bipartite* $(\varepsilon m, 2\Delta+2)$*-expander.*

(5) *Every vertex of* $Y_{i,2}$ *has at least* $\frac{\delta}{200}pm$ *neighbours in* W_i.

(6) *If* T_ℓ *is a child of* T_i *in the cluster tree, then every vertex of* W_i *has at least* $\Delta+1$ *neighbours in* $Y_{\ell,2}$.

Before proving Claim 5.5.1, let us show how to use it in order to finish the proof of Theorem 5.0.2. Assume that we have ordered $[t]$ so that if T_i is below T_ℓ, with respect to the root of T, then $i \leqslant \ell$. Starting with the subtree containing r, we will embed $(T_i)_{i\in[t]}$ following this ordering. Let us denote by φ the partial embedding of T. For every embedded subtree (T_i, r_i) we will ensure that

(*a*) $\varphi(r_i) \in W_s$ for some $s \leqslant i$, and

(*b*) $\varphi(T_{i,j} \setminus \{r_i\}) \subseteq Y_{i,j}$ for $j \in \{1,2\}$.

Suppose we are about to embed a subtree T_ℓ which is a child of some subtree T_i that was already embedded satisfying (a) and (b). Let $v_i \in V(T_i)$ be the parent of r_ℓ and note that v_i is embedded into some vertex $\varphi(v_i) \in Y_{i,2}$ (since v_i is adjacent to r_ℓ and every root has colour 1).

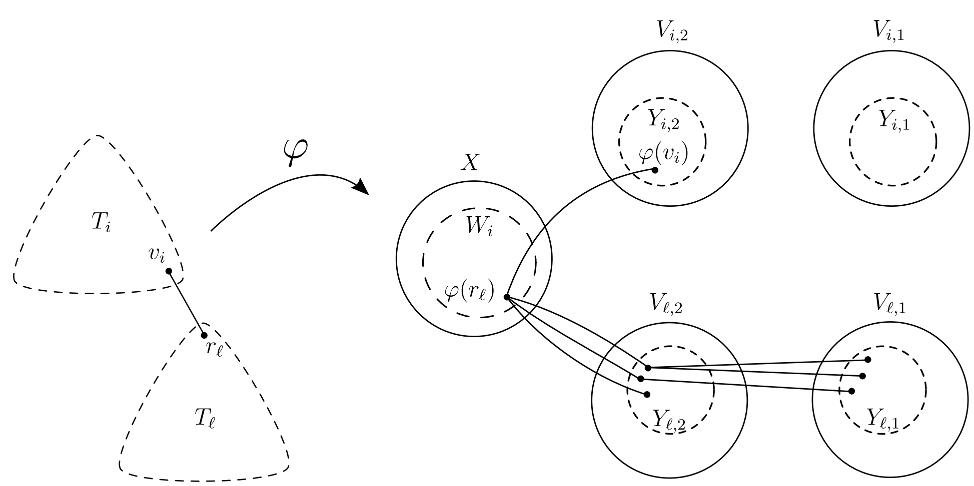

Figure 5.3: Embedding of T_ℓ

Then, because of Claim 5.5.1 (5)

$$|N_{G'}(\varphi(v_i)) \cap W_i| \geqslant \frac{\delta}{200}pm \geqslant (1-\varepsilon)\frac{\delta C_0}{200k} \geqslant \frac{8\Delta}{\beta} \geqslant 2t$$

and therefore at least one neighbour of $\varphi(v_i)$ has not been used during the embedding. We choose any unused vertex $w_\ell \in W_i \cap N_{G'}(\varphi(v_i))$ and set $\varphi(r_\ell) = w_\ell$ (when we embed T_1, we choose any vertex vetex $w_1 \in W_1$ as the image of $r_1 = r$). By Claim 5.5.1 (4) we know that $G'[Y_{i,1}, Y_{i,2}]$ is a bipartite $(\varepsilon m, 2\Delta + 2)$-expander, we will prove now that

$$G'[Y_{\ell,1} \cup \{w_\ell\}, Y_{\ell,2}] \text{ is a bipartite } (\varepsilon m + 1, \Delta + 1)\text{-expander.}$$

Indeed, since $G'[Y_{i,1}, Y_{i,2}]$ is a bipartite $(\varepsilon m, 2\Delta+2)$-expander is easy to see that the expansion conditions hold for every set $X \subseteq Y_{\ell,1} \cup Y_{\ell,2}$. Let $X' \subseteq Y_{\ell,1}$ non-empty and let us consider $X = X' \cup \{w_\ell\}$. If $|X'| \leqslant \varepsilon m$ then we have

$$|N_{G'}(X) \cap Y_{\ell,2}| \geqslant (2\Delta + 2)|X'| \geqslant (\Delta + 1)|X|,$$

where the first inequality follows because $G'[Y_{\ell,1}, Y_{\ell,2}]$ is bipartite $(\varepsilon m, 2\Delta + 2)$-expander. Similarly, if $|X'| \geqslant \varepsilon m$ then we have

$$|N_{G'}(X) \cap Y_{\ell,2}| \geqslant |N_{G'}(X') \cap Y_{\ell,2}| \geqslant |Y_{\ell,2}| - (\varepsilon m + 1).$$

Finally, if $X = \{w_\ell\}$ then by Claim 5.5.1 (6) we know that $|N_{G'}(w_\ell) \cap Y_{\ell,2}| \geqslant \Delta + 1$, and therefore $G'[Y_{\ell,1} \cup \{w_\ell\}, Y_{\ell,2}]$ is a bipartite $(\varepsilon m + 1, \Delta + 1)$-expander.

To complete the embedding of T_ℓ, note that because of Claim 5.5.1 (3) we have

$$|Y_{\ell,j}| - (2\Delta + 1)(\varepsilon m + 1) \geqslant |T_{\ell,j}| + 13\Delta\varepsilon m - 6\Delta\varepsilon m \geqslant |T_{\ell,j}|$$

for $j \in \{1, 2\}$. Thus, using Lemma 5.2.2 we may extend φ to T_ℓ, embedding T_ℓ into $(Y_{\ell,1} \cup \{w_\ell\}, Y_{\ell,2})$ so that $\varphi(T_{\ell,j} \setminus \{r_\ell\}) \subseteq Y_{\ell,j}$ for $j \in \{1, 2\}$ and w_ℓ is fixed as the image of r_ℓ (we remark that Claim 5.5.1 (2) allows us to ensure that at every step of the embedding we are using unused vertices). □

Proof of Claim 5.5.1. Let σ be a permutation on $[t]$ such that for all $1 \leqslant i < j \leqslant t$ we have

$$|T_{\sigma(i),2}| - |T_{\sigma(i),1}| \geqslant |T_{\sigma(j),2}| - |T_{\sigma(j),1}|.$$

Recall that we chose colour 2 for the larger partition class of $V(T)$. Therefore, for every $\ell \in [t]$ we have

$$\sum_{i=1}^{\ell} (|T_{\sigma(i),2}| - |T_{\sigma(i),1}|) \geqslant 0. \tag{5.6}$$

The proof of Claim 5.5.1 will be done in two stages. In the first stage, for each $i \in [t]$ the subtree T_i will be assigned to a pair of sets $(X_{i,1}, X_{i,2})$, contained in some edge from $\mathcal{M} \cup E(\mathcal{H})$, such that $|X_{i,j}| = |T_{i,j}| + 16\Delta\varepsilon m$ for $j \in \{1, 2\}$. In the second stage, we will remove some vertices from each set in order to find the sets $W_i \subseteq X$ and $Y_{i,j} \subseteq X_{i,j}$ satisfying the properties (1) – (6) from Claim 5.5.1.

Stage 1 (Assignation): In this stage we will prove that for each $i \in [t]$, there exist an edge $V_{i,1}V_{i,2} \in \mathcal{M} \cup E(\mathcal{H})$ and sets $X_{i,j} \subseteq V_{i,j}$, for $j \in \{1, 2\}$, such that

(A) $X_{i,j} \cap X_{\ell,j'} = \emptyset$ if $\{i,j\} \neq \{\ell, j'\}$;

(B) $|X_{i,j}| = |T_{i,j}| + 16\Delta\varepsilon m$; and

(C) if $(V_{i,1}, V_{i,2}) \in E(\mathcal{H})$ then $V_{i,2} \in \mathcal{Y}$.

The assignment will be done in two steps following the order given by σ. At step 1 we assign trees to edges from $\mathcal{H}$ until we use a large proportion of $\mathcal{Y} \cup \mathcal{Z}$, and at step 2 we will use edges from $\mathcal{M}$ ensuring that the clusters from each edge of $\mathcal{M}$ are used in a balanced way.

Step 1: We will assume that $|\mathcal{M}| \leqslant (\varrho + \frac{\delta}{16})k$, as otherwise we just skip this step. Let us set $Q = (\varrho + \frac{\delta}{4})k - |V(\mathcal{M})|$ and note that we have

$$|\mathcal{Y}| \geqslant Q \geqslant \frac{\delta}{16}k \quad \text{and} \quad d_{\mathcal{H}}(Y) \geqslant \frac{Q}{2} \text{ for all } Y \in \mathcal{Y}.$$

We will choose sets in $\mathcal{Y} \cup \mathcal{Z}$ until we have assigned at least $(1 - \frac{\delta}{16})Qm$ vertices to $\mathcal{Y} \cup \mathcal{Z}$. Following the order of σ, assume that we have made the assignation up to some $0 \leqslant \ell \leqslant t-1$ and we are about to assign the tree $T_{\sigma(\ell+1)}$. Suppose that there are $Y \in \mathcal{Y}$ such that

$$\sum_{X_{\sigma(i),2} \subseteq Y} |X_{\sigma(i),2}| \leqslant m - (\Delta^4 \beta n + 16\Delta\varepsilon m), \tag{5.7}$$

and $Z \in N_{\mathcal{H}}(Y)$ with

$$\sum_{X_{\sigma(i),1} \subseteq Z} |X_{\sigma(i),1}| \leqslant m - (\Delta^4 \beta n + 16\Delta\varepsilon m). \tag{5.8}$$

Since $|T_{\sigma(\ell+1)}| \leqslant \Delta^4 \beta \varrho n$, we can select sets $X_{\sigma(\ell+1),1} \subseteq Z$ and $X_{\sigma(\ell+1),2} \subseteq Y$, disjoints from the previously chosen sets, such that $|X_{\sigma(\ell+1),j}| = |T_{\sigma(\ell+1),j}| + 16\Delta\varepsilon m$ for $j \in \{1,2\}$. So, if there is no $Y \in \mathcal{Y}$ satisfying (5.7), then we have

$$\begin{aligned}
\sum_{i=1}^{\ell} |T_{\sigma(i)}| \geqslant \sum_{i=1}^{\ell} |T_{\sigma(i),2}| = \sum_{i=1}^{\ell} \Big(|X_{\sigma(i),2}| - 16\Delta\varepsilon m\Big) \\
&\geqslant |\mathcal{Y}|m - t \cdot 16\Delta\varepsilon m - k \cdot (\Delta^4 \beta n + 16\Delta\varepsilon m) \\
&\geqslant |\mathcal{Y}|m - \frac{\delta^2}{16^2}km \\
&\geqslant \left(1 - \frac{\delta}{16}\right)Qm.
\end{aligned}$$

This means that we have already used enough vertices from $\mathcal{Y} \cup \mathcal{Z}$. On the other hand, if every Y satisfying (5.7) has no neighbours satisfying (5.8), then we may use (5.6) to deduce

$$\begin{aligned}
\sum_{i=1}^{\ell} |T_\sigma(i)| \geqslant 2\sum_{i=1}^{\ell} |T_{\sigma(i),1}| = 2\sum_{i=1}^{\ell} \Big(|X_{\sigma(i),1}| - 16\Delta\varepsilon m\Big) \\
&\geqslant 2d_{\mathcal{H}}(Y)m - t \cdot 32\Delta\varepsilon m - k \cdot 2(\Delta^4 \beta n + 16\Delta\varepsilon m) \\
&\geqslant Qm - \frac{\delta^2}{16^2}km \\
&\geqslant \left(1 - \frac{\delta}{16}\right)Qm.
\end{aligned}$$

This means that if at step $\ell+1 \in [t]$ we could not find a pair (Y, Z) satisfying (5.7) and (5.8), then we have used vertices at least $(1-\frac{\delta}{16})Qm$ vertices from $\mathcal{Y} \cup \mathcal{Z}$ at step ℓ.

Step 2: Let $0 \leqslant \ell_0 \leqslant t$ be such that $T_{\sigma(1)}, \ldots, T_{\sigma(\ell_0)}$ have been assigned to $\mathcal{Y} \cup \mathcal{Z}$, satisfying (A),(B) and (C), and

$$\left(1-\frac{\delta}{16}\right)Qm \leqslant \sum_{i=1}^{\ell_0} |T_{\sigma(i)}| \leqslant \left(1-\frac{\delta}{16}\right)Qm + \Delta^4\beta\varrho n. \tag{5.9}$$

Assume that $\ell_0 < t$, otherwise we are done. For $\ell_0 + 1 \leqslant i \leqslant t$ we will assign each $T_{\sigma(i)}$ to some edge $AB \in \mathcal{M}$. At each step we will ensure that for every edge $AB \in \mathcal{M}$ we have

$$\left| \sum_{X_{\sigma(i),j} \subseteq A} |X_{\sigma(i),j}| - \sum_{X_{\sigma(i),j} \subseteq B} |X_{\sigma(i),j}| \right| \leqslant \Delta^4\beta\varrho n. \tag{5.10}$$

Suppose we are about to assign a subtree $T_{\sigma(\ell)}$, for some $\ell \geqslant \ell_0 + 1$, and that (5.10) holds at step $i = \ell - 1$ (note that (5.10) holds trivially at step ℓ_0). Suppose that there is an edge $AB \in \mathcal{M}$ such that

$$\max\left\{ \sum_{X_{\sigma(i),j} \subseteq A} |X_{\sigma(i),j}|, \sum_{X_{\sigma(i),j} \subseteq B} |X_{\sigma(i),j}| \right\} \leqslant m - (\Delta^4\beta\varrho n + 16\Delta\varepsilon m). \tag{5.11}$$

Assuming that $\sum_{X_{\sigma(i),j} \subseteq A} |X_{\sigma(i),j}| \leqslant \sum_{X_{\sigma(i'),j'} \subseteq B} |X_{\sigma(i'),j'}|$, we let $j^\star = \operatorname*{argmax}_{j \in \{1,2\}} |T_{\sigma(\ell),j}|$ and then we may take sets

- $X_{\sigma(\ell),j^\star} \subseteq A$ with $|X_{\sigma(\ell),j^\star}| = |T_{\sigma(\ell),j^\star}| + 16\Delta\varepsilon m$, and
- $X_{\sigma(\ell),3-j^\star} \subseteq B$ with $|X_{\sigma(\ell),3-j^\star}| = |T_{\sigma(\ell),3-j^\star}| + 16\Delta\varepsilon m$.

disjoints from the previously chosen sets. Note that we have assigned the larger colour class of $T_{\sigma(\ell)}$ to the less occupied cluster in $\{A, B\}$. Furthermore, since (5.10) holds at step $\ell - 1$ and as $|T_{\sigma(\ell)}| \leqslant \Delta^4\beta\varrho n$, the assignment of $T_{\sigma(\ell)}$ implies that (5.10) holds at step ℓ. So suppose that (5.11) does not hold at step $\ell - 1$ for any $AB \in \mathcal{M}$. Then we have

$$\begin{aligned} \sum_{i=\ell_0+1}^{\ell-1} |T_{\sigma(i)}| &\geqslant |V(\mathcal{M})|m - t \cdot 32\Delta\varepsilon m - k \cdot (3\Delta^4\beta\varrho n + 32\Delta\varepsilon m) \\ &\geqslant |V(\mathcal{M})|m - \frac{\delta}{16}km, \end{aligned}$$

which together with (5.9) yields

$$\begin{aligned} \sum_{i=1}^{\ell-1} |T_{\sigma(i)}| &\geqslant \left(1-\frac{\delta}{16}\right)Qm + |V(\mathcal{M})|m - \frac{\delta}{16}km \\ &\geqslant \left(1-\frac{\delta}{16}\right)\left(\varrho + \frac{\delta}{4}\right)km - \frac{\delta}{16}km \\ &\geqslant \left(\varrho + \frac{\delta}{8}\right)km \\ &\geqslant \left(\varrho + \frac{\delta}{16}\right)n, \end{aligned}$$

which is impossible since $|T| = \varrho n$. This implies that we can make the assignation for each $\ell \in [t]$.

Stage 2 (Cleaning): Assume that the cluster tree is ordered according to a BFS starting from the subtree which the root of T. Starting with a leaf of the cluster tree, suppose that we have found the sets $Y_{i,j}$ satisfying properties $(1)-(6)$ for all subtrees T_i below T_ℓ in the order of the cluster tree. Let us define

$$W_\ell := \{v \in X : d(v, Y_{i,2}) \geqslant \Delta + 1 \text{ for all } i \text{ such that } T_i \text{ is a child of } T_\ell\},$$

we want to prove that W_ℓ has a reasonable size. Given a child T_i of T_ℓ in the cluster tree, we have that

$$|Y_{i,2}| \geqslant |T_{i,j}| + 13\Delta\varepsilon m \geqslant (\Delta+1)\varepsilon m$$

and therefore, since $(X, V_{i,2})$ is (ε, p)-regular, by Lemma 5.1.1 there are at most $(\Delta+1)\varepsilon m$ vertices in X with less than $\Delta + 1$ neighbours in $Y_{i,2}$. Since the auxiliary tree has maximum degree at most Δ^4, then W_ℓ has at least

$$|X| - (\Delta+1)\Delta^4\varepsilon|X| \geqslant \frac{m}{2}$$

vertices. Now, since $(X, V_{\ell,2})$ is (ε, p)-regular, then by Lemma 5.1.1 the pair $(W_\ell, V_{\ell,2})$ is $(2\varepsilon, p)$-regular with p-density at least $\frac{\delta}{100} - \varepsilon$. By Lemma 5.1.1 there are at most $2\varepsilon m$ vertices of $V_{\ell,2}$ with less than

$$\left(\frac{\delta}{100} - 3\varepsilon\right) p|W_\ell| \geqslant \frac{\delta}{200} pm$$

neighbours in W_ℓ. We remove each such vertex from $X_{\ell,2}$ thus obtaining a set $X'_{\ell,2}$ such that every vertex in $X'_{\ell,2}$ has at least $\frac{\delta}{200}pm$ neighbours in W_ℓ. Now, we need to find an expander subgraph of $(X_{\ell,1}, X'_{\ell,2})$. Since $(V_{\ell,1}, V_{\ell,2})$ is (ε, p)-regular with $d_p(V_{\ell,1}, V_{\ell,2}) \geqslant \frac{\delta}{100}$ and

$$|X_{\ell,1}|, |X'_{\ell,2}| \geqslant 16\Delta\varepsilon m - 2\varepsilon m \geqslant (4\Delta+6)\varepsilon m,$$

we may use Lemma 5.2.3 to obtain a pair $(Y_{\ell,1}, Y_{\ell,2})$, with $Y_{\ell,1} \subseteq X_{\ell,1}$ and $Y_{\ell,2} \subseteq X'_{\ell,2}$, such that $G'[Y_{\ell,1}, Y_{\ell,2}]$ is bipartite $(\varepsilon m, 2\Delta+2)$-expander and satisfies $|Y_{\ell,j}| \geqslant |X_{\ell,j}| - 3\varepsilon m \geqslant |T_{\ell,j}| + 13\Delta\varepsilon m$ for $j \in \{1, 2\}$. $\square$

5.6 Applications in Ramsey theory

We end this chapter with some quick applications of Theorem 1.4.3 in Ramsey theory. For $s \geqslant 2$ and graphs $H_1, \ldots, H_s$, we say that a graph G is $(H_1, \ldots, H_k)$-Ramsey, denoted by $G \to (H_1, \ldots, H_s)$, if for every colouring of the edges of G with s colours there exists $i \in [s]$ such that G contains a copy of H_i in colour i. Given families of graphs $\mathcal{F}_1, \ldots, \mathcal{F}_s$, we say that G is $(\mathcal{F}_1, \ldots, \mathcal{F}_s)$-Ramsey, denoted by $G \to (\mathcal{F}_1, \ldots, \mathcal{F}_s)$, if for every colouring of the edges of G with s colours there exists $i \in [s]$ such that G contains a copy of every member of $\mathcal{F}_i$ in colour i. Furthermore, we write $G \xrightarrow{s} H$ when $H_1 = \cdots = H_s = H$ and $G \xrightarrow{s} \mathcal{F}$ when $\mathcal{F}_1 = \cdots = \mathcal{F}_s = \mathcal{F}$, in which case we say that G is (H, s)-Ramsey and $(\mathcal{F}, s)$-Ramsey respectively.

For a graph H and $s \geqslant 2$, one asks for threshold probability $p^* = p(n)$ such that if $p \gg p^*$ then, with high probability, the random graph $G(n,p)$ is (H,s)-Ramsey. The first result regarding this question, proved by Frankl and Rödl [49] and Łuczak, Ruciński and Voigt [87], states that for $p \geqslant C\sqrt{n}$ the random graph is $(K_3, 2)$-Ramsey with high probability. The systematic study of Ramsey properties of the random graph was initiated by Rödl and Ruciński [97, 98] who proved the following result.

Theorem 5.6.1. *Let $s \geqslant 2$ and let H be a graph that is not a forest consisting of stars and paths of length 3. Then there exist positive constants c and C such that*

$$\lim_{n\to\infty} \mathbb{P}[(G(n,p) \xrightarrow{s} H] = \begin{cases} 1 & \text{if } p \geqslant Cn^{-\frac{1}{m_2(H)}}, \\ 0 & \text{if } p \leqslant cn^{-\frac{1}{m_2(H)}}, \end{cases}$$

where

$$m_2(H) = \max\left\{\frac{e(H') - 1}{|H| - 2} : H' \subseteq H \text{ and } |H'| \geqslant 3\right\}.$$

In the modern probabilistic combinatorics, there are at least 3 different ways to prove Theorem 5.6.1: the *transference principle* of Conlon and Gowers [37], the *multi-round exposure method* of Schacht [102], and the *hypergraph container method* developed by Balogh, Morris, and Samotij [14], and, independently, by Saxton and Thomasson [101]. However, as far as we understand, none of this methods works for graphs whose vertices grows linearly in n.

The aim of this section is to make some progress in this area studying the Ramsey number of linear sized trees in the random graph, in particular, we prove a random analogue of the Erdős-Graham conjecture [44]. As in the dense case (Corollary 4.5.1), we may deduce the multicolour Ramsey number of trees in random graphs from Theorem 1.4.3.

Corollary 5.6.2. *Let $s, \Delta \geqslant 2$ and let $\varepsilon > 0$. Then there exists a constant $C > 0$ such that if $p \geqslant \frac{C}{n}$ we have*

$$\lim_{n\to\infty} \mathbb{P}[G((1+\varepsilon)sn, p) \xrightarrow{s} T(n, \Delta)] = 1.$$

Proof. Let $N = (1+\varepsilon)sn$ and let $0 < \delta < \frac{\varepsilon}{s}$ be fixed. For a given colouring of the edges of $G(N,p)$ with s colours, we denote by G_i the graph induced by all the edges in colour $i \in [s]$. By Theorem 1.4.3, we may assume that $e(G_i) \leqslant (\frac{1}{s} + \delta)e(G(sn,p))$ for all $i \in [s]$. Therefore

$$\sum_{i=1}^{r} e(G_i) \leqslant \sum_{i=1}^{r} \left(\frac{1}{s} + \delta\right) e(G(sn,p)) \leqslant (1 + s\delta)e(G(sn,p)) < e(G(N,p)),$$

which is a contradiction. □

We remark that Corollary 5.6.2 is sharp up to the value of C. However, for larger p (say $p \gg \frac{\log n}{n}$) we believe that the error term $O(\varepsilon n)$ in the size of the host graph can be improved.

For instance, for trees with a given ratio in the size of its colour classes, Corollary 5.6.2 if far from best possible. Indeed, Letzter [78] proved that, with high probability, $G((\frac{3}{2}+\varepsilon)n,p)$ is $(P_n,2)$-Ramsey provided $pn\to\infty$, where P_n denotes the path of length n. In a forthcoming work we will extend Letzter's result to arbitrary bounded degree trees.

A very interesting consequence of Corollary 5.6.2 is an upper bound for the multicolour *size Ramsey number* of bounded degree trees. Given a graph H and an integer $s\geqslant 2$, the s-colour size Ramsey number $\hat{r}_s(H)$ of H is the smallest integer m so that there exists a graph G with m edges such that every s-colouring of $E(G)$ yields a monochromatic copy of H. In the case of trees, it was conjecture in 1983 by Beck [16] that $\hat{r}_2(T)=O(Dn)$ for any fixed tree $T\in\mathcal{T}(n,D)$. This conjecture was settled by Friedman and Pippenger [51] proving that $\hat{r}_s(T)=O(n)$ for every $s\geqslant 2$ and every tree T with n vertices and bounded degree. We remark that this result also follows from Corollary 5.6.2 as the random graph $G(n,\frac{C}{n})$ has roughly $\frac{C}{n}\binom{n}{2}=O(n)$ edges.

Part II

Extremal Combinatorics on Words

Chapter 6

Introduction

The systematic study of combinatorics on words began in 1906 with the work of Thue [107] about the structure of *square-free* words. Although Thue's theorem appeared one year before Mantel's theorem, this field has received considerable less attention than extremal graph theory. However, after the appearance of Lothaire's book "Combinatorics on words" [81] in 1983, this topic has grown exponentially. We do not aim to describe in details this field of research. Our goal is rather to collect some of the recent results of what could be called *extremal combinatorics on words*.

A word $\boldsymbol{w}$ of length n is an ordered sequence $\boldsymbol{w} = (w_1 w_2 \dots w_n) \in \Sigma^n$, where Σ is a fixed size alphabet. The set of all possible finite words over Σ is denoted by $\Sigma^* = \bigcup_{n \geqslant 1} \Sigma^n$. We will usually represent a word in two equivalent ways, either as an order tuple or as a sequence. For example, the tuple (1011) and the sequence 1011 represent the same word. Moreover, we will usually denote words in bold letters in order to make a difference with its letters. Regarding the alphabet Σ, we are only interested in its cardinality. For $q = 2$ we will assume $\Sigma = \{0, 1\}$, and $\Sigma = [q]$ whenever $|\Sigma| = q \geqslant 3$.

In extremal combinatorics one usually deals with the problem of finding substructures maximising or minimising some parameters. So, as a first step, let us define what kind of substructures we will be interested in. There are at least three types of substructures that often appear in the literature (we will follow the nomenclature of Lothaire's book).

A *substring* or *factor* of a word $\boldsymbol{w} \in \Sigma^n$ is a sequence of consecutive characters in $\boldsymbol{w}$. A *subsequence* or *subword* of $\boldsymbol{w}$ is some word $\boldsymbol{v} \in \Sigma^\ell$ such that there are indices $1 \leqslant i_1 < \dots < i_\ell \leqslant n$ so that $w_{i_j} = v_j$ for each $j \in [\ell]$. In particular, every substring is a subword but not vice versa. Finally, a *pattern* P of $\boldsymbol{w}$ is a word $P = p_1 \dots p_m \in \mathcal{A}^m$, where $\mathcal{A} = \{a_1, \dots, a_t\}$ is an auxiliary alphabet, such that there are words $\boldsymbol{u}_1, \dots, \boldsymbol{u}_t$ over Σ and a substitution rule $\varphi(a_i) = \boldsymbol{u}_i$, for $i \in [t]$, such that $\varphi(P) = \varphi(p_1) \dots \varphi(p_m)$ yields a factor in $\boldsymbol{w}$. For instance, the word $\boldsymbol{w} = 011011011$ contains $\boldsymbol{v} = 011$ as a factor, $\boldsymbol{u} = 111111$ as a subword, and contains the pattern $P = xxx$ by replacing $x = 011$.

Let us now describe two problems that illustrate what we think as extremal combinatorics on words.

6.1 Longest common subsequence

Let $\boldsymbol{w} = (w_1 \dots w_n)$ and $\boldsymbol{u} = (u_1 \dots u_n)$ be two words of length n chosen uniformly at random from Σ^n. The *longest common subsequence (LCS) problem* asks for the maximum $\ell \leqslant n$ such that there exist sequences $1 \leqslant i_1 < \cdots < i_\ell \leqslant n$ and $1 \leqslant j_1 < \cdots < j_\ell \leqslant n$ such that $w_{i_s} = u_{j_s}$ for all $s \in [\ell]$. We write $\mathrm{LCS}_q(n)$ to denote the random variable which is equal to the length of the longest common subsequence between two random words chosen uniformly from Σ^n, where Σ is the alphabet on q symbols.

In 1975, Chvátal and Sankoff [34] proved that the expected value of $\frac{1}{n}\mathrm{LCS}_q(n)$ converges as $n \to \infty$. Indeed, it is easy to see that for every $n, m \in \mathbb{N}$ we have

$$\mathbb{E}[\mathrm{LCS}_q(n+m)] \geqslant \mathbb{E}[\mathrm{LCS}_q(n)] + \mathbb{E}[\mathrm{LCS}_q(m)]$$

and, therefore, by Fekete's supper additive lemma [48] we have that

$$\gamma_q := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{LCS}_q(n)] \text{ exists.}$$

The value γ_q is known as the Chvátal–Sankoff constant and is an open problem to determine the exact value of γ_q. In 2005, Kiwi, Loebl, and Matoušek [68] determined the asymptotic behaviour of γ_q showing that

$$\lim_{q\to\infty} \gamma_q \sqrt{q} = 2,$$

that is, for large q the expected length of the longest common subsequence is roughly $\frac{2}{\sqrt{q}}n$.

6.2 Twins in words

For a word $\boldsymbol{w} \in \Sigma^*$, let $\mathrm{LT}(\boldsymbol{w})$ be the maximum integer m so that there are two disjoint identical subwords of $\boldsymbol{w}$, each of length m. Such subwords are called *twins*. For $n \in \mathbb{N}$, we define

$$\mathrm{LT}(n, \Sigma) = \min\{\mathrm{LT}(\boldsymbol{w}) : \boldsymbol{w} \in \Sigma^n\}.$$

Thus, every word $\boldsymbol{w} \in \Sigma^n$ has twins of length $\mathrm{LT}(n, \Sigma)$. For $q \geqslant 2$, we observe that

$$\mathrm{LT}(n, [q]) \geqslant \left\lfloor \frac{n}{q+1} \right\rfloor.$$

Indeed, we start by splitting any word $\boldsymbol{w} \in [q]^n$ into $\ell = \lfloor \frac{n}{q+1} \rfloor$ substrings, each of length $q+1$. By the pigeonhole principle, there is at least one letter which has two occurrences a_i, a_i' the i-th substring, for each $i \in [\ell]$. Then, the words $\boldsymbol{v} = a_1 \dots a_\ell$ and $\boldsymbol{v}' = a_1' \dots a_\ell'$ are identical and disjoint subsequences of length $\lfloor \frac{n}{q+1} \rfloor$.

Axenovich, Person and Puzynina [12] improved this trivial lower bound showing that $\mathrm{LT}(n, [q]) \geqslant \frac{n}{q} - o(n)$, which is tight for binary alphabets up to the lower order error terms. Their proof is based in a *regularity lemma for words* which allows to split any large enough word into a bounded number of *quasi-random* substrings (see Chapter 8 for details). For larger alphabets the best lower bound is due to Bukh and Zhou [30], who showed that $\mathrm{LT}(n,[3]) \geqslant 0.34n - o(n)$ and

$$\mathrm{LT}(n,[q]) \geqslant \frac{n}{3^{4/3}q^{2/3}} - \left(\frac{q}{3}\right)^{1/3} \qquad \text{for } q \geqslant 3.$$

6.3 Our contributions

In this work, we study three problems in extremal combinatorics on words. In Chapter 7 we solved the *universality problem* for words and d-dimensional arrays. We ask for the minimum integer $f_d(q,k)$ so that there exists a d-array over an alphabet on q symbols so that there exists a d-dimensional array of order $f_d(q,k)$ containing a copy of every d-array of order k. In particular, $f_1(q,k)$ denotes the minimum length of a word over an alphabet on q symbols that contains, as a subsequence, a copy of every word of length k. We also study the probabilistic version of this problem. That is, for $k \in \mathbb{N}$, we ask for the smallest $n = n(k)$ so that, with high probability as $k \to \infty$, a random d-array of order n contains a copy of every d-array of order k.

In Chapter 8 we study two intimately related topics: *quasi-randomness* and *limit structures*. We study the notion of quasi-randomness of words that appeared in the work of Axenovich, Person, and Puzinina [12] and the work of Cooper [38]. We prove a result in the spirit of Chung–Graham–Wilson's theorem [33] for quasi-random graphs, giving a list of properties equivalent to those that a quasi-random word enjoys. In the second part of this chapter, we develop a theory of convergent word sequences in the vein of what has been done for other discrete structures, such as graphs [82] and permutations [64]. We prove that a sequence of binary words that converges in certain sense can be "represented" by a Lebesgue measurable function $f : [0,1] \to [0,1]$, and that every measurable function $f : [0,1] \to [0,1]$ can be approximated, in a appropriate sense, by a convergent sequence of binary words. Moreover, most of our results can be straightforwardly extended to larger alphabets.

Chapter 7

Universal arrays

Based on joint work with Daniel A. Quiroz and Nicolás Sanhueza-Matamala [91].

A *universal* mathematical structure is one which contains all possible substructures of a particular form. Famous examples of universal structures are De Bruijn sequences [40], which are periodic words that contain, exactly once, every possible word of a fixed size as a substring. Universal structures where perhaps first considered in a general sense by Rado [93], who studied the existence of universal graphs, hypergraphs and functions for various notions of containment.

The study of universal (finite) graphs has received particular attention, and for these the containment relation of choice has been that of induced subgraphs. Thus, a graph G is said to be *k-universal* if G contains every graph on k vertices as an induced subgraph. Two problems have been at the centre of the study of k-universal graphs. The first one is that of finding the minimum n such that there exists an n-vertex k-universal graph. In 1965, Moon [89] gave, through a simple counting argument, a lower bound of $2^{(k-1)/2}$ for that value of n. Recently, Alon [3] showed that this lower bound is asymptotically tight, essentially settling this 50-year-old problem. More so, in a later paper, Alon and Sherman [8] gave an asymptotically tight bound for the hypergraph generalisation of this problem. The second central problem in the study of k-universal graphs is the "random" analogue of the previous question, that is, finding the minimum n such that "almost every" n-vertex graph is k-universal. After works of Bollobás and Thomason [25], and Brightwell and Kohayakawa [29], Alon [3] has essentially settled this problem as well.

Finding a k-universal graph is equivalent to finding an adjacency matrix which "contains" the adjacency matrices of all k-vertex graphs. Here we are considering that an adjacency matrix M contains another matrix M', if we can obtain M' from M by iteratively applying the following operation: choose a value i and delete the i-th row and the i-th column. It is thus natural to consider square matrices together with the notion of containment given by the operation of choosing values i, j and deleting row i and column j, and its associated notion of universality. More generally, we shall consider the analogue of this notion of containment for "d-dimensional arrays" for all $d \geqslant 1$.

Given an alphabet Σ and positive integers $d, n_1, \dots, n_d$, a *d-dimensional array of size*

$n_1, \ldots, n_k$ *over* Σ is a collection of symbols $a_{i_1,i_2,\ldots,i_d} \in \Sigma$ indexed by the vectors $(i_1, i_2, \ldots, i_d) \in [n_1] \times [n_2] \times \cdots \times [n_d]$. With regard to the alphabet Σ we are only interested in its cardinality, and will assume $\Sigma = [q]$, whenever $|\Sigma| = q$. Thus Σ will usually be clear in the context and, for short, we will just talk about *d-arrays* of a certain size. A d-array of *order n* is a d-array of size $n, n, \ldots, n$, and note that 1-arrays of size n are just words of length n.

For general d-arrays and a corresponding notion of universality, we study the analogue of the two questions settled by Alon on the graph case (the "deterministic" and "random" questions). Whenever $d \geqslant 2$, we obtain asymptotically tight bounds for both questions (See Theorem 7.0.3 and Corollary 7.0.4, below) by extending a method used by Alon in the graph case. However, this technique does not seem (directly) to work when $d = 1$, that is, for the case of words. For this case we develop different tools which allow us to show tight bounds for both problems (See Theorems 7.0.1 and 7.0.2).

Let us first define the notion of containment we will consider for general d-arrays, which is a generalisation of the containment notion for matrices discussed above. For fixed d, let $A = (a_{i_1,i_2,\ldots,i_d})$ be a d-array of size $n_1, \ldots, n_d$. We define the *coordinate restriction operation* on A as follows. Choose some $j \in [d]$ and $\ell \in [n_j]$. Delete all the symbols whose j-th coordinate is ℓ, to obtain a d-array of size $n_1, n_2, \ldots, n_{j-1}, n_j - 1, n_{j+1}, \ldots, n_d$. We say a d-array A *contains* a d-array A' if we can obtain A' by iteratively applying coordinate restriction operations, and consider universal d-arrays under this containment notion.

For fixed $d, k \geqslant 1$, and a fixed alphabet Σ, we say a d-array over Σ is *k-universal* if it contains every d-array A on Σ of size $n_1, n_2, \ldots, n_d$, where $n_j \leqslant k$ for all $j \in [d]$. Note that if we want to show that a given d-array is k-universal, it is enough to show that it contains every d-array of order k. We let $f_d(q, k)$ be the minimum n such that there exists a k-universal d-array of order n over the q-symbol alphabet.

Our results in the case of words are the following.

Theorem 7.0.1. *Let $k \geqslant 1$ and $q \geqslant 2$ be integers. Then $f_1(q, k) = qk$.*

This result establishes the gap between the notions of subword and substring. While a minimal k-universal word has size qk, a De Bruijn sequence has size q^k. We also obtain the following "threshold" behaviour for randomly chosen words to be k-universal.

Theorem 7.0.2. *Let $q \geqslant 2$ be a fixed integer and $c_q := q + q/2 + q/3 + \ldots + 1$. Consider a uniformly chosen word $\boldsymbol{w}$ of length $n = n(k)$ over the q-symbol alphabet. Then for every $\varepsilon > 0$ we have*

$$\mathbb{P}[\, \boldsymbol{w} \textit{ is } k\textit{-universal} \,] \to \begin{cases} 0 & \textit{if } n \leqslant (1-\varepsilon)c_q k, \textit{ and} \\ 1 & \textit{if } n \geqslant (1+\varepsilon)c_q k, \end{cases}$$

where the limit is taken as $k \to \infty$.

In particular, for the 2-symbol alphabet, we have $f_1(2, k) = 2k$, while roughly $3k$ symbols are necessary and sufficient for a typical binary word of that length to be k-universal. This last statement answers a question of Biers-Ariel, Godbole and Kelley [20].

The following theorem and its corollary are our results for general d-arrays with $d \geqslant 2$.

Theorem 7.0.3. *Let $d, q \geqslant 2$ be fixed integers. For every $\varepsilon > 0$, a uniformly chosen d-array of order $n = (1+\varepsilon)\frac{k}{e}q^{\frac{k^{d-1}}{d}}$ over the q-symbol alphabet is k-universal with high probability as $k \to \infty$.*

Furthermore, a simple counting argument gives $f_d(q,k) \geqslant \frac{k}{e}q^{\frac{k^{d-1}}{d}}$ (see Section 7.2). Thus we obtain the following.

Corollary 7.0.4. *Let $d, q \geqslant 2$ be fixed integers. We have $f_d(q,k) = (1+o(1))\frac{k}{e}q^{\frac{k^{d-1}}{d}}$.*

We point out that the cases $d = 1$ and $d \geqslant 2$ behave in completely different manners. In the case $d = 1$, the case of words, the value of n in the random version is considerably larger than $f_1(q,k)$ (a similar scenario holds for the graph case [3]). In contrast, for d-arrays with $d \geqslant 2$ the order which is necessary for random d-arrays to be k-universal is asymptotically equal to $f_d(q,k)$.

7.1 Universal words

In this section we prove Theorems 7.0.1 and 7.0.2. We will use $\Sigma = [q]$ as the fixed q-symbol alphabet. We recall the standard notation used to work with words. Given a word $\boldsymbol{w}$ and an integer k, $\boldsymbol{w}^k$ is the k-fold concatenation of $\boldsymbol{w}$ with itself k times.

Proof of Theorem 7.0.1. No word $\boldsymbol{w}$ on at most $qk-1$ symbols can be k-universal: by the pigeonhole principle, one of the q symbols of Σ (which we can assume is 1) must appear less than k times in $\boldsymbol{w}$, but then the word 1^k is not contained in $\boldsymbol{w}$. On the other hand, the word $(12\cdots q)^k$ has length qk and is clearly k-universal. □

To prove Theorem 7.0.2, we will need a few tools. Given any word w on Σ^*, define $U_\Sigma(\boldsymbol{w})$ as the minimal prefix of $\boldsymbol{w}$ which contains all symbols of Σ if it exists, or $U_\Sigma(\boldsymbol{w}) = \boldsymbol{w}$ otherwise. Define $T_\Sigma(\boldsymbol{w})$ as $\boldsymbol{w}$ with the prefix $U_\Sigma(\boldsymbol{w})$ removed. Given a word $\boldsymbol{w}$, we can greedily decompose it in a unique way as $\boldsymbol{w} = \boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_\ell\boldsymbol{u}'$ such that for all $i \in [\ell]$, $\boldsymbol{u}_i = U_\Sigma(\boldsymbol{u}_i\boldsymbol{u}_{i+1}\cdots\boldsymbol{u}_\ell\boldsymbol{u}')$ and $T_\Sigma(\boldsymbol{u}_i\boldsymbol{u}_{i+1}\cdots\boldsymbol{u}_\ell\boldsymbol{u}') = \boldsymbol{u}_{i+1}\cdots\boldsymbol{u}_\ell\boldsymbol{u}'$, each $\boldsymbol{u}_i$ contains all the symbols of Σ and $\boldsymbol{u}'$ (possibly empty) does not contain all the symbols of Σ. We say $\boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_\ell\boldsymbol{u}'$ is the Σ-*universal decomposition of* $\boldsymbol{w}$ and we let $\nu_\Sigma(\boldsymbol{w}) = \ell$. We can use $\nu_\Sigma(\boldsymbol{w})$ to characterise k-universal words, as follows.

Proposition 7.1.1. *A word $\boldsymbol{w} \in \Sigma^*$ is k-universal if and only if $\nu_\Sigma(\boldsymbol{w}) \geqslant k$.*

Proof. Suppose $\boldsymbol{w}$ satisfies $\nu_\Sigma(\boldsymbol{w}) \geqslant k$. Then $\boldsymbol{w}$ has as a prefix a substring $\boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_k$ where each of the words $\boldsymbol{u}_i$ contains all of the symbols from Σ. Then any word $\boldsymbol{v} \in \Sigma^k$ can be found greedily as a subword in $\boldsymbol{w}$ by finding the i-th symbol of $\boldsymbol{v}$ inside the word $\boldsymbol{u}_i$.

In the other direction, suppose $\nu_\Sigma(\boldsymbol{w}) = k' < k$ and let $\boldsymbol{w} = \boldsymbol{u}_1\cdots\boldsymbol{u}_{k'}\boldsymbol{u}'$ be the Σ-universal decomposition of $\boldsymbol{w}$. Since each $\boldsymbol{u}_i$ is a minimal prefix of $\boldsymbol{u}_i\cdots\boldsymbol{u}_{k'}\boldsymbol{u}'$ that contains all the symbols of Σ, it must have the form $\boldsymbol{u}_i = \boldsymbol{v}_i\sigma_i$, where σ_i is a symbol in Σ and $\boldsymbol{v}_i$ does not use the symbol σ_i. Further, let $\sigma_{k'+1}$ be any symbol in Σ which does not appear in $\boldsymbol{u}'$ (which exists by definition). We claim that $\boldsymbol{w}$ does not contain the word $\boldsymbol{w}' = \sigma_1\sigma_2\cdots\sigma_{k'}\sigma_{k'+1}$. Since $k'+1 \leqslant k$, this readily implies that $\boldsymbol{w}$ is not k-universal.

To find a contradiction, suppose that $\boldsymbol{w}'$ is contained in $\boldsymbol{w}$. The first symbol of $\boldsymbol{w}'$ is σ_1, and the first time σ_1 appears in $\boldsymbol{w}$ is at the end of $\boldsymbol{u}_1$, thus the remaining symbols must appear after the end of $\boldsymbol{u}_1$. That means the word $\sigma_2\cdots\sigma_{k'}\sigma_{k'+1}$ is contained in $\boldsymbol{u}_2\cdots\boldsymbol{u}_{k'}\boldsymbol{u}$. Using the same reasoning, we see that for all $j \leqslant k'$, the j-th symbol of $\boldsymbol{w}'$ appears in $\boldsymbol{w}$ only after the last symbol of $\boldsymbol{u}_j$. Therefore, the last symbol of $\boldsymbol{w}'$, which is $\sigma_{k'+1}$, appears as a symbol in $\boldsymbol{u}'$, a contradiction. $\square$

We will need to estimate $\nu_\Sigma(\boldsymbol{w})$ for a uniformly chosen random word $\boldsymbol{w}$. We will appeal to the well-known "coupon-collector problem". Given a q-sized set Q and a sequence $X_1, X_2, \ldots$ of independent and uniformly chosen random variables $X_i \in Q$ for all $i \geqslant 1$, define the random variable T as the minimum integer such that $\{X_1, \ldots, X_T\} = Q$. It is known that T can be written as the sum of q independent geometric random variables $T = G_1 + \cdots + G_q$, where G_j has parameter j/q for each $j \in [q]$, and from this it is deduced that $\mathbb{E}[T] = c_q := q + q/2 + q/3 + \cdots + 1$.

Now we are ready for the proof of Theorem 7.0.2.

Proof of Theorem 7.0.2. Let Σ be the q-symbol alphabet. To estimate $\nu_\Sigma(\boldsymbol{w})$ of a random word $\boldsymbol{w}$, we will couple $\boldsymbol{w}$ with a word created from "coupon-collector" experiments, as follows. Define a random string $U \in \Sigma^*$ using the following process. Initially, let $U = \sigma_0$ be a word of length 1, where σ_0 is chosen uniformly from Σ. If U already has all the symbols of Σ, stop. Otherwise, choose uniformly and independently a symbol $\sigma \in \Sigma$ and update U by appending σ at the end. Clearly, the length $|U|$ of U distributes as in the "coupon-collector problem" and thus $\mathbb{E}[|U|] = c_q$. Given $k > 0$, let $U_1, \ldots, U_k$ be independent random strings, each of them distributed as U, and let $U^{(k)} = U_1U_2\cdots U_k$ be their concatenation. Crucially, we have $\nu_\Sigma(U^{(k)}) = k$, and each strict prefix $\boldsymbol{u}$ of $U^{(k)}$ satisfies $\nu_\Sigma(\boldsymbol{u}) < k$.

Given $k, n > 0$, we construct a (random) word $\boldsymbol{w}$ in Σ^n as follows: if $|U^{(k)}| \geqslant n$ then let $\boldsymbol{w}$ be the first n symbols of $U^{(k)}$; otherwise, construct $\boldsymbol{w}'$ from $U^{(k)}$ by appending $n - |U^{(k)}|$ fresh random symbols at the end of $U^{(k)}$. Note that each symbol of $\boldsymbol{w}$ is chosen independently and uniformly over the symbols of Σ, so $\boldsymbol{w}$ corresponds exactly to a word on Σ^n chosen uniformly at random. By construction it is clear that, for all $k, n > 0$,

$$\mathbb{P}[\,\boldsymbol{w} \text{ is } k\text{-universal}\,] = \mathbb{P}[\nu_\Sigma(\boldsymbol{w}) \geqslant k] = \mathbb{P}[|U^{(k)}| \leqslant n], \tag{7.1}$$

where the first equality is due to Proposition 7.1.1.

To estimate the last probability, note that $|U^{(k)}| = \sum_{i=1}^{\ell} |U_i|$ and recall that each of the $|U_i|$ has expectation equal to c_q. Thus, by the (Weak) Law of Large Numbers, we have that, for all $\varepsilon > 0$,

$$\mathbb{P}[(1-\varepsilon)c_q k \leqslant |U^{(k)}| \leqslant (1+\varepsilon)c_q k] \to 1, \tag{7.2}$$

whenever k goes to infinity. In particular, if $n \leqslant (1-\varepsilon)c_q k$ then $\mathbb{P}[|U^{(k)}| \leqslant n] \to 0$; and if $n \geqslant (1+\varepsilon)c_q k$ then $\mathbb{P}[|U^{(k)}| \leqslant n] \to 1$. By (7.1), the result follows. $\square$

7.2 Universal d-arrays

As before, let $\Sigma = [q]$ be the q-symbol alphabet. For integers $d, k \geqslant 1$, we write $\mathcal{A}_d(\Sigma, k)$ for the set of all d-arrays of order k over Σ. In this section, we prove Theorem 7.0.3 and stablish the lower bound for $f_d(q, k)$ which implies Corollary 7.0.4. To do so, we first need the following well-known estimates for binomial coefficients, most of which follow from Stirling's approximation. For all $n, k \geqslant 1$,

$$k! \geqslant \left(\frac{k}{e}\right)^k \quad \text{and} \quad \binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k. \tag{7.3}$$

Further, if $k \to \infty$ as $n \to \infty$, while $k = o(\sqrt{n})$,

$$\binom{n}{k} = (1 + o(1))\frac{1}{\sqrt{2\pi k}}\left(\frac{en}{k}\right)^k, \tag{7.4}$$

and if $k = \Omega(n)$ then

$$\log_2 \binom{n}{k} = (1 + o(1))H(\tfrac{k}{n})n, \tag{7.5}$$

where $H(x) = -x \log_2 x - (1 - x) \log_2(1 - x)$ is the binary entropy.

The lower bound for $f_d(q, k)$ when $d \geqslant 2$ is given by the following counting argument. Notice that there are q^{k^d} q-symbol d-arrays of order k. Therefore, a q-symbol d-array of order n must satisfy

$$\binom{n}{k}^d \geqslant q^{k^d}$$

in order to contain all arrays of order k. By (7.3) and the definition of $f_d(q, k)$ we obtain

$$\left(\frac{ef_d(q, k)}{k}\right)^{kd} \geqslant \left(\frac{f_d(q, k)}{k}\right)^d \geqslant q^{k^d}$$

and thus we have

$$f_d(q, k) \geqslant \frac{k}{e}q^{k^{d-1}/d}. \tag{7.6}$$

In light of Theorem 7.0.1, we know that for $d = 1$ the lower bound obtained here is considerably far from being tight. But we will show that it is asymptotically tight for all $d \geqslant 2$. In fact, it is asymptotically tight for the random version of the problem.

In order to prove Theorem 7.0.3 we follow an approach taken by Alon [3] in the study of universal graphs. Before diving into the proof let us first give a rough outline.

Given $k \in \mathbb{N}$ sufficiently large and $n = (1+o(1))\frac{k}{e}q^{k^{d-1}/d}$, let $A \in \mathcal{A}_d(\Sigma, n)$ be a uniformly chosen d-array of order n over Σ. For a fixed array $M \in \mathcal{A}_d(\Sigma, k)$, we consider the random variable X that counts the number of copies of M in A. Since there are q^{n^d} d-arrays of order n over Σ, it is enough to prove that $\mathbb{P}[X = 0] = o(q^{-k^d})$ and then use a union bound in order to conclude. However, it is not easy to prove this directly. Instead, we consider the random variable Y that counts the number of disjoint copies of M in A. It is clear that $Y = 0$ if and only if $X = 0$. Therefore, it is enough to estimate $\mathbb{P}[Y = 0]$. The random variable Y has the advantage that it is 1-Lipschitz, meaning that changing the value of one entry of the random array may change the value of Y in at most 1. Therefore, we may use (a known consequence of) Talagrand's inequality in order to upper bound $\mathbb{P}[Y = 0]$. However, to be able to use this tool, we need estimates on the expected number of pairs of copies of M in A which overlap in some entries, which amounts to studying the variance of X. Grasping the asymptotic behaviour of this variance turns out to be the most technical part of our proof.

Theorem 7.2.1 (Talagrand's inequality [9, Theorem 7.7.1])**.** *Let $\Omega = \prod_{i\in[r]} \Omega_i$ be a product probability space, with the product probability measure, and let $h : \Omega \to \mathbb{R}$ be a 1-Lipschitz function, that is, $|h(x) - h(y)| \leqslant 1$ when x and y differ in at most one coordinate. For $f\colon \mathbb{N} \to \mathbb{N}$, suppose that h is f-certifiable, that is, if $x \in \Omega$ is such that $h(x) \geqslant s$ then there exists a set $I \subseteq [r]$ of size at most $f(s)$ such that if a vector $y \in \Omega$ coincides with x on I, then $h(y) \geqslant s$. Then, for $Y(x) = h(x)$ and all b, t, we have*

$$\mathbb{P}[Y \leqslant b - t\sqrt{f(b)}] \cdot \mathbb{P}[Y \geqslant b] \leqslant e^{-t^2/4}.$$

Proof of Theorem 7.0.3. Let $d, q \geqslant 2$, $\varepsilon > 0$, $k \in \mathbb{N}$ (which we can assume to be large) and $n = (1+\varepsilon)\frac{k}{e}q^{\frac{k^{d-1}}{d}}$. Let $M \in \mathcal{A}_d(\Sigma, k)$ be a fixed d-array of order k over the q-symbol alphabet Σ, and let A be a uniformly chosen array from $\mathcal{A}_d(\Sigma, n)$. Our aim is to find a good upper bound on the probability that A does not contain M, i.e., one allowing us to use a union bound to prove the result.

Let $\mathcal{T}$ denote the collection of subsets of $[n]^d$ of the form $T = T_1 \times \cdots \times T_d$, where $|T_i| = k$ for each $1 \leqslant i \leqslant d$. Given $T \in \mathcal{T}$, let $T(A)$ be the subarray of A with entries $a_{i_1,\ldots,i_d}$ with $i_1 \in T_1, \ldots, i_d \in T_d$. Let X_T be the indicator function of the event that T induces a copy of M, and let $X = \sum_{T\in\mathcal{T}} X_T$ be the number of copies of M in A. Since for every $T \in \mathcal{T}$ we have $\mathbb{E}[X_T] = q^{-k^d}$, by linearity of the expectation we have

$$\mu := \mathbb{E}[X] = \binom{n}{k}^d q^{-k^d} \geqslant 16k^{2d}\log q, \tag{7.7}$$

where the last inequality follows from the choice of n, the assumption that k is large, and (7.4).

It will be crucial to show that we have

$$\mathrm{Var}(X) \leqslant (1 + o(1))\mu. \tag{7.8}$$

To this end, we investigate (the expectation of) the random variable

$$Z := \sum_{T,T'} X_T X_{T'},$$

where the sum ranges over the pairs of distinct $T, T' \in \mathcal{T}$ which intersect in at least one cell. For $i_1, \ldots, i_d \in [k]$, we write

$$\Delta_{i_1,\ldots,i_d} := \sum_{T,T'\in\mathcal{T}_{i_1,\ldots,i_d}} \mathbb{E}[X_T X_{T'}],$$

where $\mathcal{T}_{i_1,\ldots,i_d}$ denotes the collection of indices $T, T' \in \mathcal{T}$ such that $|T_j \cap T'_j| = i_j$ for all $j \in [d]$, i.e., T and T' intersect on exactly i_j indices on the j-th coordinate. Therefore, if $\Delta := \mathbb{E}[Z]$, then we have

$$\Delta := \sum_{T,T'} \mathbb{E}[X_T X_{T'}] = \sum_{(i_1,\ldots,i_d)\neq(k,\ldots,k)} \Delta_{i_1,\ldots,i_d}. \tag{7.9}$$

Given $i \in [k]$, we define

$$\Lambda_i = \binom{n}{k}\binom{k}{i}\binom{n-k}{k-i} \quad \text{and} \quad L_d(i) = q^{\frac{i^d}{d}\left(1-(k/i)^{d-1}\right)} \frac{1}{(k-i)!}\binom{k}{i}\left(\frac{(1+\varepsilon)k}{e}\right)^{k-i}.$$

In order to prove (7.8) we will use the following two claims.

Claim 7.2.2. *For all $i_1, \ldots, i_d \in [k]$ we have*

$$\frac{\Delta_{i_1,\ldots,i_d}}{\mu} \leqslant \prod_{j\in[d]} L_d(i_j).$$

Proof of Claim 7.2.2. Let $i_1, \ldots, i_k$ be given. First, note that the total number of pairs T, T' which intersect on i_j entries on the j-th coordinate is exactly equal to $\prod_{j\in[d]} \Lambda_{i_j}$. Moreover, the union of two subarrays T and T' of this type together span exactly $2k^d - i_1 \cdots i_d$ cells. Then $X_T X_{T'} = 1$ holds if and only if in each one of those cells the correct symbol is attained, which implies

$$\Delta_{i_1,\ldots,i_d} \leqslant q^{-(2k^d - i_1\cdots i_d)} \prod_{j\in[d]} \Lambda_{i_j}.$$

By the AM-GM inequality we have $i_1 \cdots i_d \leqslant \left(\frac{1}{d}\sum_{j=1}^d i_j\right)^d$. Thus we have

$$\frac{\Delta_{i_1,\ldots,i_d}}{\mu} \leqslant \frac{q^{-(2k^d - i_1\cdots i_d)} \prod_{j\in[d]} \Lambda_{i_j}}{\binom{n}{k}^d q^{-k^d}} \leqslant q^{-k^d + (\frac{1}{d}\sum_{j\in[d]} i_j)^d} \prod_{j\in[d]} \binom{k}{i_j}\binom{n}{k-i_j}.$$

Using Jensen's inequality (with the convex function $x \mapsto x^d$) we further have $\left(\frac{1}{d}\sum_{j=1}^d i_j\right)^d \leqslant \frac{1}{d}(\sum_{j=1}^d i_j^d)$. Using that $(k-i)!\binom{n}{k-i} \leqslant (n-i)^{k-i}$ and replacing $n = (1+\varepsilon)\frac{k}{e}q^{k^{d-1}/d}$ we have

$$\begin{aligned}
\frac{\Delta_{i_1,\ldots,i_d}}{\mu} &\leqslant q^{-k^d + \frac{1}{d}\sum_{j\in[d]} i_j^d} \prod_{j\in[d]} \binom{k}{i_j} \frac{n^{k-i_j}}{(k-i_j)!} \\
&\leqslant \prod_{j\in[d]} q^{\frac{1}{d}(i_j^d - k^d)} \frac{1}{(k-i_j)!}\binom{k}{i_j}\left(\frac{(1+\varepsilon)k}{e}\right)^{k-i_j} q^{\frac{1}{d}k^{d-1}(k-i_j)} \\
&= \prod_{j\in[d]} L_d(i_j),
\end{aligned}$$

as desired

Claim 7.2.3. *If $1 \leqslant i \leqslant k-1$, then $L_d(i) = o(k^{-d})$.*

Proof of Claim 7.2.3. Without loss of generality we may assume that $8\varepsilon \leqslant \log q$, as otherwise we can restrict to a smaller array. Setting $i = k - j$ and by the Bernoulli inequality $(1 + \frac{j}{k-j})^{d-1} \geqslant 1 + j\frac{d-1}{k-j}$ we have

$$\begin{aligned} L_d(k-j) &= q^{\frac{(k-j)^d}{d}(1-(1+j/(k-j))^{d-1})} \frac{1}{j!}\binom{k}{j}\left(\frac{(1+\varepsilon)k}{e}\right)^j \\ &\leqslant q^{-\frac{d-1}{d}j(k-j)^{d-1}} \frac{1}{j!}\binom{k}{j}\left(\frac{(1+\varepsilon)k}{e}\right)^j. \end{aligned}$$

We now split into two cases. Assume first that $j \geqslant (1-\beta)k$, where $\beta \in (0,1)$ is small enough so that $H(\beta) \leqslant \frac{\log q}{32}$. This choice of j allows us to use (7.5) to obtain

$$\log_2 \binom{k}{j} = (1+o(1))kH(\tfrac{j}{k}) \leqslant 2kH(\tfrac{j}{k}),$$

and, since $H(x)$ is decreasing in $(\frac{1}{2}, 1)$, it also guarantees

$$H(\tfrac{j}{k}) \leqslant H(1-\beta) = H(\beta).$$

These two observations and (7.3) give us

$$\begin{aligned} \tfrac{1}{j}\log L_d(k-j) &\leqslant -\tfrac{d-1}{d}(k-j)^{d-1}\log q + \tfrac{2k}{j\log_2 e}H(\tfrac{j}{k}) + 2\log(\tfrac{k}{j}) + \log(1+\varepsilon) \\ &\leqslant -\tfrac{1}{2}(k-j)^{d-1}\log q + 4H(\beta) + 2\log(\tfrac{k}{j}) + \varepsilon. \end{aligned}$$

We use the fact that $\log(\frac{k}{j}) = \log(1 + \frac{k-j}{j}) \leqslant \frac{k-j}{j}$, together with our choices of ε and β to obtain

$$\frac{\log L_d(k-j)}{j(k-j)} \leqslant -\frac{1}{2}\log q + \frac{4H(\beta)}{k-j} + \frac{4}{k} + \frac{\varepsilon}{k-j} \leqslant -\frac{1}{4}\log q.$$

Therefore, for $j \geqslant (1-\beta)k$ we have

$$L_d(k-j) \leqslant e^{-j(k-j)\log q/4}.$$

We are left to consider the case $j \leqslant (1-\beta)k$. Similarly, by (7.3) we have $\log j! \geqslant j\log j - j$ and $\binom{k}{j} \leqslant \left(\frac{ek}{j}\right)^j$, and then

$$\frac{\log L_d(k-j)}{j} \leqslant -\frac{1}{2}(k-j)^{d-1} - \log j + 1 + 2\log(\tfrac{k}{j}) + \varepsilon \leqslant -\frac{\beta^{d-1}k}{2} + 1 + 2\log k + \varepsilon.$$

Therefore, in this range we have

$$L_d(k-j) \leqslant e^{-\beta^{d-1}jk/4}.$$

The claim follows.

Since the sum in (7.9) is over all the $k^d - 1$ tuples $(i_1, \ldots, i_d)$ in $[k]^d$ distinct from $(k, \ldots, k)$, then Claim 7.2.2 and Claim 7.2.3 together imply that

$$\Delta = o(\mu). \tag{7.10}$$

Now, since X is a sum of zero-one random variables, we have

$$\operatorname{Var}(X) \leqslant \mathbb{E}[X] + \sum_{T,T' \in \mathcal{T}} \operatorname{Cov}(X_T, X_{T'}).$$

In the sum we only need to consider the pairs $T, T' \in \mathcal{T}$ with non-trivial intersection (otherwise the variables $X_T, X_{T'}$ are independent and thus their covariance is zero). Further, we have $\operatorname{Cov}(X_T, X_{T'}) \leqslant \mathbb{E}(X_T X_{T'})$. Therefore, by (7.10) we have

$$\operatorname{Var}(X) \leqslant \mu + \Delta = (1 + o(1))\mu,$$

and so we have finally proved (7.8).

By Chebyshev's inequality, and equations (7.7) and (7.8) we have

$$\mathbb{P}[|X - \mu| \geqslant \tfrac{1}{4}\mu] \leqslant \frac{16 \operatorname{Var}(X)}{\mu^2} \leqslant \frac{32}{\mu} \to 0.$$

Therefore, $X \geqslant \frac{3}{4}\mu$ with probability at least $\frac{3}{4}$. Likewise, by Markov's inequality and (7.10) we have

$$\mathbb{P}[Z \geqslant \tfrac{1}{5}\mu] \leqslant \frac{5\mathbb{E}[Z]}{\mu} = \frac{5\Delta}{\mu} \to 0,$$

and therefore $Z \leqslant \frac{1}{4}\mu$ with probability at least $\frac{3}{4}$. In particular, both events hold at the same time with probability at least $\frac{1}{2}$.

Let Y denote the random variable that counts the maximum number of disjoint copies of M in A. Since $X \geqslant \frac{3}{4}\mu$ and $Z \leqslant \frac{1}{4}\mu$ hold with probability at least $\frac{1}{2}$, then, by conditioning on this event, we deduce that

$$\mathbb{P}[Y \geqslant \tfrac{1}{2}\mu] \geqslant \frac{1}{2}. \tag{7.11}$$

Notice also that $X = 0$ if and only if $Y = 0$.

We are now ready to use Talagrand's inequality to finish the proof. Note that $h(A) := Y$ is 1-Lipschitz, since by switching the value of one entry one can add or remove at most 1 copy of M (the one using that entry). Moreover, $h(A)$ is f-certifiable for $f(s) = sk^d$. Using $b = \frac{1}{2}\mu$ and $t = k^{-d/2}\sqrt{\frac{1}{2}\mu}$, Talagrand's inequality and (7.11) give us

$$\mathbb{P}[X = 0] = \mathbb{P}[Y = 0] \leqslant 2e^{-\mu k^{-d}/8}.$$

Finally, we use a union bound over all the possible choices of $M \in \mathcal{A}_d(\Sigma, k)$, to deduce that the probability that A is not k-universal is at most

$$2q^{k^d} e^{-\mu k^{-d}/8} \leqslant 2q^{k^d} e^{-2k^d \log q} = 2q^{-k^d} \to 0$$

where the inequality comes from (7.7). The result follows. □

Remark 7.2.4. *The constant error term ε in Theorem 7.0.3 can be improved to a term $\Omega(\frac{\log k}{k})$. This can be seen by checking that replacing $\varepsilon = C\frac{\log k}{k}$ (with C being a large constant) is enough for (7.7) to hold. This does not change the rest of the calculations.*

Chapter 8

Quasi-random words and limits of convergent word sequences

Based on joint work with Hiệp Hàn and Marcos Kiwi [60].

8.1 Introduction

Roughly speaking, quasi-random structures are deterministic objects which share many characteristic properties of their random counterparts. Formalizing this concept has turned out to be tremendously fruitful in several areas, among others, number theory, graph theory, extremal combinatorics, the design of algorithms and complexity theory. This often follows from the fact that if an object is quasi-random, then it immediately enjoys many other properties satisfied by its random counterpart.

Seminal work on quasi-randomness concerned graphs [33, 96, 106]. Subsequently, other combinatorial objects were considered, which include subsets of $\mathbb{Z}_n$ [32, 57], hypergraphs [1, 31, 58, 108], finite groups [59], and permutations [38]. Curiously, in the rich history of quasi-randomness, *words*, i.e., sequences of letters from a finite alphabet, one of the most basic combinatorial object with many applications, do not seem to have been explicitly investigated. We overcome this apparent neglect, put forth a notion of quasi-random words and show it is equivalent to several other properties.

In contrast to the classical topic of quasi-randomness, the research of limits for discrete structures was launched rather recently by Chayes, Lovász, Sós, Szegedy and Vesztergombi [26, 84], and has become a very active topic of research since. Central to the area is the notion of convergent graph sequences $(G_n)_{n\to\infty}$, i.e., sequences of graphs which, roughly speaking, become more and more "similar" as $|V(G_n)|$ grows. For convergent graph sequences, Lovász and Szegedy [84] show the existence of natural limit objects, called *graphons*, endow the space of these structures with a metric and establish the equivalence of their notion of convergence and convergence on such a metric. Among many other consequences, it follows that quasi-random graph sequences, with edge density $p + o(1)$, converge to the constant p graphon.

In this paper, we continue the lines of previously mentioned investigations and study quasi-randomness for words and limits of convergent word sequences. Not only in the literature of quasi-randomness but also in the one concerning limits of discrete structures, explicit investigation of this fundamental object has not been considered so far.

8.1.1 Quasi-random words

Concerning quasi-randomness for words, our central notion is that of uniform distribution of letters over intervals. Specifically, a word $\boldsymbol{w} = (w_1 \dots w_n) \in \{0,1\}^n$ is called *(d,ε)-uniform* if for every interval $I \subseteq [n]$ we have

$$\sum_{i \in I} w_i = |\{i \in I \colon w_i = 1\}| = d|I| \pm \varepsilon n. \tag{8.1}$$

We say that $\boldsymbol{w}$ is *ε-uniform* if $\boldsymbol{w}$ is (d,ε)-uniform for some d. Thus, uniformity states that up to an error term of εn the number of 1-entries of $\boldsymbol{w}$ in each interval I is roughly $d|I|$, a property which binomial random words with parameter d satisfy with high probability. This notion of uniformity has been studied by Axenovich, Person and Puzynina in [12], where a regularity lemma for words was established and applied to the problem of finding twins in words. In a different context, it has been studied by Cooper [38] who gave a list of equivalent properties. A word $(w_1 \dots w_n) \in \{0,1\}^n$ can also be seen as the set $W = \{i \colon w_i = 1\} \subseteq \mathbb{Z}_n$ and from this point of view uniformity should be compared to the classical notion of quasi-randomness of subsets of $\mathbb{Z}_n$, studied by Chung and Graham in [32] and extended to the notion of U_k-uniformity by Gowers in [57]. With respect to this line of research we note that our notion of uniformity is strictly weaker than all of the ones studied in [32, 57]. Indeed, the weakest of them concerns U_2-uniformity and may be rephrased as follows: $W \subseteq \mathbb{Z}_n$ has U_2-norm at most $\varepsilon > 0$ if for all $A \subseteq \mathbb{Z}_n$ and all but εn elements $x \in \mathbb{Z}$ we have $|W \cap (A+x)| = |W|\frac{|A|}{n} \pm \varepsilon n$ where $A + x = \{a + x \colon a \in A\}$. Thus, e.g., the word $0101\dots01$ is uniform in our sense but its corresponding set does not have small U_2-norm.

Analogous to the graph case there is a counting property related to uniformity. Given a word $\boldsymbol{w} = (w_1 \dots w_n)$ and a set of indices $I = \{i_1, \dots, i_\ell\} \subseteq [n]$, where $i_1 < i_2 < \dots < i_\ell$, let $\mathrm{sub}(I, \boldsymbol{w})$ be the length ℓ subsequence $\boldsymbol{u} = (u_1 \dots u_\ell)$ of $\boldsymbol{w}$ such that $u_j = w_{i_j}$. We show that uniformity implies adequate subsequence count, i.e., for any fixed $\boldsymbol{u}$ the number of subsequences equal to $\boldsymbol{u}$ in a large uniform word $\boldsymbol{w}$, denoted by $\binom{\boldsymbol{w}}{\boldsymbol{u}}$, is roughly as expected from a random word with same density of 1-entries as $\boldsymbol{w}$. It is then natural to ask whether the converse also holds and one of our main results concerning quasi-random words states that uniformity is indeed already enforced by counting of subsequences of length three. Let $\|\boldsymbol{w}\|_1 = \sum_{i \in [n]} w_i$ denote the number of 1-entries in $\boldsymbol{w}$, then our result reads as follows.

Theorem 8.1.1. *For every $\varepsilon > 0$, $d \in [0,1]$, and $\ell \in \mathbb{N}$, there is an n_0 such that for all $n > n_0$ the following holds.*

- *If $\boldsymbol{w} \in \{0,1\}^n$ is (d,ε)-uniform, then for each $\boldsymbol{u} \in \{0,1\}^\ell$*

$$\binom{\boldsymbol{w}}{\boldsymbol{u}} = d^{\|\boldsymbol{u}\|_1}(1-d)^{\ell - \|\boldsymbol{u}\|_1}\binom{n}{\ell} \pm 5\varepsilon n^\ell.$$

- *Conversely, if $\boldsymbol{w} \in \{0,1\}^n$ is such that for all $\boldsymbol{u} \in \{0,1\}^3$ we have*

$$\binom{\boldsymbol{w}}{\boldsymbol{u}} = d^{\|\boldsymbol{u}\|_1}(1-d)^{3-\|\boldsymbol{u}\|_1}\binom{n}{3} \pm \varepsilon n^3,$$

then $\boldsymbol{w}$ is $(d, 18\varepsilon^{1/3})$-uniform.

Note that in the second part of the theorem the density of 1-entries is implicitly given. This is because $\binom{\boldsymbol{w}}{(111)} = \binom{\|\boldsymbol{w}\|_1}{3}$, and therefore the condition $\binom{\boldsymbol{w}}{(111)} \approx d^3\binom{n}{3}$ implies that $\|\boldsymbol{w}\|_1 \approx dn$. We also note that length three subsequences in the theorem cannot be replaced by length two subsequences and in this sense the result is best possible. Indeed, the word $(0\dots01\dots10\dots0)$ consisting of $(1-d)\frac{n}{2}$ zeroes followed by dn ones followed by $(1-d)\frac{n}{2}$ zeroes contains the "right" number of every length two subsequences without being uniform.

We also study a property called Equidistribution and show that it is equivalent to uniformity. Together with Theorem 8.1.1 (and its direct consequences) and a result from Cooper [38, Theorem 2.2] this yields a list of equivalent properties stated in Theorem 8.1.2. To state the result let $\boldsymbol{w}[j]$ denote the j-th letter of the word $\boldsymbol{w}$. Furthermore, by the Cayley digraph $\Gamma = \Gamma(\boldsymbol{w})$ of a word $\boldsymbol{w} = (w_1 \dots w_n)$ we mean the graph on the vertex set $\mathbb{Z}_n$ in which i and j form an edge if and only if $w_{i-j \pmod n} = 1$. Given a word $\boldsymbol{u} \in \{0,1\}^{\ell+1}$, a sequence of vertices $(v_1 \dots v_{\ell+1})$ is an increasing $\boldsymbol{u}$-path in $\Gamma = \Gamma(\boldsymbol{w})$ if the numbers $i_1, \dots, i_\ell \in [n]$ defined by $v_{k+1} = v_k + i_k \pmod n$ satisfy $i_1 < \dots < i_\ell$ and for each $k \in [\ell]$ the pair $v_k v_{k+1}$ is an edge in Γ if $u_k = w_{i_k} = 1$ and a non-edge if $u_k = w_{i_k} = 0$.

Henceforth, we define the Lipschitz norm of a function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ by

$$\|f\|_{\mathrm{Lip}} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{\min\{1 - |x-y|, |x-y|\}}.$$

Theorem 8.1.2. *For a sequence $(\boldsymbol{w}_n)_{n\to\infty}$ of words $\boldsymbol{w}_n \in \{0,1\}^n$ such that $\|\boldsymbol{w}_n\|_1 = dn + o(n)$ for some $d \in [0,1]$, the following are equivalent:*

- *(Uniformity) $(\boldsymbol{w}_n)_{n\to\infty}$ is $(d, o(1))$-uniform.*
- *(Counting) For all $\ell \in \mathbb{N}$ and all $\boldsymbol{u} \in \{0,1\}^\ell$ we have*

$$\binom{\boldsymbol{w}_n}{\boldsymbol{u}} = d^{\|\boldsymbol{u}\|_1}(1-d)^{\ell-\|\boldsymbol{u}\|_1}\binom{n}{\ell} + o(n^\ell).$$

- *(Minimizer) For all $\boldsymbol{u} \in \{0,1\}^3$ we have*

$$\binom{\boldsymbol{w}_n}{\boldsymbol{u}} = d^{\|\boldsymbol{u}\|_1}(1-d)^{3-\|\boldsymbol{u}\|_1}\binom{n}{3} + o(n^3).$$

- *(Exponential sums) For any fixed $\alpha > 0$ and for all $k \in [n-1]$ we have*

$$\tfrac{1}{n}\textstyle\sum_{j\in[n]} \boldsymbol{w}_n[j] \cdot \exp\left(\frac{2\pi i}{n} kj\right) = o(1)|k|^\alpha.$$

- *(Equidistribution) For every Lipschitz function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$*

$$\tfrac{1}{n}\textstyle\sum_{j\in[n]} \boldsymbol{w}_n[j] \cdot f(\frac{j}{n}) = d \int_{\mathbb{R}/\mathbb{Z}} f + o(1)\|f\|_{\mathrm{Lip}}.$$

- *(Cayley graph) For all $\boldsymbol{u} \in \{0,1\}^3$ the number of increasing $\boldsymbol{u}$-paths in $\Gamma(\boldsymbol{w}_n)$ is*
$$d^{\|\boldsymbol{u}\|_1}(1-d)^{3-\|\boldsymbol{u}\|_1} n\binom{n}{3} + o(n^4).$$

We will say that a word sequence is *quasi-random* if it satisfies one of (hence all) the properties of Theorem 8.1.2.

8.1.2 Convergent word sequences and word limits

Over the last two decades it has been recognized that quasi-randomness and limits of discrete structures are intimately related subjects. Being interesting in their own right, limit theories have also unveiled many connections between various branches of mathematics and theoretical computer science. Thus, as a natural continuation of the investigation on quasi-randomness, we study convergent word sequences and their limits, a topic which, to the best of our knowledge, has only been briefly mentioned by Szegedy [104].

The notion of convergence we consider is specified in terms of convergence of subsequence densities. Given $\boldsymbol{w} \in \{0,1\}^n$ and $\boldsymbol{u} \in \{0,1\}^\ell$, let $t(\boldsymbol{u}, \boldsymbol{w})$ be the density of occurrences of $\boldsymbol{u}$ in $\boldsymbol{w}$, i.e.,
$$t(\boldsymbol{u}, \boldsymbol{w}) = \binom{\boldsymbol{w}}{\boldsymbol{u}}\binom{n}{\ell}^{-1}.$$
Alternatively, if we define $\mathrm{sub}(\ell, \boldsymbol{w}) := \mathrm{sub}(I, \boldsymbol{w})$ for I uniformly chosen among all subsets of $[n]$ of size ℓ, then $t(\boldsymbol{u}, \boldsymbol{w}) = \mathbb{P}[\mathrm{sub}(\ell, \boldsymbol{w}) = \boldsymbol{u}]$.

A sequence of words $(\boldsymbol{w}_n)_{n\to\infty}$ is called *convergent* if for every finite word $\boldsymbol{u}$ the sequence $\big(t(\boldsymbol{u}, \boldsymbol{w}_n)\big)_{n\to\infty}$ converges. In what follows, we will only consider sequences of words such that the length of the words tend to infinity. This, however, is not much of a restriction since convergent word sequences with bounded lengths must be constant eventually and limits considerations for these sequences are simple.[1]

We show that convergent word sequences have natural limit objects, which turn out to be Lebesgue measurable functions of the form $f : [0,1] \to [0,1]$. Formally, write $f^1 = f$ and $f^0 = 1 - f$ for a function $f : [0,1] \to [0,1]$ and for a word $\boldsymbol{u} \in \{0,1\}^\ell$ define
$$t(\boldsymbol{u}, f) = \ell! \int_{0\leqslant x_1 < \cdots < x_\ell \leqslant 1} \prod_{i\in[\ell]} f^{u_i}(x_i)\, \mathrm{d}x_1 \ldots \mathrm{d}x_\ell. \tag{8.2}$$
We say that $(\boldsymbol{w}_n)_{n\to\infty}$ *converges to* f and that f is the *limit* of $(\boldsymbol{w}_n)_{n\to\infty}$, if for every word $\boldsymbol{u}$ we have
$$\lim_{n\to\infty} t(\boldsymbol{u}, \boldsymbol{w}_n) = t(\boldsymbol{u}, f).$$
In particular, $(\boldsymbol{w}_n)_{n\to\infty}$ is convergent in this case. Furthermore, let $\mathcal{W}$ be the set of all Lebesgue measurable functions of the form $f : [0,1] \to [0,1]$ in which, moreover, functions are identified when they are equal almost everywhere. We show that each convergent word sequence converges to a unique $f \in \mathcal{W}$ and that, conversely, for each $f \in \mathcal{W}$ there is a word sequence which converges to f.

[1] Word sequences with bounded lengths contain a subsequence of infinite length which is constant and due to convergence all members of the original sequence must agree with this constant eventually.

Theorem 8.1.3 (Limits of convergent word sequences)**.**

- *For each convergent word sequence $(\boldsymbol{w}_n)_{n\to\infty}$ there is an $f \in \mathcal{W}$ such that $(\boldsymbol{w}_n)_{n\to\infty}$ converges to f. Moreover, if $(\boldsymbol{w}_n)_{n\to\infty}$ converges to g then f and g are equal almost everywhere.*
- *Conversely, for every $f \in \mathcal{W}$ there is a word sequence $(\boldsymbol{w}_n)_{n\to\infty}$ which converges to f.*

Theorem 8.1.3 can be phrased in topological terms as follows. Given a word $\boldsymbol{u}$, one can think of $t(\boldsymbol{u}, \cdot)$ as a function from $\mathcal{W}$ to $[0, 1]$. Then, endow $\mathcal{W}$ with the initial topology with respect to the family of maps $t(\boldsymbol{u}, \cdot)$, with $\boldsymbol{u} \in \{0, 1\}^{\ell}$ and $\ell \in \mathbb{N}$, that is, the smallest topology that makes all these maps continuous. We show that this topology is actually metrisable and, moreover, compact (thereby proving Theorem 8.1.3).

The overall approach we follow is in line with what has been done for graphons [84] and permutons [64]. Nevertheless, there are important technical differences, specially concerning the (in our case, more direct) proofs of the equivalence between distinct notions of convergence which avoid compactness arguments. Instead, we rely on Bernstein polynomials and their properties as used in the (constructive) proof the Stone–Weierstrass approximation theorem.

In contrast with other technically more involved limit theories, say the ones concerning graph sequences [84] and permutation sequences [64], the simplicity of the underlying combinatorial objects we consider (words) yields concise arguments, elegant proofs, simple limit objects, and requires the introduction of far fewer concepts. Yet despite the technically comparatively simpler theory, many interesting aspects common to other structures and some specific to words appear in our investigation. As an illustration, we work out the implications for testing of the class of so-called *hereditary* word properties and address the question concerning *finite forcibility* for words, i.e., which word limits are completely determined by a finite number of prescribed subsequence densities.

8.1.3 Testing hereditary word properties

The concept of self-testing/correcting programs was introduced by Blum et al. [22, 23] and greatly expanded by the concept of graph property testing proposed by Goldreich, Goldwasser and Ron [56] (for an in depth coverage of the property testing paradigm, the reader is referred to the book by Goldreich [55]). An insightful connection between testable graph properties and regularity was established by Alon and Shapira [6] and further refined in [4, 7]. It was then observed that similar and related results can be obtained via limit theories (for the case of testing graph properties, the reader is referred to [85], and for the case of (weakly) testing permutation properties, to [65]). Thus, it is not surprising that analogue results can be established for word properties. On the other hand, it is noteworthy that such consequences can be obtained very concisely and elegantly.

We next state our main result concerning testing word properties. Formally, for $\boldsymbol{u}, \boldsymbol{w} \in \{0, 1\}^n$ let $d_1(\boldsymbol{w}, \boldsymbol{u}) = \frac{1}{n}\sum_{i\in[n]} |w_i - u_i|$. A *word property* is simply a collection of words.

A word property $\mathcal{P}$ is said to be *testable* if there is another word property $\mathcal{P}'$ (called *test property for* $\mathcal{P}$) satisfying the following conditions:

(Completeness) For every $\boldsymbol{w} \in \mathcal{P}$ of length n and every $\ell \in [n]$, $\mathbb{P}[\mathrm{sub}(\ell, \boldsymbol{w}) \in \mathcal{P}'] \geqslant \frac{2}{3}$.

(Soundness) For every $\varepsilon > 0$ there is an $\ell(\varepsilon) \geqslant 1$ such that if $\boldsymbol{w} \in \{0,1\}^n$ with $d_1(\boldsymbol{w}, \mathcal{P}) = \min_{\boldsymbol{u} \in \mathcal{P} \cap \{0,1\}^n} d_1(\boldsymbol{w}, \boldsymbol{u}) \geqslant \varepsilon$, then $\mathbb{P}[\mathrm{sub}(\ell, \boldsymbol{w}) \in \mathcal{P}'] \leqslant \frac{1}{3}$ for all $\ell(\varepsilon) \leqslant \ell \leqslant n$.

Variants of the notion of testability can be considered. However, the one stated is sort of the most restrictive. On the other hand, the notion can be strengthened by replacing the $2/3$ in the completeness part by $1 - \varepsilon$ and $1/3$ in the soundness part by ε. The notion can be weakened letting the test property $\mathcal{P}'$ depend on ε. These variants do not change the concept of testability.

A word property $\mathcal{P}$ is called *hereditary* if for each $\boldsymbol{w} \in \mathcal{P}$, every subsequence $\boldsymbol{u}$ of $\boldsymbol{w}$ also belongs to $\mathcal{P}$.

Theorem 8.1.4. *Every hereditary word property is testable.*

Since the notion of testability given above is very restrictive (it consists in sampling uniformly a constant number of characters from the word being tested) it straightforwardly yields efficient (polynomial time) testing procedures.

Examples of hereditary properties are: (1) the collection $\mathcal{P}_{\mathcal{F}}$ of words that do not contain as subsequence any word in $\mathcal{F}$ where $\mathcal{F}$ is a family of words ($\mathcal{F}$ might even be infinite), and (2) for given $\mathcal{P}_1, ..., \mathcal{P}_k$ hereditary word properties, the collection $\mathcal{P}_{\mathrm{col}}$ of words that can be k-coloured (i.e., each of its letters assigned a colour in $[k]$) so that for all $c \in [k]$ the induced c coloured sub-word is in $\mathcal{P}_c$.

8.1.4 Finite forcibility

Finite forcibility was introduced by Lovász and Sós [83] while studying a generalization of quasi-random graphs. For an in depth investigation of finitely forcible graphons we refer to the work of Lovász and Szegedy [86]. We say that $f \in \mathcal{W}$ is *finitely forcible* if there is a finite list of words $\boldsymbol{u}_1, \ldots \boldsymbol{u}_m$ such that any function $h : [0,1] \to [0,1]$ which satisfies $t(\boldsymbol{u}_i, h) = t(\boldsymbol{u}_i, f)$ for all $i \in [m]$ must agree with f almost everywhere. A direct consequence of Theorem 8.1.1 concerning quasi-random words is that the constant functions are finitely forcible (by words of length three). We can generalize this result as follows:

Theorem 8.1.5. *Piecewise polynomial functions are finitely forcible. Specifically, if there is an interval partition* $\{I_1, ..., I_k\}$ *of* $[0,1]$*, polynomials* $P_1(x), ..., P_k(x)$ *of degrees* $d_1, ..., d_k$*, respectively, and* $f \in \mathcal{W}$ *is such that* $f(x) = P_i(x)$ *for all* $i \in [k]$ *and* $x \in I_i$*, then there is a list of words* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$*, with* $m \leqslant 2^{1+2k+2\sum_i d_i} + 2^{\binom{k}{2}(1+\max_i d_i)}$ *such that any function* $h : [0,1] \to [0,1]$ *which satisfies* $t(\boldsymbol{u}_i, h) = t(\boldsymbol{u}_i, f)$ *for all* $i \in [m]$ *must agree with* f *almost everywhere.*

8.1.5 Permutons from words limits

Given $n \in \mathbb{N}$, we denote by $\mathfrak{S}_n$ the set of permutations of order n and $\mathfrak{S} = \bigcup_{n \geqslant 1} \mathfrak{S}_n$ the set of all finite permutations. Also, for $\sigma \in \mathfrak{S}_n$ and $\tau \in \mathfrak{S}_k$ we let $\Lambda(\tau, \sigma)$ be the number of copies of τ in σ, that is, the number of k-tuples $1 \leqslant x_1 < \cdots < x_k \leqslant n$ such that for every $i, j \in [k]$

$$\sigma(x_i) \leqslant \sigma(x_j) \quad \text{iff} \quad \tau(i) \leqslant \tau(j).$$

The density of copies of τ in σ, denoted by $t(\tau, \sigma)$, is the probability that σ restricted to a randomly chosen k-tuple of $[n]$ yields a copy of τ. A sequence $(\sigma_n)_{n \to \infty}$ of permutations, with $\sigma_n \in \mathfrak{S}_n$ for each $n \in \mathbb{N}$, is said to be convergent if $\lim_{n\to\infty} t(\tau, \sigma_n)$ exists for every permutation $\tau \in \mathfrak{S}$. Hoppen et al. [64] proved that every convergent sequence of permutations converges to a suitable analytic object called *permuton*, which are probability measures on the Borel σ-algebra on $[0,1] \times [0,1]$ with uniform marginals, the collection of which they denote by $\mathcal{Z}$, and also extend the map $t(\tau, \cdot)$ to the whole of $\mathcal{Z}$. Then, they define a metric $d_\square$ on $\mathcal{Z}$ so that for all $\tau \in \mathfrak{S}$ the maps $t(\tau, \cdot)$ are continuous with respect to $d_\square$. They also show that $(\mathcal{Z}, d_\square)$ is compact and, as a consequence, establish that convergence as defined above and convergence in $d_\square$ are equivalent. In particular, they prove that for every convergent sequence of permutations $(\sigma_n)_{n\to\infty}$ there is a permuton $\mu \in \mathcal{Z}$ such that $t(\tau, \sigma_n) \to t(\tau, \mu)$ for all $\tau \in \mathfrak{S}$. We give new proofs of these two results by using a more direct approach based on Theorem 8.1.3.

8.2 Quasi-randomness

In this section we give the proof of the second part of Theorem 8.1.1 and Theorem 8.1.2. We start by establishing an inverse form of the Cauchy–Schwarz inequality which is used to prove the second part of Theorem 8.1.1, that controlling the density of subsequences of length three is enough to guarantee uniformity. An alternative demonstration of the second part of Theorem 8.1.1 can be extracted from the proof of Theorem 8.1.5 (see Remark 8.5.2).

Then, after recalling some basic facts and terminology about Fourier analysis and Lipschitz functions, we proceed to prove the equivalence of the quasi-random properties listed in Theorem 8.1.2.

Lemma 8.2.1. *If $\boldsymbol{g} = (g_1, \ldots, g_n), \boldsymbol{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ and $\varepsilon \in (0,1)$ are such that*

$$\langle \boldsymbol{g}, \boldsymbol{h} \rangle^2 \geqslant \|\boldsymbol{g}\|^2 \|\boldsymbol{h}\|^2 - \varepsilon n^3 \|\boldsymbol{h}\|^2,$$

then all but at most $\varepsilon^{1/3} n$ indices $i \in [n]$ satisfy $g_i = \frac{\langle \boldsymbol{g}, \boldsymbol{h} \rangle}{\langle \boldsymbol{h}, \boldsymbol{h} \rangle} h_i \pm \varepsilon^{1/3} n$.

Proof. Let $\boldsymbol{z}$ be the projection of $\boldsymbol{g}$ onto the plane orthogonal to $\boldsymbol{h}$, i.e., $\boldsymbol{z} = \boldsymbol{g} - \frac{\langle \boldsymbol{g}, \boldsymbol{h} \rangle}{\langle \boldsymbol{h}, \boldsymbol{h} \rangle} \boldsymbol{h}$. As $\boldsymbol{z}$ and $\boldsymbol{h}$ are orthogonal, applying Pythagoras to $\boldsymbol{g} = \frac{\langle \boldsymbol{g}, \boldsymbol{h} \rangle}{\langle \boldsymbol{h}, \boldsymbol{h} \rangle} \boldsymbol{h} + \boldsymbol{z}$ yields

$$\|\boldsymbol{g}\|^2 = \tfrac{\langle \boldsymbol{g}, \boldsymbol{h} \rangle^2}{\langle \boldsymbol{h}, \boldsymbol{h} \rangle^2} \|\boldsymbol{h}\|^2 + \|\boldsymbol{z}\|^2 = \tfrac{\langle \boldsymbol{g}, \boldsymbol{h} \rangle^2}{\|\boldsymbol{h}\|^2} + \|\boldsymbol{z}\|^2.$$

The assumption then yields

$$\varepsilon n^3 \geqslant \|\boldsymbol{z}\|^2 = \sum_{i\in[n]} \left(g_i - \tfrac{\langle \boldsymbol{g},\boldsymbol{h}\rangle}{\langle \boldsymbol{h},\boldsymbol{h}\rangle} h_i\right)^2 . \tag{8.3}$$

Thus, the conclusion of the lemma must hold, otherwise $\|\boldsymbol{z}\|^2 > \varepsilon^{1/3} n(\varepsilon^{1/3}n)^2 = \varepsilon n^3$, contradicting (8.3). □

Proof (of the second part of Theorem 8.1.1). Given $\varepsilon > 0$ let $n > n_0$ be sufficiently large. By a word containing $*$ we mean the family of words obtained by replacing $*$ by 0 or 1, e.g., $\boldsymbol{u} = (*u_2u_3)$ denotes the family $\{(0u_2u_3),(1u_2u_3)\}$. For a word $\boldsymbol{u}$ containing $*$, let $\binom{\boldsymbol{w}}{\boldsymbol{u}} = \sum_{\boldsymbol{u}'} \binom{\boldsymbol{w}}{\boldsymbol{u}'}$ where the sum ranges over the family mentioned above. Given a word $\boldsymbol{w} = (w_1 \dots w_n) \in \{0,1\}^n$ which satisfies the assumption of the theorem we have

$$\binom{\boldsymbol{w}}{11*} \leqslant d^2\binom{n}{3} + 2\varepsilon n^3 \quad \text{and} \qquad \binom{\boldsymbol{w}}{*1*} + \binom{\boldsymbol{w}}{1**} \geqslant 2d\binom{n}{3} - 8\varepsilon n^3. \tag{8.4}$$

We may also assume that $d \geqslant \varepsilon$, otherwise the first condition yields $\|\boldsymbol{w}\|_1 \leqslant 3\varepsilon^{1/3}n$ due to $\binom{\|\boldsymbol{w}\|_1}{3} = \binom{\boldsymbol{w}}{111}$ and the result follows trivially.

Let $\boldsymbol{g} = (g_1,\dots,g_n)$ where $g_\ell = \sum_{i\in[\ell]} w_i$ and let $\boldsymbol{h} = (1,2,\dots,n)$. Since $g_n = \|\boldsymbol{w}\|_1$, it is easily seen that $\boldsymbol{w}$ is $18\varepsilon^{1/3}$-uniform if

$$g_\ell = \frac{\langle \boldsymbol{g},\boldsymbol{h}\rangle}{\langle \boldsymbol{h},\boldsymbol{h}\rangle}\ell \pm 9\varepsilon^{1/3}n \qquad \text{for every } \ell \in [n]. \tag{8.5}$$

To show (8.5) note first that

$$g_\ell^2 = |\{(i,j)\in[\ell]^2 \colon w_i = w_j = 1\}| \leqslant |\{(i,j)\in[\ell-1]^2 \colon w_i = w_j = 1, i\neq j\}| + 3(\ell-1) + 1.$$

Hence, up to an additive error of $3(\ell-1)+1$ the quantity g_ℓ^2 is twice the number of subsequences of $\boldsymbol{w}$ equal to $(11w_\ell)$. Summing over all $\ell \in [n]$ we obtain from (8.4)

$$\|\boldsymbol{g}\|^2 = \sum_{\ell\in[n]} g_\ell^2 \leqslant 2\binom{\boldsymbol{w}}{11*} + \tfrac{3}{2}n^2 \leqslant 2d^2\binom{n}{3} + 5\varepsilon n^3. \tag{8.6}$$

Consider next, for an $\ell \in [n]$, the family S_ℓ of subsequences of $\boldsymbol{w}$ equal to $(w_iw_jw_\ell)$ or $(w_jw_iw_\ell)$, where $i,j \in [\ell-1]$, $i \neq j$, and $w_i = 1$, $w_\ell \in \{0,1\}$. Then, we have $|S_\ell| \leqslant g_\ell \cdot \ell$, since there are at most g_ℓ choices for i and each such choice of i gives rise to $(i-1)+(\ell-i-1) \leqslant \ell$ choices for j. On the other hand, $\sum_{\ell\in[n]} |S_\ell|$ counts all subsequences of $\boldsymbol{w}$ of the form $(*1*)$ and $(1{*}{*})$. Hence, (8.4) together with $\boldsymbol{h} = (1,2,\dots,n)$ yields

$$\langle \boldsymbol{g},\boldsymbol{h}\rangle^2 = \Big(\sum_{\ell\in[n]} g_\ell\cdot\ell\Big)^2 \geqslant \Big(\sum_{\ell\in[n]} |S_\ell|\Big)^2 = \left(\binom{\boldsymbol{w}}{*1*} + \binom{\boldsymbol{w}}{1**}\right)^2 \geqslant 4d^2\binom{n}{3}^2 - 32\varepsilon\binom{n}{3}n^3.$$

As $\|\boldsymbol{h}\|^2 = \sum_{i\in[n]} i^2 = \frac{1}{6}n(n+1)(2n+1) = 2\binom{n}{3} + \frac{3}{2}n^2 - \frac{n}{2}$ from (8.6) we obtain

$$\begin{aligned}\langle \boldsymbol{g},\boldsymbol{h}\rangle^2 - \|\boldsymbol{g}\|^2\|\boldsymbol{h}\|^2 &\geqslant 4d^2\binom{n}{3}^2 - 32\varepsilon\binom{n}{3}n^3 - \left(2d^2\binom{n}{3} + 5\varepsilon n^3\right)\|\boldsymbol{h}\|^2 \\ &\geqslant 2d^2\binom{n}{3}\left(\|\boldsymbol{h}\|^2 - \frac{3}{2}n^2\right) - 16\varepsilon n^3\|\boldsymbol{h}\|^2 - \left(2d^2\binom{n}{3} + 5\varepsilon n^3\right)\|\boldsymbol{h}\|^2 \\ &\geqslant -22\varepsilon n^3\|\boldsymbol{h}\|^2.\end{aligned}$$

By Lemma 8.2.1 all but at most $(22\varepsilon)^{1/3}n$ indices $i \in [n]$ satisfy $g_i = \frac{\langle \boldsymbol{g},\boldsymbol{h}\rangle}{\langle \boldsymbol{h},\boldsymbol{h}\rangle} i \pm (22\varepsilon)^{1/3}n$. In particular, for every $\ell \in [n]$ there is such an index i with $i = \ell \pm (22\varepsilon)^{1/3}n$. Thus

$$g_\ell = g_i \pm (22\varepsilon)^{1/3}n = \frac{\langle \boldsymbol{g},\boldsymbol{h}\rangle}{\langle \boldsymbol{h},\boldsymbol{h}\rangle} i \pm 2(22\varepsilon)^{1/3}n = \frac{\langle \boldsymbol{g},\boldsymbol{h}\rangle}{\langle \boldsymbol{h},\boldsymbol{h}\rangle} \ell \pm 3(22\varepsilon)^{1/3}n$$

which shows (8.5) and the second part of Theorem 8.1.1 follows. □

Remark 8.2.2. *The previous proof shows something stronger than what is claimed. Specifically, that instead of requiring the right count of all subsequences of length three it is sufficient to have* (8.4)*, i.e., the correct upper bound for the count of* $(11*)$ *and the correct lower bound for the sum of the count of* $(*1*)$ *and* $(1**)$*.*

We now turn our attention to Theorem 8.1.2 and recall here some facts from Fourier analysis on the circle. Letting $\mathrm{d}x$ correspond to the Lebesgue measure on the unit circle, for $k \in \mathbb{Z}$, the Fourier transform $\widehat{f}(k)$ of a function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is defined by

$$\widehat{f}(k) = \int_{\mathbb{R}/\mathbb{Z}} f(x) e^{-2\pi i k x}\,\mathrm{d}x.$$

Given $N \in \mathbb{N}$, the *Fejér approximation* of order N of f is defined by

$$\sigma_N f(x) = \sum_{|n|\leqslant N} \Big(1 - \frac{|n|}{N+1}\Big) \widehat{f}(n) e^{2\pi i n x}.$$

Lemma 8.2.3 (Proposition 1.2.12 from [92])**.** *There is a constant* $C > 0$ *such that for any Lipschitz function* $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ *and for every* $M \geqslant 2$ *one has*

$$\|f - \sigma_M f\|_\infty \leqslant C\|f\|_{\mathrm{Lip}} \frac{\log M}{M}.$$

Lemma 8.2.4 (Theorem 1.5.3 from [92])**.** *There is a constant* $c > 0$ *such that for any Lipschitz function* $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ *and for every* $m \neq 0$ *one has*

$$|\widehat{f}(m)| \leqslant \frac{c\|f\|_{\mathrm{Lip}}}{|m|}.$$

We are now in the position to prove Theorem 8.1.2.

Proof (of Theorem 8.1.2). The equivalence between the Uniformity, Counting, and Minimizer properties follow from Theorem 8.1.1. The equivalence between the Cayley graph and Counting properties follows by noting that there is a one-to-n correspondence between subsequences in $\boldsymbol{w}_n$ equal to $\boldsymbol{u}$ and increasing $\boldsymbol{u}$-paths in $\Gamma(\boldsymbol{w}_n)$. To see this, simply note that $(v_1, ..., v_{\ell+1})$ is an increasing $\boldsymbol{u}$-path in $\Gamma(\boldsymbol{w}_n)$ if and only if $(v_1+a, ..., v_{\ell+1}+a)$ is an increasing $\boldsymbol{u}$-path in $\Gamma(\boldsymbol{w}_n)$, for all $a \in [n]$ (where arithmetic over vertices is modulo n). The equivalence between the properties Uniformity and Exponential sums was shown by Cooper in [38, Theorem 2.2] who also proved that if Exponential sums is true for a particular α_0, then it is true for all $\alpha > 0$. We next show that the properties Exponential sums and Equidistribution are equivalent. It is clear that the latter implies the former for $\alpha = 1$, and thus for all $\alpha > 0$, by Cooper's work and since $f(x) = \exp(2\pi i k x)$ integrates to 0 and has Lipschitz norm at most $2|k|$. To show the converse let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ be given. We will show that for any $\varepsilon > 0$ and for large n, the following holds for $d = \|\boldsymbol{w}\|_1/n$:

$$\left|\frac{1}{n}\sum_{j:\boldsymbol{w}_n[j]=1} f(j/n) - d\int_{\mathbb{R}/\mathbb{Z}} f\right| \leqslant \varepsilon\|f\|_{\mathrm{Lip}}.$$

Let C and c be the absolute constants from Lemma 8.2.3 and Lemma 8.2.4, respectively. Choose M large enough so that $M/\log M \geqslant 2C/\varepsilon$ and n large enough so that for all $|m| \leqslant M$ we have $\left|\sum_{j:\boldsymbol{w}_n[j]=1}\exp\left(\frac{2\pi i}{n}mj\right)\right| < \frac{\varepsilon}{2cM}n|m|$. Applying this bound we obtain

$$\begin{aligned}\sum_{j:\boldsymbol{w}_n[j]=1}\sigma_M f(j/n) &= \sum_{j:\boldsymbol{w}_n[j]=1}\sum_{|m|\leqslant M}\left(1-\frac{|m|}{M+1}\right)\widehat{f}(m)\exp\left(\tfrac{2\pi i}{n}mj\right)\\ &= \sum_{|m|\leqslant M}\left(1-\frac{|m|}{M+1}\right)\widehat{f}(m)\sum_{j:\boldsymbol{w}_n[j]=1}\exp\left(\tfrac{2\pi i}{n}mj\right)\\ &= \widehat{f}(0)\cdot dn \pm \frac{\varepsilon}{2cM}n\sum_{0<|m|\leqslant M}\left|\left(1-\frac{|m|}{M+1}\right)\widehat{f}(m)\right||m|.\end{aligned}$$

As $\widehat{f}(0) = \int_{\mathbb{R}/\mathbb{Z}} f$, we obtain from Lemma 8.2.4 that

$$\left|\frac{1}{n}\sum_{j:\boldsymbol{w}_n[j]=1}\sigma_M f(j/n) - d\int_{\mathbb{R}/\mathbb{Z}} f\right| \leqslant \frac{\varepsilon}{2cM}\sum_{0<|m|\leqslant M}\left|\left(1-\frac{|m|}{M+1}\right)\widehat{f}(m)\right||m| \leqslant \frac{\varepsilon}{2}\|f\|_{\mathrm{Lip}}.$$

By Lemma 8.2.3, triangle inequality and the choice of M we conclude

$$\begin{aligned}\left|\frac{1}{n}\sum_{j:\boldsymbol{w}_n[j]=1} f(j/n) - d\int_{\mathbb{R}/\mathbb{Z}} f\right| &\leqslant \left|\frac{1}{n}\sum_{j:\boldsymbol{w}_n[j]=1}\sigma_M f(j/n) - d\int_{\mathbb{R}/\mathbb{Z}} f\right| + C\|f\|_{\mathrm{Lip}}\frac{\log M}{M}\\ &\leqslant \frac{\varepsilon}{2}\|f\|_{\mathrm{Lip}} + \frac{\varepsilon}{2}\|f\|_{\mathrm{Lip}} = \varepsilon\|f\|_{\mathrm{Lip}}.\end{aligned}$$

This finishes the proof. $\square$

8.3 Limits of word sequences

In this section we give the proof of Theorem 8.1.3 concerning word limits. Although the overall approach is in line with what has been done for graphons [84] and permutons [64], there are important technical differences which we will stress below. Central concepts and auxiliary results involved in the proof will be introduced along the way. The section is divided into four subsections. We start by a simple reformulation of the notion of convergent word sequences in terms of convergence of a function sequence in $\mathcal{W}$. This notion is called t-convergence and we show in Lemma 8.3.1 that the limit of a t-convergent function sequence is unique, if it exists. In the second subsection, we endow $\mathcal{W}$ with the interval-distance $d_\square$ and show in Lemma 8.3.2 that convergence with respect to $d_\square$ implies t-convergence. Proposition 8.3.6 from the same subsection gives a direct proof of the converse. In the third subsection, we specify a third and last notion of convergence (convergence in distribution) based on sampling of f-random letters for a given $f \in \mathcal{W}$. We prove in Lemma 8.3.8 that this notion of convergence is equivalent to the two previously defined, and deduce the compactness of the metric space $(\mathcal{W}, d_\Box)$ in Theorem 8.3.9. In the fourth and last part, we show in Lemma 8.3.10 and Corollary 8.3.11 that every element of $f \in \mathcal{W}$ is, a.s., the limit of a convergent random word sequence.

8.3.1 Uniqueness and t-convergence

Given the nature of the limit it is convenient to first reformulate the notion of convergence in analytic terms. For a given word $\boldsymbol{w}_n = (w_1 \dots w_n)$ define the *function associated to* $\boldsymbol{w}_n$ to be the n-step 0-1-function $f_{\boldsymbol{w}_n} \in \mathcal{W}$ given by $f_{\boldsymbol{w}_n}(x) = w_{\lceil nx \rceil}$. It is then easy to see that $t(\boldsymbol{u}, f_{\boldsymbol{w}_n})$, as defined in (8.2), satisfies[2]

$$t(\boldsymbol{u}, f_{\boldsymbol{w}_n}) = t(\boldsymbol{u}, \boldsymbol{w}_n) + O\big(n^{-1}\big) \qquad \text{for every word } \boldsymbol{u}. \tag{8.7}$$

Thus the following, applied to $f_n = f_{\boldsymbol{w}_n}$, yields a reformulation of convergence of $(\boldsymbol{w}_n)_{n\to\infty}$. Given a sequence $(f_n)_{n\to\infty}$ in $\mathcal{W}$ and $f \in \mathcal{W}$, we say that

$$f_n \xrightarrow{t} f \qquad \text{if} \qquad \lim_{n\to\infty} t(\boldsymbol{u}, f_n) = t(\boldsymbol{u}, f) \quad \text{for all finite words } \boldsymbol{u}.$$

The next lemma implies that the limit, if it exists, is guaranteed to be unique. The idea of the proof goes back to a remark of Král' and Pikhurko concerning permutons (see [75, Remark 6]).

Lemma 8.3.1. *Let $f, g : [0,1] \to [0,1]$. If $t(\boldsymbol{u}, f) = t(\boldsymbol{u}, g)$ for all words $\boldsymbol{u}$, then $f = g$ almost everywhere.*

Proof. Given $k \in \mathbb{N}$, note that

$$\begin{aligned}
\int_0^1 f(x)x^k \,\mathrm{d}x &= \int_0^1 f(x)\Big(\int_0^x \mathrm{d}y\Big)^k \mathrm{d}x = \int_{y_1,\dots,y_k \leqslant x} f(x)\,\mathrm{d}y_1 \dots \mathrm{d}y_k \,\mathrm{d}x \\
&= k! \int_{y_1<\dots<y_k<x} f(x)\,\mathrm{d}y_1 \dots \mathrm{d}y_k \,\mathrm{d}x = \frac{1}{k+1} \sum_{\boldsymbol{u}\in\{0,1\}^k} t(u_1 \dots u_k 1, f) \\
&= \frac{1}{k+1} \sum_{\boldsymbol{u}\in\{0,1\}^k} t(u_1 \dots u_k 1, g) = \int_0^1 g(x)x^k \,\mathrm{d}x.
\end{aligned}$$

Thus, for each polynomial $P(x) \in \mathbb{R}[x]$ we get $\int_0^1 f(x)P(x)\,\mathrm{d}x = \int_0^1 g(x)P(x)\,\mathrm{d}x$, and by the Stone–Weierstrass theorem $\int_0^1 f(x)h(x)\,\mathrm{d}x = \int_0^1 g(x)h(x)\,\mathrm{d}x$ holds for every continuous function $h : [0,1] \to \mathbb{R}$. This implies that $f = g$ almost everywhere. □

[2]To see (8.7), split $[0,1]$ into n intervals of equal lengths. Let A denote the event that ℓ independent uniform random points of $[0,1]$ land in different intervals and let B be the event that, after reordering these points, say $x_1 < \dots < x_\ell$, we have $\big(f_{\boldsymbol{w}_n}(x_1), \dots, f_{\boldsymbol{w}_n}(x_\ell)\big) = \boldsymbol{u}$. Then, $t(\boldsymbol{u}, f_{\boldsymbol{w}_n}) = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\overline{A}]\mathbb{P}[\overline{A}]$ and we further have $\mathbb{P}[B|A] = t(\boldsymbol{u}, \boldsymbol{w}_n)$ and $\mathbb{P}[A] = \prod_{i=1}^{\ell-1}(1 - i/n) = 1 - O(n^{-1})$.

8.3.2 Interval-metric and the metric space $(\mathcal{W}, d_\square)$

In view of the equivalence of uniformity and subsequence counts shown in Theorem 8.1.1, it is natural to consider the following notions of norm, distance and convergence, which are all analogues of the notions of cut-norm, cut-distance and convergence in graph limit theory. Given $h : [0,1] \to [-1,1]$ define the interval-norm

$$\|h\|_\square = \sup_{I \subseteq [0,1]} \left| \int_I h(x)\,\mathrm{d}x \right|,$$

where the supremum is taken over all intervals $I \subseteq [0,1]$. The interval-metric $d_\square$ is then defined by

$$d_\square(f,g) = \|f - g\|_\square \qquad \text{for every } f, g : [0,1] \to [0,1],$$

and we write

$$f_n \xrightarrow{\square} f \qquad \text{if} \qquad \lim_{n\to\infty} d_\square(f_n, f) = 0.$$

The following result states that the interval-norm controls subsequence counts, in particular, $f_n \xrightarrow{\square} f$ implies $f_n \xrightarrow{t} f$. As a by-product of the lemma, we obtain the first part of Theorem 8.1.1 concerning counting subsequences in uniform words.

Lemma 8.3.2. *For $f, g \in \mathcal{W}$ and $\boldsymbol{u} \in \{0,1\}^\ell$ we have*

$$\Big| t(\boldsymbol{u}, f) - t(\boldsymbol{u}, g) \Big| \leqslant \ell^2 \cdot d_\square(f,g).$$

In particular, if $\boldsymbol{w} \in \{0,1\}^n$ is ε-uniform and $n = n(\varepsilon, \ell)$ is sufficiently large, then for some $d \in [0,1]$ we have for each $\boldsymbol{u} \in \{0,1\}^\ell$

$$\tbinom{\boldsymbol{w}}{\boldsymbol{u}} = d^{\|\boldsymbol{u}\|_1} (1-d)^{\ell - \|\boldsymbol{u}\|_1} \tbinom{n}{\ell} \pm 5\varepsilon n^\ell.$$

Proof. We first show that the second part follows from the first. Given an ε-uniform word $\boldsymbol{w} \in \{0,1\}^n$, let $f : [0,1] \to [0,1]$ be the function associated to $\boldsymbol{w}$ and let $d = \int f(t)\,\mathrm{d}t \in [0,1]$. Define $g : [0,1] \to [0,1]$ constant equal to d and recall that $g^1 = g$ and $g^0 = 1 - g$. Then, for each $\boldsymbol{u} \in \{0,1\}^\ell$

$$t(\boldsymbol{u}, g) = \ell! \int_{0 \leqslant x_1 < \cdots < x_\ell \leqslant 1} \prod_{i \in [\ell]} g^{u_i}(x_i)\,\mathrm{d}x_1 \ldots \mathrm{d}x_\ell = d^{\|\boldsymbol{u}\|_1} (1-d)^{\ell - \|\boldsymbol{u}\|_1}.$$

Since $d_\square(f,g) \leqslant 2\varepsilon$ due to uniformity of $\boldsymbol{w}$, for large n, the second part of the lemma follows from the first part and (8.7) as

$$\tbinom{\boldsymbol{w}}{\boldsymbol{u}} = t(\boldsymbol{u}, f)\tbinom{n}{\ell} \pm \varepsilon n^\ell = t(\boldsymbol{u}, g)\tbinom{n}{\ell} \pm 5\varepsilon n^\ell = d^{\|\boldsymbol{u}\|_1} (1-d)^{\ell - \|\boldsymbol{u}\|_1} \tbinom{n}{\ell} \pm 5\varepsilon n^\ell.$$

Now we turn to the proof of the first part. Let

$$X_j(x_1, \ldots, x_\ell) = \Big(f^{u_j}(x_j) - g^{u_j}(x_j) \Big) \prod_{i=1}^{j-1} f^{u_i}(x_i) \prod_{i=j+1}^{\ell} g^{u_i}(x_i).$$

Making use of a telescoping sum we write

$$
\begin{aligned}
\Big|t(\boldsymbol{u},f)-t(\boldsymbol{u},g)\Big| &= \ell!\Big|\int_{x_1<\cdots<x_\ell}\Big(\prod_{j\in[\ell]}f^{u_j}(x_j)-\prod_{j\in[\ell]}g^{u_j}(x_j)\Big)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell\Big| \\
&= \ell!\Big|\int_{x_1<\cdots<x_\ell}\sum_{j\in[\ell]}X_j(x_1,\ldots,x_\ell)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell\Big| \\
&\leqslant \ell!\sum_{j\in[\ell]}\Big|\int_{x_1<\cdots<x_\ell}X_j(x_1,\ldots,x_\ell)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell\Big|.
\end{aligned}
$$

Since $\Big|\int_{x_{j-1}}^{x_{j+1}}\Big(f^{u_j}(x_j)-g^{u_j}(x_j)\Big)\,\mathrm{d}x_j\Big|\leqslant d_\square(f,g)$ and $0\leqslant f,g\leqslant 1$, for $j\in[\ell]$ we have

$$
\Big|\int_{x_{j-1}}^{x_{j+1}}X_j(x_1,...,x_\ell)\,\mathrm{d}x_j\Big|\leqslant d_\square(f,g)\prod_{i=1}^{j-1}f^{u_i}(x_i)\prod_{i=j+1}^{\ell}g^{u_i}(x_i).
$$

Hence,

$$
\begin{aligned}
&\Big|\int_{x_1<\cdots<x_\ell}X_j(x_1,\ldots,x_\ell)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell\Big| \\
&\qquad\leqslant d_\square(f,g)\int_{\substack{x_1<\cdots<x_{j-1}\\ \leqslant x_{j+1}<\cdots<x_\ell}}\prod_{i=1}^{j-1}f^{u_i}(x_i)\prod_{i=j+1}^{\ell}g^{u_i}(x_i)\,\mathrm{d}x_1\ldots\mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1}\ldots\mathrm{d}x_\ell \\
&\qquad\leqslant\frac{1}{(\ell-1)!}d_\square(f,g)
\end{aligned}
$$

and the first part of the lemma follows. $\square$

Remark 8.3.3. *We note that the same argument extends without change to larger size alphabets in the following sense. Given an alphabet $\Sigma=\{a_1,\ldots,a_k\}$, let $\boldsymbol{f}=(f^{a_1},\ldots,f^{a_k})$ and $\boldsymbol{g}=(g^{a_1},\ldots,g^{a_k})$ be two tuples of functions $f^{a_i},g^{a_i}:[0,1]\to[0,1]$, for $i\in[k]$, such that*

$$
f^{a_1}(x)+\cdots+f^{a_k}(x)=1 \text{ and } g^{a_1}(x)+\cdots+g^{a_k}(x)=1 \text{ almost everywhere.}
$$

For a word $\boldsymbol{u}\in\Sigma^\ell$, define the density of $\boldsymbol{u}$ in $\boldsymbol{f}$ in similar manner as in (8.2)*, namely*

$$
t(\boldsymbol{u},\boldsymbol{f})=\ell!\int_{0\leqslant x_1<\cdots<x_\ell\leqslant 1}\prod_{i\in[k]}f^{u_i}(x_i)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell.
$$

Then, the proof from above yields

$$
\Big|t(\boldsymbol{u},\boldsymbol{f})-t(\boldsymbol{u},\boldsymbol{g})\Big|\leqslant\ell^2\cdot\max_{i\in[k]}d_\square(f^{a_i},g^{a_i}).
$$

Note that Lemma 8.3.2 implies that if $f_n\xrightarrow{\square}f$, then $f_n\xrightarrow{t}f$. Our goal now is to show that the converse also holds. Let $(f_n)_{n\to\infty}$ be a sequence such that $f_n\xrightarrow{t}f$. Following the proof of Lemma 8.3.1, we will use that for any polynomial $P(x)\in\mathbb{R}[x]$ we can write $\int_0^1(f_n(x)-f(x))P(x)$ as a linear combination of subsequence densities. By approximating $\mathbf{1}_{[a,b]}(x)$ by a polynomial $P_{a,b}(x)\in\mathbb{R}[x]$, with error term uniform in $0\leqslant a<b\leqslant 1$, we may show that $\int_0^1(f_n(x)-f(x))\mathbf{1}_{[a,b]}(x)$ can be approximated by $\int_0^1(f_n(x)-f(x))P_{a,b}(x)$, thence

by a linear combination of subsequence densities, implying our claim. In order to prove this approximation result, we introduce next the class of Bernstein polynomials,

$$b_{t,i}(x) = \binom{t}{i} x^i (1-x)^{t-i}, \qquad \text{for all } t \in \mathbb{N},\ i \in [t] \text{ and } x \in [0,1].$$

Since $b_{t,i}(x)$ is the probability mass function (pmf) of a binomial random variable we have that:

Fact 8.3.4. $\sum_{i=0}^{t} b_{t,i}(x) = 1, \qquad \sum_{i=0}^{t} i b_{t,i}(x) = tx \qquad$ *and* $\qquad \sum_{i=0}^{t} (tx-i)^2 b_{t,i}(x) = tx(1-x).$

Even though here we only need to approximate functions on $[0,1]$, we will consider the general case of functions on $[0,1]^k$ since it will later be useful in our study of higher dimensional combinatorial structures. For $k,t \in \mathbb{N} \setminus \{0\}$, let $\boldsymbol{i} = (i_1, \dots i_k) \in [t]^k$. Given a function $J : [0,1]^k \to \mathbb{R}$, define its Bernstein polynomial evaluated at $\boldsymbol{x} = (x_1, \dots, x_k) \in [0,1]^k$ by

$$B_{t,J}(\boldsymbol{x}) = \sum_{0 \leqslant i_1, \dots, i_k \leqslant t} J(\tfrac{\boldsymbol{i}}{t}) \prod_{j \in [k]} b_{t,i_j}(x_j).$$

We can now formally state the approximation of indicator functions we use.

Lemma 8.3.5. *For $\boldsymbol{a} = (a_1, \dots, a_k) \in [0,1]^k$ let $J = \mathbf{1}_{[0,a_1] \times \cdots \times [0,a_k]}$. If $r \in \mathbb{N}$ and $\boldsymbol{x} \in [0,1]^k$ satisfy $|x_i - a_i| > r^{-1/4}$ for all $i \in [k]$, then $|B_{r,J}(\boldsymbol{x}) - J(\boldsymbol{x})| \leqslant kr^{-1/2}$.*

Proof. Let $B = B_{r,J}$. By Fact 8.3.4 we have

$$|B(\boldsymbol{x}) - J(\boldsymbol{x})| = \Big| B(\boldsymbol{x}) - J(\boldsymbol{x}) \sum_{0 \leqslant i_1, \dots, i_k \leqslant r} \prod_{j \in [k]} b_{r,i_j}(x_j) \Big| \leqslant \sum_{0 \leqslant i_1, \dots, i_k \leqslant r} \Big| J(\tfrac{\boldsymbol{i}}{r}) - J(\boldsymbol{x}) \Big| \prod_{j \in [k]} b_{r,i_j}(x_j).$$

Let $L = \{\boldsymbol{i} \colon \|\boldsymbol{x} - \frac{\boldsymbol{i}}{r}\|_\infty > r^{-1/4}\} \subseteq (\{0\} \cup [r])^k$. As $|x_j - a_j| > r^{-1/4}$ for all $j \in [k]$, for each $\boldsymbol{i} \notin L$ we have that $J(\frac{\boldsymbol{i}}{r}) = J(\boldsymbol{x})$ and thus

$$\sum_{\boldsymbol{i} \notin L} \Big| J(\tfrac{\boldsymbol{i}}{r}) - J(\boldsymbol{x}) \Big| \prod_{j \in [k]} b_{r,i_j}(x_j) = 0.$$

For $\ell \in [k]$, let $L_\ell = \{\boldsymbol{i} \in L \colon |rx_\ell - i_\ell| > r^{3/4}\}$, and note that $L = L_1 \cup \cdots \cup L_k$. Due to $\Big| J(\frac{\boldsymbol{i}}{r}) - J(\boldsymbol{x}) \Big| \leqslant 1$ we have

$$\sum_{\boldsymbol{i} \in L} \Big| J(\tfrac{\boldsymbol{i}}{r}) - J(\boldsymbol{x}) \Big| \prod_{j \in [k]} b_{r,i_j}(x_j) \leqslant \sum_{\ell \in [k]} \sum_{\boldsymbol{i} \in L_\ell} \prod_{j \in [k]} b_{r,i_j}(x_j). \tag{8.8}$$

By Fact 8.3.4, since $b_{r,i_j}(x) \leqslant 1$, for every $x \in [0,1]$,

$$\sum_{\boldsymbol{i} \in L_k} \prod_{j \in [k]} b_{r,i_j}(x_j) \leqslant \sum_{\boldsymbol{i} \in L_k} \frac{(rx_k - i_k)^2}{r^{3/2}} b_{r,i_k}(x_k) = \frac{1}{r^{1/2}} x_k(1-x_k) \leqslant \frac{1}{r^{1/2}}.$$

The same bound holds for every L_ℓ, $\ell \in [k-1]$. Therefore, the RHS of (8.8) is at most $kr^{-1/2}$, as required. $\square$

Given two functions $f, g \in \mathcal{W}$, we have the inequality

$$\sup_{b\in[0,1]} \left| \int_0^b f(x)\,\mathrm{d}x - \int_0^b g(x)\,\mathrm{d}x \right| \leqslant d_\square(f,g) \leqslant 2 \sup_{b\in[0,1]} \left| \int_0^b f(x)\,\mathrm{d}x - \int_0^b g(x)\,\mathrm{d}x \right|. \tag{8.9}$$

The first inequality in (8.9) is direct from the definition of $d_\square$, and the second inequality follows from the identity $\int_0^b (f(x) - g(x)) = \int_a^b (f(x) - g(x)) + \int_0^a (f(x) - g(x))$.

The following proposition states that t-convergence implies convergence with respect to $d_\square$, and thus, together with Lemma 8.3.2, establishes that both notions of convergence are equivalent.

Proposition 8.3.6. *If $(f_n)_{n\to\infty}$ is a sequence in $\mathcal{W}$ which is t-convergent, then it is a Cauchy sequence with respect to $d_\square$. Moreover, if $f_n \xrightarrow{t} f$ for some $f \in \mathcal{W}$, then $f_n \xrightarrow{\square} f$.*

Proof. Given $\varepsilon > 0$, let $r = \lceil (20/\varepsilon)^4 \rceil$. For $\delta = \varepsilon/2^{3r+2}$, let n_0 be sufficiently large so that for all $n, m \geqslant n_0$ we have

$$\left| t(\boldsymbol{u}, f_n) - t(\boldsymbol{u}, f_m) \right| \leqslant \delta \quad \text{ for all } \boldsymbol{u} \in \bigcup_{s\in[r]} \{0,1\}^s. \tag{8.10}$$

Recall from the proof of Lemma 8.3.1, that for each $k \in \mathbb{N}$ we have

$$\int_0^1 f(x) x^k \,\mathrm{d}x = \frac{1}{k+1} \sum_{\boldsymbol{u}\in\{0,1\}^k} t(u_1 \ldots u_k 1, f).$$

Thus, for $k \leqslant r$ and $h = f_n - f_m$, we have

$$\left| \int_0^1 h(x) x^k \,\mathrm{d}x \right| = \frac{1}{k{+}1} \left| \sum_{\boldsymbol{u}\in\{0,1\}^k} (t(u_1 \ldots u_k 1, f_n) - t(u_1, \ldots, u_k 1, f_m)) \right| \leqslant \frac{2^k \delta}{k{+}1}.$$

For $a \in [0,1]$, let $J_a = \mathbf{1}_{[0,a]}$ and j_a be the largest index such that $\frac{j_a}{r} \leqslant a$. Then,

$$\left| \int_0^1 h(x) B_{r,J_a}(x) \,\mathrm{d}x \right| \leqslant \sum_{i=0}^{j_a} \tbinom{r}{i} \left| \int_0^1 h(x) x^i (1-x)^{r-i} \,\mathrm{d}x \right| \leqslant 2^{3r}\delta.$$

Thus, since $|h| \leqslant 1$ and $|\mathbf{1}_{[0,a]}(x) - B_{r,J_a}| \leqslant 2$, by Lemma 8.3.5, we have

$$\begin{aligned} \left| \int_0^1 h(x) \mathbf{1}_{[0,a]}(x) \,\mathrm{d}x \right| &\leqslant \left| \int_0^1 h(x) B_{r,J_a}(x) \,\mathrm{d}x \right| + \left| \int_0^1 h(x) (\mathbf{1}_{[0,a]}(x) - B_{r,J_a}(x)) \,\mathrm{d}x \right| \\ &\leqslant 2^{3r}\delta + (4r^{-1/4} + r^{-1/2}). \end{aligned}$$

The desired conclusion follows from (8.9) and by our choice of t and δ observing that

$$d_\square(f_n, f_m) \leqslant 2 \sup_{a\in[0,1]} \left| \int_0^1 h(x) \mathbf{1}_{[0,a]}(x) \,\mathrm{d}x \right| \leqslant 2^{3r+1}\delta + 10 r^{-1/4} \leqslant \varepsilon.$$

The second part follows by replacing f_m by f in (8.10), taking $h = f_n - f$, and repeating the above argument. $\square$

The compactness of the metric space $(\mathcal{W}, d_\square)$ can be easily established via the Banach–Alaoglu theorem in $L^\infty([0,1])$. Instead, we follow a different strategy laid out in the following section. This strategy has the advantage that it emphasizes the probabilistic point of view of convergence. It is based on a new model of random words that naturally arises from the theory and that may be of independent interest.

We note that one can also establish the compactness of $(\mathcal{W}, d_\square)$ by using the regularity lemma for words [12]. This approach has the advantage of being more constructive and for the sake of completeness we include it in the Section 8.6.

8.3.3 Random letters from limits and compactness of $(\mathcal{W}, d_\square)$

Consider the standard metric on $[0,1]$ and the discrete metric on $\{0,1\}$. Let $\Omega = [0,1] \times \{0,1\}$ be equipped with the L_∞-distance, which thus assigns to a pair of points in Ω the standard distance of their first coordinates if the second coordinates agree and one otherwise. Let $\mathcal{B}$ denote the Borel σ-algebra of Ω, let $f : [0,1] \to [0,1]$ be a Borel measurable function and recall that $f^1 = f$ and $f^0 = 1 - f$. Also, denote by $\mathrm{U}([0,1])$ and $\mathrm{B}(p)$ the uniform distribution over $[0,1]$ and the Bernoulli distribution with expected value $p \in [0,1]$, respectively. We say that

$$(X, Y) \in \Omega \text{ is an } f\text{-random letter} \qquad \text{if} \qquad X \sim \mathrm{U}([0,1]) \text{ and } Y \sim \mathrm{B}(f(X)).$$

Observe that an f-random letter (X, Y) is a pair of mixed[3] random variables where Y is distributed according to the conditional pmf

$$f_{Y|X}(\varepsilon|x) = \mathbb{P}[Y = \varepsilon | X = x] = f^\varepsilon(x) \qquad \varepsilon \in \{0,1\} \text{ and } x \in [0,1].$$

Then, (X, Y) has the mixed joint cumulative probability distribution

$$F(x, \varepsilon) = \mathbb{P}[X \leqslant x, Y = \varepsilon] = \int_0^x f^\varepsilon(t)\,\mathrm{d}t, \tag{8.11}$$

and thus the mixed joint pmf $f_{X,Y}(x, \varepsilon) = f^\varepsilon(x)$. The marginal probability distribution of Y is

$$\mathbb{P}[Y = \varepsilon] = F(1, \varepsilon) = \int_0^1 f^\varepsilon(t)\,\mathrm{d}t, \qquad \varepsilon \in \{0,1\},$$

hence $Y \sim \mathrm{B}(p)$ with $p = \int_0^1 f(t)\,\mathrm{d}t$. Furthermore, conditioned on Y the variable X is distributed according to the conditional pmf $f_{X|Y}$ which satisfies

$$f_{X|Y}(x|\varepsilon) \cdot \mathbb{P}[Y = \varepsilon] = f_{X,Y}(x, \varepsilon) = f^\varepsilon(x). \tag{8.12}$$

One may therefore equivalently sample (X, Y) by first choosing $Y \sim \mathrm{B}(p)$ with $p = \int_0^1 f(t)\,\mathrm{d}t$, and then choose X (conditional on Y) according to the conditional pmf $f_{X|Y}$ satisfying (8.12). By means of this sampling procedure a sequence $(f_n)_{n\to\infty}$ gives rise to a sequence $\big((X_n, Y_n)\big)_{n\to\infty}$, where each (X_n, Y_n) is the f_n-random letter, and the corresponding sequence of probability distributions $(\mathbb{P}_n)_{n\to\infty}$ is as defined in (8.11). As usual for general metric spaces (see, e.g., [21,

[3]Mixed in the sense that X is continuous while Y is discrete.

Chapter 5]), we say that $\big((X_n, Y_n)\big)_{n\to\infty}$ converges to (X, Y) in distribution if $(\mathbb{P}_n)_{n\to\infty}$ weakly converges to $\mathbb{P}$, i.e., if for all bounded continuous functions $h : \Omega \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int_\Omega h \,\mathrm{d}\mathbb{P}_n = \int_\Omega h \,\mathrm{d}\mathbb{P}. \tag{8.13}$$

From this definition we immediately have the following.

Fact 8.3.7. *If $\big((X_n, Y_n)\big)_{n\to\infty}$ converges to (X, Y) in distribution, then $(X_n)_{n\to\infty}$ (resp. $(Y_n)_{n\to\infty}$) converges to X (resp. Y) in distribution.*

We now write

$$f_n \xrightarrow{\mathrm{d}} f \qquad \text{if} \qquad \big((X_n, Y_n)\big)_{n\to\infty} \text{ converges to } (X, Y) \text{ in distribution.}$$

The next lemma shows the equivalences of convergence in $d_\square$ and convergence in distribution.

Lemma 8.3.8. *Let $f_1, f_2, \dots$ and f be functions in $\mathcal{W}$. Then, $f_n \xrightarrow{\square} f$ if and only if $f_n \xrightarrow{\mathrm{d}} f$.*

Proof. Let (X_n, Y_n) be an f_n-random letter (resp. (X, Y) be an f-random letter) with the associated probability measure $\mathbb{P}_n$ and cumulative distribution F_n (resp. $\mathbb{P}$ and F). Let

$$\|F_n - F\|_\infty = \sup_{(x,\varepsilon)\in\Omega} |F_n(x, \varepsilon) - F(x, \varepsilon)|$$

and note that by definition we have

$$\|F_n - F\|_\infty = \sup_{x\in\Omega} |F_n(x, 0) - F(x, 0)| = \sup_{x\in\Omega} |F_n(x, 1) - F(x, 1)|.$$

Now observe that

$$\|F_n - F\|_\infty \leqslant d_\square(f_n, f) \leqslant 2\|F_n - F\|_\infty, \tag{8.14}$$

where the first inequality is obvious and the second one follows because for all $\varepsilon \in \{0, 1\}$ and $0 \leqslant a < b \leqslant 1$ it holds that $\int_{[a,b]} (f_n - f)(t)\,\mathrm{d}t = (F_n - F)(b, \varepsilon) - (F_n - F)(a, \varepsilon)$. Thus, $f_n \xrightarrow{\square} f$ if and only if $\lim_{n\to\infty} \|F_n - F\|_\infty = 0$ which we claim holds if and only if

$$\lim_{n\to\infty} F_n(x, \varepsilon) = F(x, \varepsilon) \qquad \text{for all } \varepsilon \in \{0, 1\} \text{ and } x \in [0, 1]. \tag{8.15}$$

Indeed, it is clear that $\lim_{n\to\infty} \|F_n - F\|_\infty = 0$ implies (8.15). For the converse note that for each $\varepsilon \in \{0, 1\}$ we have $|f^\varepsilon| \leqslant 1$, thus for every $x, y \in [0, 1]$

$$|F(x, \varepsilon) - F(y, \varepsilon)| = \left|\int_0^x f^\varepsilon(t)\,\mathrm{d}t - \int_0^y f^\varepsilon(t)\,\mathrm{d}t\right| \leqslant |x - y|. \tag{8.16}$$

Given an integer $k > 0$, by (8.15), there is an n_k such that $\max_{i\in[k]} \left|F_n\left(\frac{i}{k}, \varepsilon\right) - F\left(\frac{i}{k}, \varepsilon\right)\right| < \frac{1}{k}$ for each $n > n_k$. For an $x \in [0, 1]$ let $i_x \in [k]$ be such that $|x - \frac{i_x}{k}| \leqslant \frac{1}{k}$. Then, by triangle inequality and (8.16), for any $x \in [0, 1]$

$$|F_n(x, \varepsilon) - F(x, \varepsilon)| \leqslant \left|F_n\left(\tfrac{i_x}{k}, \varepsilon\right) - F\left(\tfrac{i_x}{k}, \varepsilon\right)\right| + 2|x - \tfrac{i_x}{k}| \leqslant \tfrac{3}{k}$$

which thus establishes that (8.15) implies $\lim_{n\to\infty} \|F_n - F\|_\infty = 0$.

To prove the lemma we now show that (8.15) holds if and only if (X_1, Y_1), $(X_2, Y_2), \dots$ converges to (X, Y) in distribution, i.e., $\mathbb{P}_1, \mathbb{P}_2, \dots$ weakly converges to $\mathbb{P}$ as defined in (8.13). For an $h : \Omega \to \mathbb{R}$ and an $\varepsilon \in \{0, 1\}$ define the projection $h_\varepsilon : [0, 1] \to \mathbb{R}$ via $h_\varepsilon(x) = h(x, \varepsilon)$. Thus, $F_\varepsilon(x) = F(x, \varepsilon)$, $F_{n,\varepsilon}(x) = F_n(x, \varepsilon)$ and we also define $\mathbb{P}_\varepsilon$ via $\mathbb{P}_\varepsilon[A] = \mathbb{P}[A \times \{\varepsilon\}]$ for any $A \in \mathcal{B}([0, 1])$ and in the same manner define $\mathbb{P}_{n,\varepsilon}$.

For a metric space (M, d), we denote by $C(M)$ the set of continuous functions $h : M \to \mathbb{R}$. As Ω is equipped with L_∞-distance d_Ω we have $d_\Omega((x, \alpha), (y, \beta)) = \delta < 1$ if an only if $\alpha = \beta$ and $|x - y| = \delta$. Hence, $h \in C(\Omega)$ if and only if $h_0, h_1 \in C([0, 1])$. Moreover, by verifying the following for step functions h and then extending to all $h \in C(\Omega)$ by a standard limiting argument we have

$$\int_\Omega h \, \mathrm{d}\mathbb{P}_n = \sum_\varepsilon \int_{[0,1]} h_\varepsilon \, \mathrm{d}\mathbb{P}_{n,\varepsilon} \qquad \text{and} \qquad \int_\Omega h \, \mathrm{d}\mathbb{P} = \sum_\varepsilon \int_{[0,1]} h_\varepsilon \, \mathrm{d}\mathbb{P}_\varepsilon.$$

In particular,

$$\lim_{n\to\infty} \int_\Omega h \, \mathrm{d}\mathbb{P}_n = \int_\Omega h \, \mathrm{d}\mathbb{P} \quad \text{for all } h \in C(\Omega)$$

holds if and only if

$$\lim_{n\to\infty} \int_\Omega h \, \mathrm{d}\mathbb{P}_{n,\varepsilon} = \int_\Omega h \, \mathrm{d}\mathbb{P}_\varepsilon \quad \text{for all } \varepsilon \in \{0, 1\}, \text{ and all } h \in C([0, 1]).$$

In other words, $\mathbb{P}_1, \mathbb{P}_2, \dots$ converges weakly to $\mathbb{P}$ if and only if $\mathbb{P}_{1,\varepsilon}, \mathbb{P}_{2,\varepsilon}, \dots$ converges weakly to $\mathbb{P}_\varepsilon$ for all $\varepsilon \in \{0, 1\}$. As the underlying space is $[0, 1]$ it is well known that weak convergence of $\mathbb{P}_{1,\varepsilon}, \mathbb{P}_{2,\varepsilon}, \dots$ to $\mathbb{P}_\varepsilon$ is equivalent to the fact that $\lim_{n\to\infty} F_{n,\varepsilon}(x) = F_\varepsilon(x)$ holds for all x where $F_\varepsilon(x)$ is continuous. As seen from (8.16), F_ε is continuous on the entirety of $[0, 1]$. This thus shows that weak convergence of $\mathbb{P}_1, \mathbb{P}_2, \dots$ to $\mathbb{P}$ is equivalent to (8.15) and the lemma follows. $\square$

The compactness of $(\mathcal{W}, d_\square)$ now follows from Lemma 8.3.8 and classical results from measure theory, namely Prokhorov's theorem concerning the existence of weak convergent subsequences for a given sequence of measures over compact measurable spaces and Radon–Nikodym theorem concerning the existence of derivatives of measures which are absolutely continuous with respect to the Lebesgue measure.

Theorem 8.3.9. *The metric space $(\mathcal{W}, d_\square)$ is compact.*

Proof. Given a sequence $(f_n)_{n\to\infty}$ of functions $f_n \in \mathcal{W}$. Consider the sequence of f_n-random letters $\big((X_n, Y_n)\big)_{n\to\infty}$ with the corresponding sequence of probabilities $(\mathbb{P}_n)_{n\to\infty}$ on $(\Omega, \mathcal{B})$ defined by (8.11). As Ω is compact we conclude from Prokhorov's theorem (see Chapter 1, Section 5 of [21]) that there is a pair of random variables (X, Y) with joint probability measure $\mathbb{P}$ such that $(\mathbb{P}_n)_{n\to\infty}$ contains a subsequence $(\mathbb{P}_{n_i})_{i\to\infty}$ which weakly converges to $\mathbb{P}$. By Fact 8.3.7 we know that $X \sim \mathrm{U}[0, 1]$ while Y is Bernoulli. Denoting by λ the Lebesgue measure, the restriction of $\mathbb{P}$ to $Y = 1$ yields a measure μ which satisfies $\mu(A) = \mathbb{P}[X \in A, Y = 1] \leqslant \lambda(A)$ for every measurable set A. In particular, μ is absolutely continuous with respect to the Lebesgue measure λ (i.e., $\mu(A) = 0$ whenever $\lambda(A) = 0$) and the Radon–Nikodym theorem guarantees the existence of a function f such that

$$\mu([0, x]) = \int_0^x f(t) \, \mathrm{d}t = \mathbb{P}[X \leqslant x, Y = 1]$$

and thus

$$\mathbb{P}[X \leqslant x, Y = 0] = x - \mu([0,x]) = \int_0^x (1 - f(t))\,\mathrm{d}t.$$

In other words, $f_{X,Y}(x,\varepsilon) = f^{\varepsilon}(x)$ is the pmf of (X,Y) and we thus have $f_{n_i} \xrightarrow{\mathrm{d}} f$. Lemma 8.3.8 guarantees that $f_{n_i} \xrightarrow{\square} f$ as well. Lastly, it is easily seen that $f(x) \in [0,1]$ almost everywhere and we may therefore assume that $f \in \mathcal{W}$. $\square$

The last theorem thus establishes the existence of the limit object claimed in the first part of Theorem 8.1.3.

8.3.4 Random words from limits

To establish the second part of Theorem 8.1.3 we consider, for any $f \in \mathcal{W}$, a suitable sequence of random words arising from f and show that it converges to f almost surely. For $f \in \mathcal{W}$ and $\boldsymbol{x} = (x_1, ..., x_\ell) \in [0,1]^\ell$ such that $x_1 < x_2 < ... < x_\ell$ let $\boldsymbol{w} = \mathrm{sub}(\boldsymbol{x}, f)$ be the word obtained by choosing $w_i = 1$ with probability $f(x_i)$ and $w_i = 0$ with probability $1 - f(x_i)$ (making independent decisions for different x_i's). Consider now n independent f-random letters $(X_1, Y_1), \ldots, (X_n, Y_n)$. After reordering the first coordinate, i.e., taking a permutation $\sigma : [n] \to [n]$ so that $X_{\sigma(1)} < \cdots < X_{\sigma(n)}$, the *$f$-random word* $\mathrm{sub}(n,f)$ is given by

$$\mathrm{sub}(n,f) = (Y_{\sigma(1)}, \ldots, Y_{\sigma(n)}).$$

Lemma 8.3.10. *Let $f \in \mathcal{W}$ and let f_n be the function associated to the f-random word* $\mathrm{sub}(n,f)$*. For all $n \in \mathbb{N}$ and $a \geqslant \frac{1}{n}$ we have*

$$\mathbb{P}\Big[d_\square(f_n, f) \geqslant 10a\Big] \leqslant 4ne^{-2an^2}.$$

Proof. For $x \in [0,1]$ let

$$W_n(x) = \int_0^x f_n(t)\,\mathrm{d}t \quad \text{and} \quad W(x) = \int_0^x f(t)\,\mathrm{d}t.$$

Recall that by (8.9) we have $d_\square(f_n, f) \leqslant 2\|W_n - W\|_\infty$. Therefore, we only need to bound $\mathbb{P}[\|W_n - W\|_\infty \geqslant 5a]$.

Given $i \in [n]$ and $x \in [\frac{i-1}{n}, \frac{i}{n})$, since $|f_n|, |f| \leqslant 1$, we have that $|W_n(x) - W(x)| \leqslant |W_n(\frac{i}{n}) - W(\frac{i}{n})| + \frac{2}{n}$, and thus

$$\|W_n - W\|_\infty \leqslant \frac{2}{n} + \max_{i\in[n]} |W_n(\tfrac{i}{n}) - W(\tfrac{i}{n})|.$$

For $i \in [n]$, we next bound the probability that $|W_n(\frac{i}{n}) - W(\frac{i}{n})|$ is at least $3a$. Consider the sequence $(X_1, Y_1), \ldots, (X_n, Y_n)$ of f-random letters that define $\mathrm{sub}(n,f)$, and suppose that $X_{\sigma(1)} < \cdots < X_{\sigma(n)}$ for some permutation $\sigma : [n] \to [n]$. Since f_n is the function associated to $\mathrm{sub}(n,f)$ we have

$$\Big|W_n(\tfrac{i}{n}) - \frac{1}{n}\sum_{j=1}^{i} Y_{\sigma(j)}\Big| \leqslant \frac{1}{n}$$

and thus, letting $Z_i = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{X_j \leqslant \frac{i}{n}\}$ and $S_i = \frac{1}{n}\sum_{j=1}^{n} Y_j \mathbf{1}\{X_j \leqslant \frac{i}{n}\} = \frac{1}{n}\sum_{j=1}^{Z_i} Y_{\sigma(j)}$, we get

$$\Big|W_n(\tfrac{i}{n}) - S_i\Big| \leqslant \frac{1}{n} + \Big|\frac{i}{n} - Z_i\Big|. \tag{8.17}$$

On the other hand, for every $j \in [n]$ we have that

$$\mathbb{E}[Y_j \mathbf{1}\{X_j \leqslant \tfrac{i}{n}\}] = \int_0^{\frac{i}{n}} f(t)\,\mathrm{d}t = W(\tfrac{i}{n}),$$

so $\mathbb{E}[S_i] = W(\frac{i}{n})$. Using Chernoff's bound (see Theorem 2.8 and Remark 2.5 from [66]) we get

$$\mathbb{P}\Big[\Big|Z_i - \tfrac{i}{n}\Big| \geqslant a\Big] \leqslant 2e^{-2a^2 n} \quad \text{and} \quad \mathbb{P}\Big[\Big|S_i - W(\tfrac{i}{n})\Big| \geqslant a\Big] \leqslant 2e^{-2a^2 n},$$

which together with (8.17) and the fact that $a \geqslant \frac{1}{n}$, implies that

$$\mathbb{P}\Big[|W_n(\tfrac{i}{n}) - W(\tfrac{i}{n})| \geqslant 3a\Big] \leqslant \mathbb{P}\Big[|S_i - W(\tfrac{i}{n}))| \geqslant a\Big] + \mathbb{P}\Big[\Big|Z_i - \tfrac{i}{n}\Big| \geqslant a\Big] \leqslant 4e^{-2a^2 n}.$$

Putting everything together we conclude that

$$\mathbb{P}[d_\square(f_n, f) \geqslant 10a] \leqslant \mathbb{P}[\|W_n - W\|_\infty \geqslant 5a] \leqslant \sum_{i=1}^{n} \mathbb{P}\Big[|W_n(\tfrac{i}{n}) - W(\tfrac{i}{n})| \geqslant 3a\Big] \leqslant 4ne^{-2a^2 n}.$$

□

As an immediate consequence we obtain the following.

Corollary 8.3.11. *For all $f \in \mathcal{W}$, the sequence of f-random words* $(\mathrm{sub}(n, f))_{n\to\infty}$ *converges to f a.s.*

Proof. For $n \in \mathbb{N}$ let $f_n = \mathrm{sub}(n, f)$. Taking $a = n^{-\frac{1}{4}}$ in Lemma 8.3.10 and using the Borel–Cantelli lemma, it follows that $f_n \xrightarrow{\square} f$ almost surely. Then, by Lemma 8.3.2 we conclude that $f_n \xrightarrow{t} f$ almost surely, and therefore, by (8.7), $(\mathrm{sub}(n, f))_{n\to\infty}$ converges to f almost surely. □

Equipped with the results from above we now establish the second main result of this section.

Proof (of Theorem 8.1.3). The uniqueness of the limit, if it exists, follows from Lemma 8.3.1. The second part of the theorem concerning the existence of word sequences converging to any given $f \in \mathcal{W}$ follows from Corollary 8.3.11.

It is thus left to establish the existence of a limit. Consider a convergent sequence $(\boldsymbol{w}_n)_{n\to\infty}$ of words and let $(f_n)_{n\to\infty}$ be the sequence of associated functions $f_n = f_{\boldsymbol{w}_n} \in \mathcal{W}$. Because of (8.7) the sequence $(f_n)_{n\to\infty}$ is t-convergent and thus, by Proposition 8.3.6, $(f_n)_{n\to\infty}$ is a Cauchy sequence with respect to $d_\square$. The compactness of $(\mathcal{W}, d_\square)$, as guaranteed by Theorem 8.3.9, implies that there exists $f \in \mathcal{W}$ such that $d_\square(f_n, f) \to 0$. Finally, because of Lemma 8.3.2 we have that $f_n \xrightarrow{t} f$ and therefore $(\boldsymbol{w}_n)_{n\to\infty}$ converges to f. □

Concluding this section and in preparation for the next one, we show that a tail bound on $d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}})$ similar to the one of Lemma 8.3.10 holds if instead of sampling an $f_{\boldsymbol{w}}$-random word for some word $\boldsymbol{w}$, we sample a subsequence $\boldsymbol{u} = \mathrm{sub}(\ell, \boldsymbol{w})$.

Lemma 8.3.12. *Let $\boldsymbol{w} \in \{0,1\}^n$, $\ell \in [n]$ and $\frac{1}{8} \geqslant a > \frac{1}{\ell}$. Then, for the random word $\boldsymbol{u} = \mathrm{sub}(\ell, \boldsymbol{w})$ we have that*

$$\mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) \geqslant 8a] \leqslant 2\ell e^{-\frac{1}{3}\ell a^2}.$$

Proof. For $x \in [0,1]$ let $F_{\boldsymbol{u}}(x) = \int_0^x f_{\boldsymbol{u}}(t)\,\mathrm{d}t$ and $F_{\boldsymbol{w}}(x) = \int_0^x f_{\boldsymbol{w}}(t)\,\mathrm{d}t$. By an argument similar to the initial part of the proof of Lemma 8.3.10, we get that

$$\mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) \geqslant 8a] \leqslant \mathbb{P}\Big[\max_{i\in[\ell]} |F_{\boldsymbol{u}}(\tfrac{i}{\ell}) - F_{\boldsymbol{w}}(\tfrac{i}{\ell})| \geqslant 2a\Big] \leqslant \sum_{i\in[\ell]} \mathbb{P}\Big[|F_{\boldsymbol{u}}(\tfrac{i}{\ell}) - F_{\boldsymbol{w}}(\tfrac{i}{\ell})| \geqslant 2a\Big]. \quad (8.18)$$

Now, let $I_1, ..., I_n$ be indicator random variables summing up to ℓ, and observe that

$$S_i = F_{\boldsymbol{u}}(\tfrac{i}{\ell}) = \frac{1}{\ell} \sum_{j\in[n]:\frac{j}{n}\leqslant\frac{i}{\ell}} \boldsymbol{w}[j] I_j.$$

By linearity of expectation and given that $\mathbb{E}[I_j] = \frac{\ell}{n}$ for every $j \in [n]$ it follows that

$$\mathbb{E}[S_i] = \frac{1}{n} \sum_{j\in[n]:\frac{j}{n}\leqslant\frac{i}{\ell}} \boldsymbol{w}[j] = F_{\boldsymbol{w}}(\lfloor \tfrac{i}{\ell} n \rfloor \tfrac{1}{n}) = F_{\boldsymbol{w}}(\tfrac{i}{\ell}) \pm \tfrac{1}{n} = F_{\boldsymbol{w}}(\tfrac{i}{\ell}) \pm \tfrac{1}{\ell}.$$

Using that $a > \frac{1}{\ell}$, by (8.18), we get that

$$\mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) \geqslant 8a] \leqslant \sum_{i\in[\ell]} \mathbb{P}[|S_i - \mathbb{E}(S_i)| \geqslant a]. \quad (8.19)$$

Let $X_i = \ell S_i$. Note that $X_i = \sum_{j\in J_i(\boldsymbol{w})} I_j$ where $J_i(\boldsymbol{w}) = \{j \in [n] : j \leqslant \frac{i}{\ell} n, \boldsymbol{w}[j] = 1\}$. We claim that X_i is a hypergeometric distribution with parameters n, ℓ and $|J_i(\boldsymbol{w})|$ (the distribution of the number of black balls obtained by sampling without replacement ℓ balls from a set of n balls of which $|J_i(\boldsymbol{w})|$ are black). It is well known that Chernoff type tail bounds hold for these distributions (see for example [66, Theorem 2.10]). Specifically, by (2.5) and (2.6) from [66], for $\lambda = \ell|J_i(\boldsymbol{w})|/n$, and since $\lambda \leqslant \ell$, we have that

$$\mathbb{P}[S_i \leqslant \mathbb{E}[S_i] - a] = \mathbb{P}[X_i \leqslant \mathbb{E}[X_i] - \ell a] \leqslant \exp\Big(-\frac{(\ell a)^2}{2\lambda}\Big) \leqslant e^{-\frac{1}{2}\ell a^2},$$

and, since $a \leqslant \frac{1}{8} \leqslant \frac{3}{2}$,

$$\mathbb{P}[S_i \geqslant \mathbb{E}[S_i]+a] = \mathbb{P}[X_i \geqslant \mathbb{E}[X_i]+\ell a] \leqslant \exp\Big(-\frac{(\ell a)^2}{2(\lambda + \ell a/3)}\Big) \leqslant \exp\Big(-\frac{\ell a^2}{2(1 + a/3)}\Big) \leqslant e^{-\frac{1}{3}\ell a^2}.$$

The last two tail bounds together with (8.19) yield the desired conclusion. $\square$

8.4 Testing hereditary word properties

We now turn our focus to algorithmic considerations. Specifically, to the study of testable word properties and how it relates to word limits (recall that a word property $\mathcal{P}$ is simply a collection of words). The presentation below is heavily influenced by the derivation of analogous results for graphons by Lovász and Szegedy [85] (for related results concerning testability of permutation properties and limit objects see [65, 69]). First, we define the notion of closure of a word property and then give two alternative useful characterizations. Next, we shall see that there is a close connection between testability of word properties and attributes of their closures. Finally, we derive this section's main result, that is Theorem 8.1.4.

First, we define the *closure* of a word property $\mathcal{P}$, denoted $\overline{\mathcal{P}}$, as

$$\overline{\mathcal{P}} = \{f \in \mathcal{W}\colon\ \boldsymbol{w}_n \in \mathcal{P} \text{ for all } n \in \mathbb{N}, \text{ and } \boldsymbol{w}_n \xrightarrow{t} f\}.$$

Recall that property $\mathcal{P}$ is hereditary if $\operatorname{sub}(I, \boldsymbol{w}) \in \mathcal{P}$ for every $\boldsymbol{w} \in \mathcal{P}$ of length n and every $I \subseteq [n]$.

Proposition 8.4.1. *If $\mathcal{P}$ is a hereditary word property, then*

$$\overline{\mathcal{P}} = \{f \in \mathcal{W}\colon\ \mathbb{P}[\operatorname{sub}(\ell, f) \notin \mathcal{P}] = 0 \textit{ for all } \ell \geqslant 1\} = \{f \in \mathcal{W}\colon\ t(\boldsymbol{u}, f) = 0 \textit{ for all } \boldsymbol{u} \notin \mathcal{P}\}.$$

Moreover, if there is a word that does not belong to $\mathcal{P}$, then every $f \in \overline{\mathcal{P}}$ is 0-1 *valued except maybe on a set of null measure.*

Proof. The second equality holds since for each integer $\ell \geqslant 1$ we have

$$\mathbb{P}[\operatorname{sub}(\ell, f) \notin \mathcal{P}] = \sum_{\boldsymbol{u}\in\{0,1\}^\ell\setminus\mathcal{P}} \mathbb{P}[\operatorname{sub}(\ell, f) = \boldsymbol{u}] = \sum_{\boldsymbol{u}\in\{0,1\}^\ell\setminus\mathcal{P}} t(\boldsymbol{u}, f). \tag{8.20}$$

To show the first equality recall from Corollary 8.3.11 that $\big(\operatorname{sub}(\ell, f)\big)_{\ell\to\infty}$ converges to f a.s. Hence, if moreover $\mathbb{P}[\operatorname{sub}(\ell, f) \in \mathcal{P}] = 1$ holds for every ℓ, then there is a sequence of words from $\mathcal{P}$ which converges to f, showing that $f \in \overline{\mathcal{P}}$.

To show the converse, let $(\boldsymbol{w}_n)_{n\to\infty}$ be a sequence of words in $\mathcal{P}$ that converges to $f \in \overline{\mathcal{P}}$, i.e., $\lim_{n\to\infty} t(\boldsymbol{u}, \boldsymbol{w}_n) = t(\boldsymbol{u}, f)$ for every word $\boldsymbol{u}$. In particular, if $\boldsymbol{u} \notin \mathcal{P}$ then $t(\boldsymbol{u}, \boldsymbol{w}_n) = 0$ by heredity of $\mathcal{P}$ and thus $t(\boldsymbol{u}, f) = 0$. By (8.20) we then obtain $\mathbb{P}[\operatorname{sub}(\ell, f) \notin \mathcal{P}] = 0$.

Finally, suppose that $f \in \overline{\mathcal{P}}$ and that there is a $\boldsymbol{u} \in \{0,1\}^\ell \setminus \mathcal{P}$ for some ℓ. Let $\boldsymbol{X} = (X_1, ..., X_\ell)$ be uniformly chosen in $[0,1]^\ell$, then the characterization of $\overline{\mathcal{P}}$ yields

$$\begin{aligned}
0 = \mathbb{P}[\operatorname{sub}(\ell, f) \notin \mathcal{P}] &\geqslant \mathbb{P}[\operatorname{sub}(\boldsymbol{X}, f) = \boldsymbol{u}] \\
&\geqslant \int_{\substack{x_1,\dots,x_\ell\in f^{-1}(]0,1[)\\ x_1<\dots<x_\ell}} \prod_{i\in[\ell]} f^{u_i}(x_i)\,\mathrm{d}x_1 \dots \mathrm{d}x_\ell \\
&\geqslant \tfrac{1}{\ell!} \int_{x_1,\dots,x_\ell\in f^{-1}(]0,1[)} \prod_{i\in[\ell]} f^{u_i}(x_i)\,\mathrm{d}x_1 \dots \mathrm{d}x_\ell.
\end{aligned}$$

Thus, $f^{-1}(]0,1[)$ has null Lebesgue measure. □

Next, we establish two technical results that will allow us to relate testability of hereditary word properties and characteristics of their closure. In what follows, for $f, g \in \mathcal{W}$ we write $d_1(f,g) = \|f-g\|_1$ for the usual distance in $L_1([0,1])$.

Proposition 8.4.2. *If $\mathcal{P}$ is an hereditary word property and $\boldsymbol{w}$ is a word, then $d_1(\boldsymbol{w},\mathcal{P}) \leqslant d_1(f_{\boldsymbol{w}},\overline{\mathcal{P}})$.*

Proof. We may assume that there is a word not contained in $\mathcal{P}$, since the conclusion is trivial otherwise. Let $\delta > 0$, then by Proposition 8.4.1 there is a 0-1 valued $g \in \overline{\mathcal{P}}$ such that $d_1(f_{\boldsymbol{w}}, g) \leqslant d_1(f_{\boldsymbol{w}},\overline{\mathcal{P}}) + \delta$. By Proposition 8.4.1 we know that $\mathbb{P}[\mathrm{sub}(n,g) \in \mathcal{P}] = 1$, hence, if $\boldsymbol{w}' = \mathrm{sub}(\boldsymbol{X}, g)$ where $\boldsymbol{X} = (X_1, ..., X_n)$ is such that X_i is uniformly chosen in the interval $[\frac{i-1}{n}, \frac{i}{n}]$, then $\mathbb{P}[\boldsymbol{w}' \in \mathcal{P}] = 1$ as well. Since the probability that index i contributes to $d_1(\boldsymbol{w},\boldsymbol{w}')$ is $g(X_i)$ if $w_i = 0$ and $1 - g(X_i)$ if $w_i = 1$ we have

$$\mathbb{E}[d_1(\boldsymbol{w},\boldsymbol{w}')] = \|f_{\boldsymbol{w}} - g\|_1 = d_1(f_{\boldsymbol{w}}, g) \leqslant d_1(f_{\boldsymbol{w}},\overline{\mathcal{P}}) + \delta.$$

In particular, there exists $\widetilde{\boldsymbol{w}} \in \mathcal{P}$ for which $d_1(f_{\boldsymbol{w}},\overline{\mathcal{P}})+\delta \geqslant d_1(\boldsymbol{w},\widetilde{\boldsymbol{w}}) \geqslant d_1(\boldsymbol{w},\mathcal{P})$ holds. Since δ is arbitrary, the desired conclusion follows. □

Lemma 8.4.3. *If $\mathcal{P}$ is an hereditary word property and $(f_n)_{n\to\infty}$ is a sequence of functions in $\mathcal{W}$ such that $d_\square(f_n,\overline{\mathcal{P}}) \to 0$, then $d_1(f_n,\overline{\mathcal{P}}) \to 0$.*

Proof. If every word is in $\mathcal{P}$, then $\overline{\mathcal{P}} = \mathcal{W}$ and the result is obvious. Assuming otherwise, suppose that $d_1(f_n,\overline{P}) \not\to 0$. Then, there exist $\varepsilon > 0$, a sequence $(\varepsilon_n)_{n\to\infty}$ that converges to 0, and a sequence $(g_n)_{n\to\infty}$ in $\mathcal{P}$ such that for all $n \in \mathbb{N}$ we have

$$d_1(f_n, g_n) \geqslant \varepsilon \quad \text{and} \quad d_\square(f_n, g_n) \leqslant d_\square(f_n,\overline{\mathcal{P}}) + \varepsilon_n.$$

Since $\mathcal{W}$ is compact (passing to a subsequence[4]) we may assume that $g_n \overset{\square}{\to} f$ for some $f \in \overline{\mathcal{P}}$, and deduce that $f_n \overset{\square}{\to} f$. Moreover, by Proposition 8.4.1 we get that f is 0–1 valued. Consider the Lebesgue measurable sets $\Omega_b = f^{-1}(b)$ for $b \in \{0,1\}$. Then

$$d_1(f_n, f) = \|f_n - f\|_1 = \int_{\Omega_0} f_n + \int_{\Omega_1} (1 - f_n).$$

In case Ω_0, Ω_1 are intervals we conclude from $\lim_{n\to\infty} d_\square(f_n, f) = 0$ that

$$\lim_{n\to\infty} \int_{\Omega_0} f_n = \int_{\Omega_0} f = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \int_{\Omega_1} (1 - f_n) = \int_{\Omega_1} (1 - f) = 0.$$

By standard limiting arguments this extends to finite unions of intervals and finally to all Lebesgue measurable sets, and the lemma follows. □

Finally, we are ready to derive the main result of this section.

[4]The term "passing to a subsequence" means considering a subsequence instead of the original sequence. However, to avoid making the notation more cumbersome, the subsequence keeps the same name as the original sequence.

Proof (of Theorem 8.1.4). Let $\mathcal{P}$ be a hereditary word property and let $\varepsilon > 0$. By Lemma 8.4.3 there is a $\delta = \delta(\varepsilon) > 0$ such that if $d_\square(f, \overline{\mathcal{P}}) < \delta$, then $d_1(f, \overline{\mathcal{P}}) < \varepsilon$. We first observe that, by definition of $\overline{\mathcal{P}}$ and Lemma 8.3.12, there is an $n(\varepsilon) \geqslant 1$ such that for every word $\boldsymbol{w}$ of length $n \geqslant n(\varepsilon)$ the following holds:

(i) If $\boldsymbol{w}$ belongs to $\mathcal{P}$, then $d_\square(f_{\boldsymbol{w}}, \overline{\mathcal{P}}) < \frac{\delta}{4}$.

(ii) If $\boldsymbol{u} = \mathrm{sub}(\ell, \boldsymbol{w})$ and $n \geqslant \ell \geqslant n(\varepsilon)$, then $\mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) < \delta/4] \geqslant \frac{2}{3}$.

Let $\mathcal{P}'$ be the collection of words $\boldsymbol{v}$ such that $d_\square(f_{\boldsymbol{v}}, \overline{\mathcal{P}}) \leqslant \frac{\delta}{2}$ (this depends on ϵ, but this is acceptable as discussed after introducing the notion of testability). We claim that $\mathcal{P}'$ is a test property for $\mathcal{P}$ (for the given ε).

Let $\boldsymbol{w}$ be a word which we assume to be of length $n \geqslant n(\varepsilon)$.[5] Let $\boldsymbol{u} = \mathrm{sub}(\ell, \boldsymbol{w})$ where $\ell \in [n]$. In order to establish completeness, suppose that $\boldsymbol{w} \in \mathcal{P}$. By definition of $\mathcal{P}'$ and triangle inequality

$$\mathbb{P}[\boldsymbol{u} \in \mathcal{P}'] = \mathbb{P}[d_\square(f_{\boldsymbol{u}}, \overline{\mathcal{P}}) \leqslant \tfrac{\delta}{2}] \geqslant \mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) + d_\square(f_{\boldsymbol{w}}, \overline{\mathcal{P}}) < \tfrac{\delta}{2}].$$

Hence, from (i) we get $\mathbb{P}[\boldsymbol{u} \in \mathcal{P}'] \geqslant \mathbb{P}[d_\square(f_{\boldsymbol{u}}, f_{\boldsymbol{w}}) < \frac{\delta}{4}]$. By (ii) it follows that $u \in \mathcal{P}'$ with probability at least 2/3.

To prove soundness, assume $\ell \geqslant n(\varepsilon)$ and that $\boldsymbol{u} \in \mathcal{P}'$ (i.e., $d_\square(f_{\boldsymbol{u}}, \overline{\mathcal{P}}) \leqslant \delta/2$) with probability strictly larger than $\frac{1}{3}$. Together with (ii), this implies that there is at least one subsequence $\widetilde{\boldsymbol{u}}$ of $\boldsymbol{w}$ such that $d_\square(f_{\widetilde{\boldsymbol{u}}}, f_{\boldsymbol{w}}) < \delta/4$ and $d_\square(f_{\widetilde{\boldsymbol{u}}}, \overline{\mathcal{P}}) \leqslant \delta/2$. By triangle inequality $d_\square(f_{\boldsymbol{w}}, \overline{\mathcal{P}}) < \delta$, so by our choice of δ, we have $d_1(f_{\boldsymbol{w}}, \overline{\mathcal{P}}) \leqslant \varepsilon$. Thus, Proposition 8.4.2, implies that $d_1(\boldsymbol{w}, \mathcal{P}) \leqslant d_1(f_{\boldsymbol{w}}, \overline{\mathcal{P}}) < \varepsilon$ as desired. $\square$

8.5 Finite forcibility

In this section we investigate word limits that are prescribed by a finite number of subsequence densities. In particular, we prove Theorem 8.1.5 showing that piecewise polynomial functions are forcible.

The proof relies on the following lemma which shows, among other, that moments of cumulative distributions can be characterized by a finite number of subsequence densities of the distribution's mass density function.

Lemma 8.5.1. *If $f : [0,1] \to [0,1]$ is a Lebesgue measurable function and $F(x) = \int_0^x f(t)\,\mathrm{d}t$, then for each $i, j \in \mathbb{N}$ we have*

$$\int x^i F(x)^j dx = \frac{i!j!}{(i+j+1)!} \sum_{\substack{\boldsymbol{u} \in \{0,1\}^{i+j+1} \\ u_1 + \ldots + u_{i+j} \geqslant j}} t(\boldsymbol{u}, f).$$

[5] Adding to $\mathcal{P}'$ every word of length smaller than $n(\epsilon)$ preserves its hereditary property and immediately implies that both completeness and soundness are satisfied for w's of length smaller than $n(\epsilon)$.

Proof. Observe that

$$\begin{aligned}
\int x^i F(x)^j \,\mathrm{d}x &= \int \Big(\int_0^x dy\Big)^i \Big(\int_0^x f(z)dz\Big)^j \,\mathrm{d}x \\
&= \int \Big(\int_{0\leqslant y_1,\dots,y_i\leqslant x} \mathrm{d}y_1 \dots \mathrm{d}y_i\Big)\Big(\int_{0\leqslant z_1,\dots,z_j\leqslant x} \prod_{k=1}^{j} f(z_k)\,\mathrm{d}z_1\dots\mathrm{d}z_j\Big)\,\mathrm{d}x \\
&= i!j!\int \Big(\int_{0\leqslant y_1<\dots<y_i\leqslant x} \mathrm{d}y_1 \dots \mathrm{d}y_i\Big)\Big(\int_{0\leqslant z_1<\dots<z_j\leqslant x} \prod_{k=1}^{j} f(z_k)\,\mathrm{d}z_1\dots\mathrm{d}z_j\Big)\,\mathrm{d}x \\
&= i!j! \sum_{S\subseteq[i+j]:|S|=j} \int_{0\leqslant x_1<\dots<x_{i+j}\leqslant x} \prod_{s\in S} f(x_s)\,\mathrm{d}x_1\dots\mathrm{d}x_{i+j}\,\mathrm{d}x.
\end{aligned}$$

Since

$$1 = \prod_{s\in[i+j]\setminus S}\Big(f(x_s)+(1-f(x_s))\Big) = \sum_{U\subseteq[i+j]:S\subseteq U}\Big(\prod_{s\in U\setminus S} f(x_s)\Big)\Big(\prod_{s\notin U}(1-f(x_s))\Big),$$

we get

$$\begin{aligned}
\int x^i F(x)^j \,\mathrm{d}x &= i!j! \sum_{U\subseteq[i+j]:|U|\geqslant j} \binom{|U|}{j} \int_{0\leqslant x_1<\dots<x_{i+j}\leqslant x} \prod_{s\in U} f(x_s) \prod_{s\notin U}(1-f(x_s))\,\mathrm{d}x_1\dots\mathrm{d}x_{i+j}\,\mathrm{d}x \\
&= \frac{i!j!}{(i+j+1)!} \sum_{\substack{\boldsymbol{u}\in\{0,1\}^{i+j+1} \\ u_1+\dots+u_{i+j}\geqslant j}} \binom{\|u\|_1}{j} t(\boldsymbol{u}, f).
\end{aligned}$$

□

We next prove this section's main result concerning the finite forcibility of piecewise polynomial functions.

Proof (of Theorem 8.1.5). Let $P_1(x), \dots, P_k(x)$ be polynomials where P_i is of degree d_i and let $\{I_1, \dots, I_k\}$ be an interval partition of $[0, 1]$ such that $f(x) = P_i(x)$ for all $x \in I_i$. Let

$$Q_i(x) = \int_{I_i\cap[0,x]} P_i(t)\,\mathrm{d}t + \sum_{j\in[k]:I_j\subseteq[0,x]} \int_{I_j} P_j(t)\,\mathrm{d}t.$$

Then, $F(x) = \int_0^x f(t)\,\mathrm{d}t$ is continuous and $F(x) = Q_i(x)$ for each $i \in [k]$.

Next, let $d = \sum_{i\in[k]} \deg(Q_i) = k + \sum_{i\in[k]} d_i$ and define the polynomial

$$P(x, y) = \Big(y - Q_1(x)\Big)^2\Big(y - Q_2(x)\Big)^2 \dots \Big(y - Q_k(x)\Big)^2 = \sum_{1\leqslant i+j\leqslant 2d} c_{ij} x^j y^i$$

for some coefficients c_{ij}. Note that $\int_0^1 P\Big(x, F(x)\Big)\,\mathrm{d}x = 0$. Moreover, Lemma 8.5.1 guarantees that there is a list of words of length at most $2d+1$, say, $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ with $m \leqslant 2^{2d+1}$, such that the fact $\int_0^1 P\Big(x, F(x)\Big)\,\mathrm{d}x = 0$ already follows from the prescription of the values $t(\boldsymbol{u}_i, f)$, $i \in [m]$. Thus, if $h \in \mathcal{W}$ is such that $t(\boldsymbol{u}_i, h) = t(\boldsymbol{u}_i, f)$ for all $i \in [m]$, then $H(x) = \int_0^x h(t)\,\mathrm{d}t$ is continuous and satisfies $0 = \int_0^1 P\Big(x, H(x)\Big)\,\mathrm{d}x$. This implies that $P\Big(x, H(x)\Big) = 0$ everywhere, and by the definition of $P(x,y)$ we conclude that for each $x \in [0,1]$ there is an $\ell = \ell(x) \in [k]$ such that $H(x) = Q_\ell(x)$. Suppose that $\ell(x) = j$ for some x and $\ell(x') = j' \neq j$ for some $x' > x$. As H is continuous this can only happen if Q_j intersects $Q_{j'}$ in the interval $[x, x']$. On the other hand, two polynomials Q_i and Q_j have at most $\max\{\deg(Q_i), \deg(Q_j)\}$ intersection points, thus there are at most $t = \binom{k}{2}(1 + \max_{i\in[k]} d_i)$ intersection points of $Q_1, \dots, Q_k$ in total. Let these points be ordered by the first coordinate. Then, each H from above can be associated to a subsequence of intersection points, thus there are at most 2^t functions H such that $P\Big(x, H(x)\Big) = 0$ everywhere, implying at most that many functions $h : [0,1] \to [0,1]$ such that $t(\boldsymbol{u}_i, h) = t(\boldsymbol{u}_i, f)$ for all $i \in [m]$. To finish the proof note that by uniqueness of word limits, see Theorem 8.1.3, we can find for each h, which differs from f by a non-zero measure set, a word $\boldsymbol{u}_h$ such that $t(\boldsymbol{u}_h, f) \neq t(\boldsymbol{u}_h, h)$. Thus, f is uniquely determined by the densities of at most $m + 2^t \leqslant 2^{1+2k+2\sum_i d_i} + 2^{\binom{k}{2}(1+\max_i d_i)}$ words. $\square$

Remark 8.5.2. *The same proof for $k = 1$ and $P_1(x) = a$ being constant yields an alternative proof of the second part of Theorem 8.1.1. In this case*

$$P\Big(x, F(x)\Big) = \Big(F(x) - ax\Big)^2 = F(x)^2 - 2axF(x) + a^2x^2$$

and by Lemma 8.5.1, the fact $\int_0^1 P\Big(x, F(x)\Big)\,\mathrm{d}x = 0$ is determined by densities of words of length three.

8.6 Regularity lemma for words

In this section we give an alternative proof of Theorem 8.3.9 based on the regularity lemma for words introduced by Axenovich, Puzynina and Person in [12] to study the twins problem. For completeness, we give an (analytic) proof of the regularity lemma.

A measurable partition $\mathcal{P}$ of $[0,1]$ is a partition in which each atom is a measurable set of positive measure. Moreover, we say that $\mathcal{P}$ is an interval partition if every atom in $\mathcal{P}$ is a non-degenerate interval. In what follows, we will only consider measurable partitions with a finite number of atoms, and given a partition $\mathcal{P}$ we denote by $|\mathcal{P}|$ its number of atoms. Given two partitions $\mathcal{P}$ and $\mathcal{Q}$ we say that $\mathcal{Q}$ refines $\mathcal{P}$, which we denote by $Q \preceq P$, if for every $P \in \mathcal{P}$ there are atoms $Q_1, \dots, Q_k \in \mathcal{Q}$ such that $P = Q_1 \cup \dots \cup Q_k$. The common refinement of $\mathcal{P}$ and $\mathcal{Q}$ is the partition

$$\mathcal{P} \wedge \mathcal{Q} = \{A \cap B : A \in \mathcal{P}, B \in \mathcal{Q} \text{ such that } A \cap B \neq \emptyset\}.$$

Moreover, given a measurable set A we define the refinement of $\mathcal{P}$ by A as the common refinement of $\mathcal{P}$ and the partition $\{A, A^c\}$.

Let $f : [0,1] \to \mathbb{R}$ be a measurable function and let $\mathcal{P}$ be a partition. The conditional expectation of f with respect to $\mathcal{P}$ is the function $\mathbb{E}[f|\mathcal{P}]$ defined as

$$\mathbb{E}[f|\mathcal{P}](x) = \sum_{P\in\mathcal{P}} \frac{\mathbf{1}_{P(x)}}{\lambda(P)} \int_P f(t)\,\mathrm{d}t,$$

for all $x \in [0,1]$. The energy of $\mathcal{P}$ with respect to f is defined by

$$\mathcal{E}_f(\mathcal{P}) = \int_0^1 \Big(\mathbb{E}[f|\mathcal{P}](x)\Big)^2 \,\mathrm{d}x.$$

Note that $\mathcal{E}_f(\mathcal{P}) \leqslant \|f\|_\infty^2$. The following is a well known (easily derived) result about conditional expectations.

Lemma 8.6.1. *Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions such that $\mathcal{Q} \preceq \mathcal{P}$. Given any measurable function $f : [0,1] \to \mathbb{R}$, we have*

$$\int_0^1 \mathbb{E}[f|\mathcal{P}](t)\mathbb{E}[f|\mathcal{Q}](t)\,\mathrm{d}t = \int_0^1 \Big(\mathbb{E}[f|\mathcal{P}](t)\Big)^2 \,\mathrm{d}t.$$

Our next result shows that every $[0,1]$-valued measurable function over the interval $[0,1]$ can be approximated by a step function, which is supported on a partition of "bounded complexity" (a somewhat related result by Feige et al., the so called Local Repetition Lemma, was obtained in [47, Lemma 2.4]).

Theorem 8.6.2. *(Weak regularity lemma) Let $\varepsilon > 0$ and let $\mathcal{P}$ be an interval partition of $[0,1]$. For every measurable function $f : [0,1] \to [0,1]$ there exists an interval partition $\mathcal{P}_\varepsilon \preceq \mathcal{P}$ such that $\|f - \mathbb{E}[f|\mathcal{P}_\varepsilon]\|_\square \leqslant \varepsilon$ and $|\mathcal{P}_\varepsilon| \leqslant |\mathcal{P}| + 2\varepsilon^{-2}$.*

Proof. Set $\mathcal{P}_1 = \mathcal{P}$ and suppose that $\|f - \mathbb{E}[f|\mathcal{P}_1]\|_\square > \varepsilon$, as otherwise the result is trivial. For $k \geqslant 1$, assume we have defined a sequence of interval partitions $\mathcal{P}_k \preceq \cdots \preceq \mathcal{P}_1$ such that $\|f - \mathbb{E}[f|\mathcal{P}_k]\|_\square > \varepsilon$. This implies that there is an interval $I_{k+1} \notin \mathcal{P}_k$ such that

$$\left| \int_{I_{k+1}} (f - \mathbb{E}[f|\mathcal{P}_k])(t)\,\mathrm{d}t \right| > \varepsilon. \tag{8.21}$$

Define $\mathcal{P}_{k+1}$ as the smallest interval partition that contains the refinement of $\mathcal{P}_k$ by I_{k+1}. Since either I_{k+1} can split two distinct intervals of $\mathcal{P}_k$ into two subintervals each, or split a single interval of $\mathcal{P}_k$ into three subintervals, we have that $|\mathcal{P}_{k+1}| \leqslant |\mathcal{P}_k| + 2$. From (8.21) and by the Cauchy-Schwartz inequality, we deduce that

$$\begin{aligned}
\varepsilon^2 &< \left(\int_{I_{k+1}} \Big(\mathbb{E}[f|\mathcal{P}_{k+1}](t) - \mathbb{E}[f|\mathcal{P}_k](t)\Big)\,\mathrm{d}t \right)^2 \\
&\leqslant \int_0^1 (\mathbb{E}[f|\mathcal{P}_{k+1}](t) - \mathbb{E}[f|\mathcal{P}_k](t))^2\,\mathrm{d}t \\
&= \int_0^1 \Big(\mathbb{E}[f|\mathcal{P}_{k+1}](t)\Big)^2 \mathrm{d}t - \int_0^1 \Big(\mathbb{E}[f|\mathcal{P}_k](t)\Big)^2 \mathrm{d}t,
\end{aligned}$$

where the last equality follows from Lemma 8.6.1. Thus we have

$$1 \geqslant \|f\|_\infty^2 \geqslant \mathcal{E}_f(\mathcal{P}_{k+1}) \geqslant \mathcal{E}_f(\mathcal{P}_k) + \varepsilon^2,$$

and so, after at most ε^{-2} iterations, one finds some $\ell \leqslant \varepsilon^{-2}+1$ which satisfies $\|f - \mathbb{E}[f|\mathcal{P}_\ell]\|_\square \leqslant \varepsilon$. Since $|\mathcal{P}_k| \leqslant |\mathcal{P}_{k+1}| + 2$ for every $k \in [\ell]$, we get the claimed upper bound for $|\mathcal{P}_\ell|$. $\square$

Lemma 8.6.3 (Theorem 35.5 from [21])**.** *Let $f : [0,1] \to \mathbb{R}$ be an integrable function, and let $(\mathcal{P}_i)_{i\in\mathbb{N}}$ be a sequence of partitions such that $\mathcal{P}_{i+1} \preceq \mathcal{P}_i$ for all $i \in \mathbb{N}$. Then the sequence $(\mathbb{E}[f|\mathcal{P}_i])_{i\in\mathbb{N}}$ converges a.e. to $\mathbb{E}[f|\mathcal{P}_\infty]$, where $\mathcal{P}_\infty$ is the smallest σ-algebra containing each atom in $(\mathcal{P}_i)_{i\in\mathbb{N}}$.*

Now we are ready to provide an alternative proof of Theorem 8.3.9.

Proof (of Theorem 8.3.9). Let $(f_n)_{n\in\mathbb{N}}$ be any sequence in $\mathcal{W}$. By the Banach–Alaoglu theorem we may assume that $(f_n)_{n\in\mathbb{N}}$ converges weakly to some $f \in \mathcal{W}$. We claim that there are a collection of subsequences $(f_{n,k})_{n\in\mathbb{N}}$, for $k \in \mathbb{N}$, satisfying the following properties.

(i) $(f_{n,k})_{n\in\mathbb{N}}$ is a subsequence of $(f_{n,k-1})_{n\in\mathbb{N}}$, with $f_{n,0} = f_n$ for all $n \in \mathbb{N}$.

(ii) For $k \geqslant 2$, there is an interval partition $\mathcal{P}_k \preceq \mathcal{P}_{k-1}$ such that $|\mathcal{P}_k| \leqslant m_k$ and $\|f_{n,k} - \mathbb{E}[f_{n,k}|\mathcal{P}_k]\|_\square \leqslant \frac{1}{k}$ for every $n \in \mathbb{N}$.

(iii) For all $k \in \mathbb{N}$, the sequence $(\mathbb{E}[f_{n,k}|\mathcal{P}_k])_{n\in\mathbb{N}}$ converges a.e. to $f_k^* = \mathbb{E}[f|\mathcal{P}_k]$.

Assume we have constructed the sequence up to step k. We apply Theorem 8.6.2, with $\varepsilon_k = \frac{1}{k+1}$ and initial partition $\mathcal{P}_k$, to the sequence $(f_{n,k})_{n\in\mathbb{N}}$ so that for every $n \in \mathbb{N}$ we get an interval partition $\mathcal{P}_{n,k} \preceq \mathcal{P}_k$, with $|\mathcal{P}_{n,k}| \leqslant m_{k+1}$ for some positive integer m_{k+1} independent of n, and such that $\|f_{n,k} - \mathbb{E}[f_{n,k}|\mathcal{P}_{n,k}]\|_\square \leqslant \frac{1}{k+1}$. For $n \in \mathbb{N}$, let $J_{n,k} = \{a_{n,1} = 0 < \cdots < a_{n,\ell_n} = 1\}$ be the set of points that define the intervals of $\mathcal{P}_{n,k}$. Note that $\ell_n \leqslant m_{k+1}$. By the pigeonhole principle there is an integer $\ell \leqslant m_{k+1}$ and a subsequence $(f_{n,k+1})_{n\in\mathbb{N}}$ such that $\ell_n = \ell$ for all $n \in \mathbb{N}$. Moreover, since $[0,1]$ is compact we may even assume that $a_{n,i} \to a_i$ for each $i \in [\ell]$, where $a_1 = 0 \leqslant \ldots \leqslant a_\ell = 1$. Let $\mathcal{P}_{k+1} \preceq \mathcal{P}_k$ be the partition defined by $J_k = \{a_1 < \cdots < a_\ell\}$. Note that (i) and (ii) hold because of the definition of $(f_{n,k+1})_{n\in\mathbb{N}}$. Furthermore, because $\mathcal{P}_{k+1}$ is finite and since $(f_{n,k+1})_{n\in\mathbb{N}}$ converges weakly to f we conclude that (iii) also holds. On the other hand, by Lemma 8.6.3 we deduce that the sequence $(f_k^*)_{k\in\mathbb{N}}$ converges a.e. to $f_\infty = \mathbb{E}[f|\mathcal{P}_\infty]$. We claim that $\lim_{k\to\infty} d_\square(f_{k,k}, f_\infty) \to 0$. Indeed, Given $\varepsilon > 0$ by (ii), (iii) and the dominated convergence theorem, for large k we have

$$d_\square(f_\infty, f_{k,k}) \leqslant d_\square(f_\infty, f_k^*) + d_\square(f_{k,k}, \mathbb{E}[f_{k,k}|\mathcal{P}_k]) + d_\square(\mathbb{E}[f_{k,k}|\mathcal{P}_k], f_k^*) \leqslant \frac{\varepsilon}{3} + \frac{1}{k} + \frac{\varepsilon}{3} \leqslant \varepsilon.$$

□

8.7 Permutons from words limits

In this section we re-derive two key results proven by Hoppen et al. [64] concerning permutation sequences and show they can be obtained as consequences of our results concerning convergent word sequences. This leads to an alternative proof of the existence of permutons. Overall, our approach gives a simpler proof for the existence of permutons due mostly to the simpler objects (words and measurable transformations of the unit interval) on which our analysis is carried out, and the rather direct implication concerning permutons presented below. Moreover, we give a direct proof (avoiding compactness arguments) of the equivalence between t-convergence and convergence in the respective cut-distance, which we believe is both technically original and of independent interest.

First, recall that for $n \in \mathbb{N}$, we write $\mathfrak{S}_n$ for the set of permutations of order n and $\mathfrak{S}$ for the set of all finite permutations. Also, for $\sigma \in \mathfrak{S}_n$ and $\tau \in \mathfrak{S}_k$ we write $\Lambda(\tau, \sigma)$ for the number of copies of τ in σ, that is, the number of k-tuples $1 \leqslant x_1 < \cdots < x_k \leqslant n$ such that for every $i, j \in [k]$

$$\sigma(x_i) \leqslant \sigma(x_j) \qquad \text{iff} \qquad \tau(i) \leqslant \tau(j).$$

The density of copies of τ in σ, denoted by $t(\tau, \sigma)$, was defined as the probability that σ restricted to a randomly chosen k-tuple of $[n]$ yields a copy of τ, that is

$$t(\tau, \sigma) = \begin{cases} \binom{n}{k}^{-1} \Lambda(\tau, \sigma) & \text{if } n \geqslant k, \\ 0 & \text{otherwise.} \end{cases}$$

Following [64, Definition 1.2], a sequence $(\sigma_n)_{n \to \infty}$ of permutations, with $\sigma_n \in \mathfrak{S}_n$ for each $n \in \mathbb{N}$, is said to be convergent if $\lim_{n \to \infty} t(\tau, \sigma_n)$ exists for every permutation $\tau \in \mathfrak{S}$. A permuton is a probability measure μ on the Borel σ-algebra on $[0,1] \times [0,1]$ that has uniform marginals, that is, for every measurable set $A \subseteq [0,1]$ one has

$$\mu(A \times [0,1]) = \mu([0,1] \times A) = \lambda(A).$$

The collection of permutons is denoted by $\mathcal{Z}$. It turns out that every permutation may be identified with a permuton which preserves the sub-permutation densities. Indeed, given a permutation $\sigma \in \mathfrak{S}_n$ we define the permuton μ_σ associated to σ in the following way. First, for $i, j \in [n]$ define

$$B_{i,j} = B_i \times B_j \qquad \text{where} \qquad B_i = \begin{cases} \left[\frac{i-1}{n}, \frac{i}{n}\right) & \text{if } i \neq n, \\ \left[\frac{n-1}{n}, 1\right] & \text{otherwise.} \end{cases}$$

and note that $B_{i,j}$ has Lebesgue measure $\lambda^{(2)}(B_{i,j}) = \frac{1}{n^2}$ for every $i, j \in [n]$. For every measurable set $E \subseteq [0,1]^2$ we let

$$\mu_\sigma(E) = \sum_{i=1}^{n} n \lambda^{(2)}(B_{i,\sigma(i)} \cap E) = \int_E n \mathbf{1}\{\sigma(\lceil nx \rceil) = \lceil ny \rceil\} \,\mathrm{d}x \,\mathrm{d}y.$$

It is easy to see that $\mu_\sigma \in \mathcal{Z}$.

We next argue that the densities of sub-permutations is preserved by μ_σ. First, let us explain what we mean by sub-permutation densities for a permuton. Given $\mu \in \mathcal{Z}$ and $k \in \mathbb{N}$, we sample k points $(X_1, Y_1), \ldots, (X_k, Y_k)$, where each (X_i, Y_i) is sampled independently accordingly to μ. Then, if $\sigma, \pi \in \mathfrak{S}_k$ are two permutations such that

$$X_{\pi(1)} \leqslant \ldots \leqslant X_{\pi(k)} \quad \text{and} \quad Y_{\sigma(1)} \leqslant \ldots \leqslant Y_{\sigma(k)},$$

we define the random sub-permutation $\mathrm{sub}(k, \mu) \in \mathfrak{S}_k$ by $\mathrm{sub}(k, \mu) = \sigma \pi^{-1}$.

Henceforth, let $\mu^{(k)} = \mu \otimes \cdots \otimes \mu$ be the k-fold product measure on $([0,1] \times [0,1])^k$. Given a permutation $\tau \in \mathfrak{S}_k$, the density of τ in μ, denoted by $t(\tau, \mu)$, is defined as the probability that $\mathrm{sub}(k, \mu)$ is isomorphic to τ, that is

$$t(\tau, \mu) = k! \int \mathbf{1}\{x_1 < \cdots < x_k, y_{\tau^{-1}(1)} < \cdots < y_{\tau^{-1}(k)}\} \,\mathrm{d}\mu^{(k)}.$$

It is easily shown (see [64, Lemma 3.5] for a proof) that given any permutations $\sigma \in \mathfrak{S}_n$ and $\tau \in \mathfrak{S}_k$ we have

$$|t(\tau,\sigma) - t(\tau,\mu_\sigma)| \leqslant \binom{k}{2}\frac{1}{n}. \tag{8.22}$$

In particular, (8.22) implies that a sequence of permutations $(\sigma_n)_{n\to\infty}$ converges if and only if $(t(\tau,\mu_{\sigma_n}))_{n\to\infty}$ is convergent for every permutation $\tau \in \mathfrak{S}$, and thus we may talk about permutations and permutons as the "same" object. We say that a sequence of permutons $(\mu_n)_{n\to\infty}$ is t-convergent if $(t(\tau,\mu_n))_{n\to\infty)}$ converges for every $\tau \in \mathfrak{S}$.

As in the case of words one can define a metric $d_\square$ on $\mathcal{Z}$ so that for all $\tau \in \mathfrak{S}$ the maps $t(\tau,\cdot)$ are Lipschitz continuous with respect to $d_\square$. Indeed, given two permutons $\mu,\nu \in \mathcal{Z}$ define

$$d_\square(\mu,\nu) = \sup_{I,J\subseteq[0,1]} |\mu(I\times J) - \nu(I\times J)|,$$

where the supremum is taken over all intervals in $[0,1]$. In order to prove that $t(\tau,\cdot)$ is Lipschitz continuous with respect to $d_\square$ we need the following result which is the permuton analogue of Lemma 8.3.2.

Lemma 8.7.1. *Given a permutation $\tau \in \mathfrak{S}_k$, for all permutons $\mu,\nu \in \mathcal{Z}$ we have*

$$|t(\tau,\mu) - t(\tau,\nu)| \leqslant k^2 d_\square(\mu,\nu).$$

Proof. Define

$$E^\tau = \{(\vec{x},\vec{y}) \in [0,1]^k \times [0,1]^k : x_1 < \cdots < x_k, y_{\tau^{-1}(1)} < \cdots < y_{\tau^{-1}(k)}\}. \tag{8.23}$$

Then, we have $t(\tau,\mu) = k!\mu^{(k)}(E^\tau)$ and $t(\tau,\nu) = k!\nu^{(k)}(E^\tau)$. For $j \in [k]$, let

$$Q_j = \mu^{(j)} \otimes \nu^{(k-j)} - \mu^{(j-1)} \otimes \nu^{(k-j+1)}$$

and note that

$$\frac{1}{k!}|t(\tau,\mu) - t(\tau,\nu)| = |\mu^{(k)}(E^\tau) - \nu^{(k)}(E^\tau)| = \left|\sum_{j=1}^k Q_j(E^\tau)\right| \leqslant \sum_{j=1}^k |Q_j(E^\tau)|.$$

Let $j \in [k]$ be fixed. Given $(\vec{x},\vec{y})$, let $E_j^\tau(\vec{x},\vec{y}) = [x_{j-1},x_{j+1}] \times [y_{\tau^{-1}(j-1)}, y_{\tau^{-1}(j+1)}]$ if $x_1 < \cdots < x_{j-1} < x_{j+1} < \cdots < x_k$ and $y_{\tau^{-1}(1)} < \cdots < y_{\tau^{-1}(j-1)} < y_{\tau^{-1}(j+1)} < \cdots < y_{\tau^{-1}(k)}$, and $E_j^\tau(\vec{x},\vec{y}) = \emptyset$ otherwise. Thus $\left|\mu(E_j^\tau(\vec{x},\vec{y})) - \nu(E_j^\tau(\vec{x},\vec{y}))\right| \leqslant d_\square(\mu,\nu)$ for all $(\vec{x},\vec{y})$ and then, we have that

$$\begin{aligned}
|Q_j(E^\tau)| &= \left|\int \Big(\mu(E_j^\tau(\vec{x},\vec{y})) - \nu(E_j^\tau(\vec{x},\vec{y}))\Big)\,\mathrm{d}\mu^{(j-1)} \otimes \nu^{(k-j)}\right| \\
&\leqslant \int \left|\mu(E_j^\tau(\vec{x},\vec{y})) - \nu(E_j^\tau(\vec{x},\vec{y}))\right| \mathrm{d}\mu^{(j-1)} \otimes \nu^{(k-j)} \\
&\leqslant \int_{x_1<\cdots<x_{j-1}<x_{j+1}<\cdots<x_k} \left|\mu(E_j^\tau(\vec{x},\vec{y})) - \nu(E_j^\tau(\vec{x},\vec{y}))\right| \mathrm{d}\mu^{(j-1)} \otimes \nu^{(k-j)} \\
&\leqslant \frac{1}{(k-1)!} d_\square(\mu,\nu).
\end{aligned}$$

Finally, summing for each $j \in [k]$ we obtain the bound. $\square$

In Hoppen et al. [64], the compactness of $(\mathcal{Z}, d_\square)$ is established and, as a consequence, also the equivalence between t-convergence and convergence in $d_\square$. In particular, they prove that for every convergent sequence of permutations $(\sigma_n)_{n\to\infty}$ there is a permuton $\mu \in \mathcal{Z}$ such that $t(\tau, \sigma_n) \to t(\tau, \mu)$ for all $\tau \in \mathfrak{S}$. The goal of this section is to give a new proof of these two results by using a more direct approach based on Theorem 8.1.3 and the permuton analogue of Proposition 8.3.6 based on Bernstein polynomials.

We start with a permuton analogue of Lemma 8.3.1. A similar result was proved by Glebov, Grzesik, Klimošová and Král' [53, Theorem 3] by using a probabilistic interpretation.

Lemma 8.7.2. *Let $\mu \in \mathcal{Z}$ be a permuton and let $i, j \in \mathbb{N}$. There exist a set $S_{i,j}$ of permutations of order $i + j + 1$ and positive numbers $(C_\tau)_{\tau\in S_{i,j}}$ such that*

$$\int_{[0,1]^2} x^i y^j \,\mathrm{d}\mu(x, y) = \sum_{\tau\in S_{i,j}} C_\tau t(\tau, \mu).$$

Proof. We proceed as in the proof of Lemma 8.3.1. First, since μ has uniform marginals we have that

$$x^i = \Big(\int_{[0,x]\times[0,1]} \mathrm{d}\mu(x', y')\Big)^i = \int_{[0,1]^{2i}} \mathbf{1}\{x_1, \dots, x_i \leqslant x\} \,\mathrm{d}\mu(x_1, y_1) \dots \mathrm{d}\mu(x_i, y_i)$$

and similarly

$$y^j = \int_{[0,1]^{2j}} \mathbf{1}\{y_{i+1}, \dots, y_{i+j} \leqslant y\} \,\mathrm{d}\mu(x_{i+1}, y_{i+1}) \dots \mathrm{d}\mu(x_{i+j}, y_{i+j}).$$

Whence, setting

$$G_U(\vec{x}, x) = \mathbf{1}\{x_1, \dots, x_i \leqslant x\} \prod_{u\in U} \mathbf{1}\{x_{i+u} \leqslant x\} \prod_{u\notin U} \mathbf{1}\{x \leqslant x_{i+u}\}$$

and

$$H_S(\vec{y}, y) = \mathbf{1}\{y_{i+1}, \dots, y_{i+j} \leqslant y\} \prod_{s\in S} \mathbf{1}\{y_s \leqslant y\} \prod_{s\notin S} \mathbf{1}\{y \leqslant y_s\},$$

by the Fubini–Tonelli theorem, we have

$$\begin{aligned} x^i y^j &= \int_{[0,1]^{2(i+j)}} \mathbf{1}\{x_1, \dots, x_i \leqslant x\}\mathbf{1}\{y_{i+1}, \dots, y_{i+j} \leqslant y\} \,\mathrm{d}\mu^{(i+j)}(\vec{x}, \vec{y}) \\ &= \sum_{U\subseteq[j]} \sum_{S\subseteq[i]} \int_{[0,1]^{2(i+j)}} G_U(\vec{x}, x) H_S(\vec{y}, y) \,\mathrm{d}\mu^{(i+j)}(\vec{x}, \vec{y}). \end{aligned}$$

Finally, by reordering the position of the coordinates below and above x, respectively, we have

$$\int_{[0,1]^2} x^i y^j \,\mathrm{d}\mu(x, y) = \sum_{k\in[j]} \sum_{\ell\in[i]} \binom{j}{k}\binom{i}{\ell} \frac{(i + k)!(j - k)!}{(i + j + 1)!} \sum_{\sigma\in\mathfrak{S}_{i+j+1}:\sigma(i+k+1)\geqslant j+1} t(\sigma, \mu).$$

$\square$

As pointed out in [75], the previous result can be used to prove the uniqueness of the limit of a sequence of permutations as we did for limits of words by using Lemma 8.3.1. Indeed, suppose that $\mu, \nu \in \mathcal{Z}$ are two permutons such that $t(\sigma, \mu) = t(\sigma, \nu)$ for every finite permutation $\sigma \in \mathfrak{S}$. By Lemma 8.7.2 we deduce that for every continuous function $h : [0,1]^2 \to \mathbb{R}$ we have

$$\int_{[0,1]^2} h(x,y)\,\mathrm{d}\mu(x,y) = \int_{[0,1]^2} h(x,y)\,\mathrm{d}\nu(x,y),$$

which implies that $\mu = \nu$. On the other hand, Lemma 8.7.2 can also be used to establish the permuton analogue of Proposition 8.3.6, that t-convergence implies the convergence with respect to $d_\square$.

Proposition 8.7.3. *If $(\mu_n)_{n\to\infty}$ is a sequence in $\mathcal{Z}$ which is t-convergent, then it is a Cauchy sequence with respect to $d_\square$. Moreover, if $\mu_n \xrightarrow{t} \mu$ for some $\mu \in \mathcal{Z}$, then $\mu_n \xrightarrow{\square} \mu$.*

Proof. Let $\varepsilon > 0$ be fixed and let $r = \lceil (80/\varepsilon)^4 \rceil$. Let $S_{i,j} \subseteq \mathfrak{S}_{i+j+1}$ and C_τ be as in the statement of Lemma 8.7.2, define $C = \max\{C_\tau : \tau \in S_{i,j}, i, j \leqslant r\}$, and let

$$\delta = \frac{\varepsilon}{C(2r+1)!2^{4r+3}}.$$

Let n_0 be sufficiently large so that for all $n, m \geqslant n_0$ we have

$$|t(\tau,\mu_n) - t(\tau,\mu_m)| \leqslant \delta \qquad \text{for all } \tau \in \bigcup_{i\in[r]} \mathfrak{S}_i. \tag{8.24}$$

Hence, for $i, j \leqslant r$ and $\nu = \mu_n - \mu_m$, by Lemma 8.7.2 and since $|\mathfrak{S}_{i+j+1}| \leqslant (2r+1)!$, we have

$$\left| \int_{[0,1]^2} x^i y^j \,\mathrm{d}\nu(x,y) \right| = \left| \sum_{\tau \in S_{i,j}} C_\tau (t(\tau,\mu_n) - t(\tau,\mu_m)) \right| \leqslant C(2r+1)!\delta.$$

For $a, b \in [0,1]$, let $J_{a,b} = \mathbf{1}_{[0,a]\times[0,b]}$ and let j_a, j_b be the largest indices such that $\frac{j_a}{r} \leqslant a$ and $\frac{j_b}{r} \leqslant b$. Recall that the Bernstein polynomial of $J_{a,b}$ is denoted by $B_{r,J_{a,b}}$ and observe that

$$\begin{aligned}
\left| \int B_{r,J_{a,b}}(x,y)\,\mathrm{d}\nu(x,y) \right| &\leqslant \sum_{i=0}^{i_a} \sum_{j=0}^{j_b} \binom{r}{i}\binom{r}{j} \left| \int x^i (1-x)^{r-i} y^j (1-y)^{r-j} \,\mathrm{d}\nu(x,y) \right| \\
&\leqslant \sum_{0\leqslant i,j \leqslant r} \sum_{k=0}^{r-i} \sum_{\ell=0}^{r-j} \binom{r}{i}\binom{r}{j}\binom{r-i}{k}\binom{r-j}{\ell} \left| \int x^{i+k} y^{j+\ell} \,\mathrm{d}\nu(x,y) \right| \\
&\leqslant C2^{4r}(2r+1)!\delta.
\end{aligned}$$

Now, by Lemma 8.3.5 we have

$$\begin{aligned}
|\nu([0,a]\times[0,b])| &= \left| \int \mathbf{1}_{[0,a]\times[0,b]}(x,y)\,\mathrm{d}\nu(x,y) \right| \\
&\leqslant \left| \int B_{r,J_{a,b}}(x,y)\,\mathrm{d}\nu(x,y) \right| + \left| \int (\mathbf{1}_{[0,a]\times[0,b]}(x,y) - B_{r,J_{a,b}}(x,y))\,\mathrm{d}\nu(x,y) \right| \\
&\leqslant C2^{4r}(2r+1)!\delta + (8r^{-1/4} + 2r^{-1/2}),
\end{aligned}$$

where the last inequality follows since μ_n and μ_m have uniform marginals. Putting everything together, by our choice of r, δ and ν, we have

$$d_\square(\mu_n, \mu_m) \leqslant 4 \sup_{a,b\in[0,1]} |\nu([0,a] \times [0,b])| \leqslant C2^{4r+2}(2r+1)!\delta + 40r^{-1/4} \leqslant \varepsilon.$$

For the second part just replace μ_m by μ in (8.24) and choose $\nu = \mu_n - \mu$. Then, repeat the above argument. $\square$

We can now give the alternative proof of the result of Hoppen et al. [64] concerning the existence of a limit (permuton) for a convergent permutation sequence. Note that this limit is unique as discussed right after the proof of Lemma 8.7.2.

Theorem 8.7.4 (Hoppen et al. [64, Theorem 1.6])**.** *For every convergent sequence of permutations $(\sigma_n)_{n\to\infty}$ there exists a permuton $\mu \in \mathcal{Z}$ such that $\sigma_n \xrightarrow{t} \mu$.*

Proof. Let $(\sigma_n)_{n\to\infty}$ be given and let $(\mu_n)_{n\to\infty}$ be the sequence of corresponding permutons. Given $x \in [0,1]$ and $n \in \mathbb{N}$, we define

$$f_{n,x}(y) = \int_0^x n\mathbf{1}\{\sigma_n(\lceil nt\rceil) = \lceil ny\rceil\}\,\mathrm{d}t \quad \text{ for all } y \in [0,1].$$

It is easy to see that

(i) $f_{n,x}(\cdot) \leqslant f_{n,x'}(\cdot)$ for all $x \leqslant x'$,

(ii) $f_{n,0}(\cdot) = 0$ for all $n \in \mathbb{N}$, and

(iii) $f_{n,1}(\cdot) = 1$ for all $n \in \mathbb{N}$.

We claim that $(f_{n,x})_{n\to\infty}$ converges for all $x \in [0,1]$. Indeed, by Proposition 8.7.3, $(\mu_n)_{n\to\infty}$ is a Cauchy sequence with respect to $d_\square$, and for every interval $I \subseteq [0,1]$

$$\left|\int_I (f_{n,x}(t) - f_{m,x}(t))\,\mathrm{d}t\right| = \Big|\mu_n([0,x] \times I) - \mu_m([0,x] \times I)\Big| \leqslant d_\square(\mu_n, \mu_m).$$

Thus $(f_{n,x})_{n\to\infty}$ is a Cauchy sequence in $(\mathcal{W}, d_\square)$ and therefore, by Theorem 8.3.9, it has a limit $f_x \in \mathcal{W}$. Furthermore, note that for all $x \in [0,1]$ we have

$$\int_0^1 f_x(t)\,\mathrm{d}t = \lim_{n\to\infty} \int_0^1 f_{n,x}(t)\,\mathrm{d}t = \lim_{n\to\infty} \frac{\lceil nx\rceil}{n} = x \tag{8.25}$$

and, because of (i), for all $a, x, x' \in [0,1]$,

$$\left|\int_0^a f_x(t)\,\mathrm{d}t - \int_0^a f_{x'}(t)\,\mathrm{d}t\right| \leqslant |x - x'|. \tag{8.26}$$

Given $s \in [0,1]$ and given an interval $I \subseteq [0,1]$, we set

$$\tilde{\mu}([0,s] \times I) = \int_I f_s(t)\,\mathrm{d}t.$$

Because of (i), (ii) and (iii), $\tilde{\mu}$ is well defined and so by standard limiting arguments we can extend $\tilde{\mu}$ to a unique probability measure μ on $[0,1]\times[0,1]$. Observe that because of (iii) we have that $f_1(\cdot)=1$ almost everywhere. This together with (8.25) imply that μ has uniform marginals and therefore $\mu\in\mathcal{Z}$. To conclude that $\sigma_n \xrightarrow{t} \mu$, by Lemma 8.7.1, it is enough to show that $d_\square(\sigma_n,\mu)\to 0$. If not, then there are $\varepsilon>0$ and sequences $(x_n)_{n\to\infty}$ and $(a_n)_{n\to\infty}$ such that, without loss of generality, for all n sufficiently large we have

$$\int_0^{a_n} f_{n,x_n}(t)\,\mathrm{d}t \geqslant \mu([0,x_n]\times[0,a_n])+\varepsilon=\int_0^{a_n} f_{x_n}(t)\,\mathrm{d}t+\varepsilon.$$

Moreover, because of (8.26) and by compactness of $[0,1]$ we can find $a,x\in[0,1]$ such that (passing to a subsequence) for all n sufficiently large we have

$$\int_0^{a} f_{n,x}(t)\,\mathrm{d}t \geqslant \int_0^{a} f_{x}(t)\,\mathrm{d}t+\frac{\varepsilon}{2},$$

contradicting the fact that $(f_{n,x})_{n\to\infty}$ converges to f_x. $\square$

8.8 Non-binary words.

Let Σ be a finite alphabet. For a word $\boldsymbol{w}\in\Sigma^n$ and an interval $I\subseteq[n]$ let $N_a(\boldsymbol{w},I)$ denote the number of occurrences of $a\in\Sigma$ in $\mathrm{sub}(I,\boldsymbol{w})$ and let $N_a(\boldsymbol{w})=N_a(\boldsymbol{w},[n])$. Moreover, as for the binary alphabet case, denote by $\binom{\boldsymbol{w}}{\boldsymbol{u}}$ the number of subsequences of $\boldsymbol{w}$ which coincide with $\boldsymbol{u}$. A sequence $(\boldsymbol{w}_n)_{n\to\infty}$ of words $\boldsymbol{w}_n\in\Sigma^n$ is called $o(1)$-uniform if for each $a\in\Sigma$ there is a density d_a such that $N_a(\boldsymbol{w}_n,I)=d_a|I|+o(1)n$ holds for each interval $I\subseteq[n]$. We obtain the following analogue (generalization) of Theorem 8.1.3 for finite size alphabets.

Theorem 8.8.1. *Given a sequence $(\boldsymbol{w}_n)_{n\to\infty}$ of words $\boldsymbol{w}_n\in\Sigma^n$ over the finite size alphabet Σ. If $(\boldsymbol{w}_n)_{n\to\infty}$ is $o(1)$-uniform, then for each $a\in\Sigma$ there is a density $d_a\in[0,1]$ such that for every $\ell\in\mathbb{N}$ and every word $\boldsymbol{u}\in\Sigma^\ell$ we have $\binom{\boldsymbol{w}_n}{\boldsymbol{u}}=\prod_{a\in\Sigma} d_a^{N_a(\boldsymbol{u})}\binom{n}{\ell}+o(n^\ell)$. Conversely, if for some collection of densities $\{d_a\in[0,1]:a\in\Sigma\}$ we have $\binom{\boldsymbol{w}_n}{\boldsymbol{u}}=\prod_{a\in\Sigma} d_a^{N_a(\boldsymbol{u})}\binom{n}{3}+o(n^3)$ for all words $\boldsymbol{u}\in\Sigma^3$, then $(\boldsymbol{w}_n)_{n\to\infty}$ is $o(1)$-uniform.*

Proof. The first part of the theorem follows from Remark 8.3.3 by an argument similar to the one used in the proof of the first part of Lemma 8.3.2. For the second part, let us consider a letter $a\in\Sigma$ and a word $\boldsymbol{w}$ over Σ. We define the binary word $\boldsymbol{w}^a$ as the word obtained by replacing each letter a in $\boldsymbol{w}$ by 1 and the remaining letters by 0. Moreover, for $\boldsymbol{u}\in\{0,1\}^\ell$ we let $\Sigma_a(\boldsymbol{u})$ be the set of words $\boldsymbol{v}\in\Sigma^\ell$ such that $\boldsymbol{v}^a=\boldsymbol{u}$. Then, it is easy to see that

$$t(\boldsymbol{u},\boldsymbol{w}^a)=\sum_{\boldsymbol{v}\in\Sigma_a(\boldsymbol{u})} t(\boldsymbol{v},\boldsymbol{w}). \tag{8.27}$$

For each $a\in\Sigma$ we can thus define the sequence $(\boldsymbol{w}_n^a)_{n\to\infty}$ of words over the alphabet $\{0,1\}$ which, because of (8.27), satisfies the counting property for subsequences of length 3. From Theorem 8.1.1 and our working hypothesis we conclude that $(\boldsymbol{w}_n^a)_{n\to\infty}$ is $o(1)$-uniform over the alphabet $\{0,1\}$ and thus we deduce that $N_a(\boldsymbol{w}_n,I)=N_1(\boldsymbol{w}_n^a,I)=d_a|I|+o(1)n$ for all intervals $I\subseteq[n]$. By repeating the above argument for each letter in Σ we conclude that $(\boldsymbol{w}_n)_{n\to\infty}$ is $o(1)$-uniform. $\square$

Similarly, one can obtain an analogue of Theorem 8.1.3 concerning limits of convergent word sequences for larger alphabets. A sequence $(\boldsymbol{w}_n)_{n\to\infty}$ of words over the alphabet $\Sigma = \{a_1, \ldots, a_k\}$ is convergent if for all $\ell \in \mathbb{N}$ and $\boldsymbol{u} \in \Sigma^\ell$ the subsequence density $\left(\binom{\boldsymbol{w}_n}{\boldsymbol{u}}/\binom{n}{\ell}\right)_{n\to\infty}$ converges. Moreover, given a k-tuple of functions $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k}) \in \mathcal{W}^k$ such that $f^{a_1}(x) + \cdots + f^{a_k}(x) = 1$ for almost all $x \in [0, 1]$, we say that $(\boldsymbol{w}_n)_{n\to\infty}$ converges to $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k})$ if for all $\ell \in \mathbb{N}$ and $\boldsymbol{u} \in \Sigma^\ell$ the subsequence density $\left(\binom{\boldsymbol{w}_n}{\boldsymbol{u}}/\binom{n}{\ell}\right)_{n\to\infty}$ converges to

$$t(\boldsymbol{u}, \boldsymbol{f}) = \ell! \int_{0 \leqslant x_1 < \cdots < x_\ell \leqslant 1} \prod_{i \in [\ell]} f^{u_i}(x_i) \, \mathrm{d}x_1 \ldots \mathrm{d}x_\ell.$$

For the case of non-binary alphabets, we obtain the following limit theorem.

Theorem 8.8.2 (Limits of convergent k-letter word sequences)**.** *Let* $\Sigma = \{a_1, \ldots, a_k\}$.

- *Each convergent sequence* $(\boldsymbol{w}_n)_{n\to\infty}$ *of words,* $\boldsymbol{w}_n \in \Sigma^n$*, converges to some vector* $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k}) \in \mathcal{W}^k$ *and* $f^{a_1}(x) + \cdots + f^{a_k}(x) = 1$ *for almost all* $x \in [0, 1]$*. Moreover, if* $(\boldsymbol{w}_n)_{n\to\infty}$ *converges to* $\boldsymbol{g} = (g^{a_1}, \ldots, g^{a_k})$*, then* $f^{a_i} = g^{a_i}$ *almost everywhere, for all* $i \in [k]$*.*
- *Conversely, for every vector* $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k}) \in \mathcal{W}^k$ *which satisfies* $f^{a_1}(x) + \cdots + f^{a_k}(x) = 1$ *for almost all* $x \in [0, 1]$ *there is a sequence* $(\boldsymbol{w}_n)_{n\to\infty}$ *of words* $\boldsymbol{w}_n \in \Sigma^n$ *which converges to* $\boldsymbol{f}$*.*

Proof. The first part follows by reducing to the size two alphabet case. Indeed, fix $a_i \in \Sigma$. For each $n \in \mathbb{N}$ we define the word $\boldsymbol{w}_n^{a_i}$ as before and thus we obtain a sequence $(\boldsymbol{w}_n^{a_i})_{n\to\infty}$ of words over the binary alphabet, which we claim is convergent. Indeed, since $(\boldsymbol{w}_n)_{n\to\infty}$ is convergent then each term in the RHS in (8.27) is convergent and thus $(t(\boldsymbol{u}, \boldsymbol{w}_n^{a_i}))_{n\to\infty}$ is convergent. Therefore, Theorem 8.1.3 implies that $(\boldsymbol{w}_n^{a_i})_{n\to\infty}$ converges to a (unique) $f^{a_i} \in \mathcal{W}$. In particular, the sequence $(f_n^{a_i})_{n\to\infty}$ of functions associated to $(\boldsymbol{w}_n^{a_i})_{n\to\infty}$ satisfies $f_n^{a_i} \xrightarrow{t} f^{a_i}$ and Proposition 8.3.6 implies that $f_n^{a_i} \xrightarrow{\square} f^{a_i}$ as well. The argument shown in Lemma 8.3.2 (see Remark 8.3.3) then yields that $(\boldsymbol{w}_n)_{n\to\infty}$ converges to $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k})$ and it is not hard to see that $f^{a_1}(x) + \cdots + f^{a_k}(x) = 1$ for almost all $x \in [0, 1]$.

To prove the second part, we exhibit a sequence of words which converges to a given $\boldsymbol{f} = (f^{a_1}, \ldots, f^{a_k})$. Consider the $\boldsymbol{f}$-random letter $(X, Y) \in [0, 1] \times \Sigma$ obtained by choosing X uniformly in $[0, 1]$ and, conditioned on $X = x$, choosing Y to be $a_i \in \Sigma$ with probability $f^{a_i}(x)$. Next, for each positive integer n choose $\boldsymbol{f}$-random letters $(X_1, Y_1), \ldots (X_n, Y_n)$ and a permutation $\sigma : [n] \to [n]$ such that $X_{\sigma(1)} \leqslant \ldots \leqslant X_{\sigma(n)}$. Then, define the $\boldsymbol{f}$-random word $\boldsymbol{w}_n = Y_{\sigma(1)} \ldots Y_{\sigma(n)}$. By fixing a letter $a_i \in \Sigma$ and replacing the $\boldsymbol{w}_n$'s by $\boldsymbol{w}_n^{a_i}$'s as above we obtain a sequence of f^{a_i}-random words over size two alphabets whose associated functions converge in the interval-norm to f^{a_i} a.s. due to Corollary 8.3.11. Then, Lemma 8.3.2 and Remark 8.3.3 imply that the $\boldsymbol{f}$-random word sequence converges to $\boldsymbol{f}$. $\square$

Chapter 9

Future perspectives

To conclude this part of the thesis, we discuss some potential future research directions.

9.1 Longest common subsequence for generalised random words

As mentioned in Chapter 8, given a word limit $f \in \mathcal{W}$ one can define an n-letter random word $\text{sub}(n, f)$ by sampling n points $x_1 < \cdots < x_n$ from $[0, 1]$ and then setting the i-th letter of $\text{sub}(n, f)$ as a Bernoulli random variable with mean $f(x_i)$. It is thus natural to define the longest common subsequence problem in this new random word model.

For $f_1, f_2 \in \mathcal{W}$, we write $\text{LCS}(f_1, f_2, n)$ to denote the random variable which is equal to the length of the longest common subsequence between $\text{sub}(n, f_1)$ and $\text{sub}(n, f_2)$. If $f_1 = f_2 = f$, then we write $\text{LCS}(f, n)$ instead of $\text{LCS}(f_1, f_2, n)$. We observe that if $f_1(x) = f_2(x) = \frac{1}{2}$ for all $x \in [0, 1]$, then $\text{LCS}(\frac{1}{2}, n) = \text{LCS}_2(n)$ for all $n \in \mathbb{N}$. On the other hand, it does not seem possible to adapt the sub-additivity argument to prove that $\frac{1}{n}\text{LCS}(f_1, f_2, n)$ converges for arbitrary $f_1, f_2 \in \mathcal{W}$. Therefore, it would be possible that $\text{LCS}(f_1, f_2, n) = o(n)$ even if $f_1 = f_2$.

Problem 9.1.1. *Characterise the functions $f \in \mathcal{W}$ such that $LCS(f, n) = \Omega(n)$.*

9.2 Turán numbers for words

Recall that for a graph H, the Turán function $\text{ex}(n, H)$ of H is defined as the maximum number of edges that a n-vertex H-free graph can have. The study of the Turán numbers is one of the central topics in extremal graph theory, however, as far as we know no such concept has been studied for words. A quite interesting problem is to study the following Turán-type problem for words (an analogue problem may be defined for permutations as well). Given two fixed words $\boldsymbol{u}$ and $\boldsymbol{v}$, we define the Turán function of the pair $(\boldsymbol{u}, \boldsymbol{v})$ as

$$\mathrm{ex}(n, \boldsymbol{u}, \boldsymbol{v}) = \max\{\tbinom{\boldsymbol{w}}{\boldsymbol{u}} : \boldsymbol{w} \in \{0,1\}^n \text{ and } t(\boldsymbol{v}, \boldsymbol{w}) = 0\}, \tag{9.1}$$

that is, $\mathrm{ex}(n, \boldsymbol{u}, \boldsymbol{v})$ is the maximum number of possible copies of $\boldsymbol{u}$ in a word of length n which does not contain a copy of $\boldsymbol{v}$. We conjecture that if $\boldsymbol{u}$ does not contain $\boldsymbol{v}$, then the unique maximiser for (9.1) should be a *blow-up* of $\boldsymbol{u}$. An even more challenging problem is to study the function

$$\mathrm{ex}_\alpha(n, \boldsymbol{u}, \boldsymbol{v}) = \max\{\tbinom{\boldsymbol{w}}{\boldsymbol{u}} : \boldsymbol{w} \in \{0,1\}^n \text{ and } t(\boldsymbol{v}, \boldsymbol{w}) \geqslant \alpha\} \tag{9.2}$$

for some fixed $\alpha \in [0,1]$. We remark here that both problems are quite natural. Indeed, let us define the continuous version of (9.2) as

$$\mathrm{ex}_\alpha(\boldsymbol{u}, \boldsymbol{v}) = \sup\{t(\boldsymbol{u}, f) : f \in \mathcal{W} \text{ and } t(\boldsymbol{v}, f) \geqslant \alpha\}. \tag{9.3}$$

Since $t(\boldsymbol{u}, \cdot)$, $t(\boldsymbol{v}, \cdot)$ are continuous and $(\mathcal{W}, d_\square)$ is compact, we know that the set $\{f \in \mathcal{W} : t(\boldsymbol{v}, f) \geqslant \alpha\}$ is compact too and so (9.3) is indeed a maximum. Therefore, in some sense the extremal functions (9.1) and (9.2) are well defined, and thus one could hope to study the Turán problem for words with analytical and elementary tools.

9.3 Twins in d-arrays

For $d \geqslant 1$ and a d-array $A = (a_{i_1,\ldots,i_d})$ of size n over an alphabet Σ, let $\mathrm{LT}(A)$ denote the largest m such that there are indices $\mathcal{I} = \{(i^{(1)}_{j_1}, \ldots, i^{(d)}_{j_d}) \in [n]^d : (j_1, \ldots, j_d) \in [m]^d\}$ and $\mathcal{L} = \{(\ell^{(1)}_{j_1}, \ldots, \ell^{(d)}_{j_d}) \in [n]^d : (j_1, \ldots, j_d) \in [m]^d\}$, where $1 \leqslant i^{(k)}_1 < \cdots < i^{(k)}_m \leqslant n$ and $1 \leqslant \ell^{(k)}_1 < \cdots < \ell^{(k)}_m \leqslant n$ are increasing sequences for each $k \in [d]$, such that

$$a_{i^{(1)}_{j_1},\ldots,i^{(d)}_{j_d}} = a_{\ell^{(1)}_{j_1},\ldots,\ell^{(d)}_{j_d}} \text{ for all } (j_1, \ldots, j_d) \in [m]^d.$$

The arrays $(a_{i^{(1)}_{j_1},\ldots,i^{(d)}_{j_d}})_{\vec{j}\in[m]^d}$ and $(a_{\ell^{(1)}_{j_1},\ldots,\ell^{(d)}_{j_d}})_{\vec{j}\in[m]^d}$ are called *twins*. For $n, d \in \mathbb{N}$ and an alphabet Σ, we define the function

$$\mathrm{LT}_d(n, \Sigma) = \min\{\mathrm{LT}(A) : A \in \mathcal{A}_d(\Sigma, n)\}.$$

Thus, by definition, every d-array of size n over Σ contains twins of size $\mathrm{LT}_d(n, \Sigma)$. As mentioned in the introduction, for $d = 1$ we know that $\mathrm{LT}_1(n, [q]) = \Omega(n)$ for all $q \geqslant 2$. It would be interesting to study the function $\mathrm{LT}_d(n, \Sigma)$ for $d \geqslant 2$.

Problem 9.3.1. *For $q, d \geqslant 2$. Prove or disprove that $LT_d(n, [q]) = \Omega(n)$.*

9.4 Universality of permutations

For $k \geqslant 2$, a *k-universal* permutation is one that contains all permutations of $\mathfrak{S}_k$. The question of the minimal n such that there exists a k-universal permutation in $\mathfrak{S}_n$ was asked by Arratia [11], who conjectured that the optimal value of n, given k, is $(1 + o(1))k^2/e^2$. The random version of this problem was posed by Alon (see [11]) who conjectured that a random permutation of order $(1 + o(1))k^2/4$ is k-universal with high probability. If true, this bound would be tight, as can be deduced from the known results on the length of the *longest increasing subsequence* of random permutations. The best known upper bound for this problem is due to Xe and Kwan [63], who recently proved that a random permutation on $O(k^2 \log\log k)$ elements is k-universal with high probability.

The study of higher dimensional permutations is ripe for further research. A *line* of a d-array $A = (a_{i_1,\dots,i_d})$ of order n is a sequence of elements obtained by choosing some $j \in [n]$ and looking at the entries $a_{i_1,\dots,i_{j-1},\ell,i_{j+1},\dots,i_d}$, for some fixed $i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_d \in [n]$ and ℓ ranging from 1 to n. Just as a usual permutation can be identified with a permutation matrix, it is possible to define a *d-dimensional permutation* (henceforth, *d-permutation*) of order n as a $(d+1)$-array of order n over $\{0, 1\}$, where each line contains a unique 1 entry (see [79, 80] for equivalent definitions and discussion).

Looking for connections with the case of permutations, we propose the following notion of "universality" for d-permutations. A *d-pattern* of order k is a sequence $(\sigma_1, \dots, \sigma_d)$ where $\sigma_\ell \in \mathfrak{S}_k$ for all $\ell \in [d]$. We say a d-permutation M of order n *contains* a d-pattern of order k if there exists a sequence $x^{(1)}, \dots, x^{(k)} \in [n]^{d+1}$ of index vectors such that $M_{x_1^{(i)} x_2^{(i)} \cdots x_{d+1}^{(i)}} = 1$ for all $i \in [k]$, $x_1^{(1)} < x_1^{(2)} < \cdots < x_1^{(k)}$ (the first coordinates of the vectors are increasing), and further, for each $\ell \in [d]$ and all $i, j \in [k]$, it holds that $x_{\ell+1}^{(i)} < x_{\ell+1}^{(j)}$ if and only if $\sigma_\ell(i) < \sigma_\ell(j)$. Note that for $d = 1$ this is equivalent to the containment of one permutation in another. We say a d-permutation M is *k-pattern-universal* if it contains all d-patterns of order k.

Linial and Simkin [80] considered "monotone subsequences of length k" in d-permutations, which expressed in our language correspond to d-patterns of order k of the form $(\sigma, \dots, \sigma)$, where σ is the identity function. They showed that the longest monotone subsequence in a random d-permutation of order n has length $\Theta(n^{d/(d+1)})$ with high probability. This implies that a random d-permutation needs to have order at least $\Omega(k^{(d+1)/d})$ to be k-pattern-universal with high probability. In analogy with the case of permutations, we believe this to be tight.

Bibliography

[1] E. Aigner-Horev, D. Conlon, H. Hàn, Y. Person, and M. Schacht. Quasirandomness in hypergraphs. *Electron. J. Combin.*, 25(3):3–34, 2018.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. On a conjecture of Loebl. In *Graph theory, combinatorics, and algorithms, Vol. 1, 2 (Kalamazoo, MI, 1992)*, Wiley-Intersci. Publ., pages 1135–1146. Wiley, New York, 1995.

[3] N. Alon. Asymptotically optimal induced universal graphs. *Geom. Funct. Anal.*, 27(1):1–32, 2017.

[4] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.*, 39(1):143–167, 2009.

[5] N. Alon, M. Krivelevich, and B. Sudakov. Embedding nearly-spanning bounded degree trees. *Combinatorica*, 27(6):629–644, 2007.

[6] N. Alon and A. Shapira. Every monotone graph property is testable. In *Proc. of the 37th Annual ACM Symposium on Theory of Computing*, STOC'05, pages 128–137. ACM, 2005.

[7] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.*, 37(6):1703–1727, 2008.

[8] N. Alon and N. Sherman. Induced universal hypergraphs. *SIAM J. Discrete Math.*, 33(2):629–642, 2019.

[9] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley Publishing, 4th edition, 2016.

[10] P. Araújo, L. Moreira, and M. Pavez-Signé. Ramsey goodness of trees in random graphs. *arXiv preprint arXiv:2001.03083*, 2020.

[11] R. Arratia. On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern. *Electron. J. Combin.*, 6(1):1, 1999.

[12] M. Axenovich, Y. Person, and S. Puzynina. A regularity lemma and twins in words. *J. Combin. Theory, Series A*, 120(4):733–743, 2013.

[13] J. Balogh, B. Csaba, and W. Samotij. Local resilience of almost spanning trees in random graphs. *Random Structures Algorithms*, 38(1-2):121–139, 2011.

[14] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, 28(3):669–709, 2015.

[15] J. Balogh, R. Morris, and W. Samotij. The method of hypergraph containers. *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 3059–3092, 2018.

[16] J. Beck. On size Ramsey number of paths, trees, and circuits. I. *J. Graph Theory*, 7(1):115–129, 1983.

[17] G. Besomi, M. Pavez-Signé, and M. Stein. Maximum and minimum degree conditions for embedding trees. *SIAM J. Discrete Math.*, 34(4):2108–2123, 2020.

[18] G. Besomi, M. Pavez-Signé, and M. Stein. On the Erdős–Sós conjecture for bounded degree trees. Accepted in Combin. Probab. Comput.

[19] G. Besomi, M. Pavez-Signé, and M. Stein. Degree conditions for embedding trees. *SIAM J. Discrete Math.*, 33(3):1521–1555, 2019.

[20] Y. Biers-Ariel, A. Godbole, and E. Kelley. Expected Number of Distinct Subsequences in Randomly Generated Binary Strings. *Discrete Math. Theor. Comput. Sci.*, Vol. 19 no. 2, Permutation Patterns 2016, June 2018.

[21] P. Billingsley. *Probability and Measure.* Wiley Series in Probability and Statistics. Wiley, 1995.

[22] M. Blum and S. Kannan. Designing programs that check their work. *J. ACM*, 42:269–291, 1995.

[23] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47:549–595, 1993.

[24] B. Bollobás. *Extremal Graph Theory.* L.M.S. Monographs. Academic Press, 1978.

[25] B. Bollobás and A. Thomason. Graphs which contain all small graphs. *European J. Combin.*, 2(1):13–15, 1981.

[26] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[27] J. Böttcher. *Large-scale structures in random graphs*, page 87–140. London Mathematical Society Lecture Note Series. Cambridge University Press, 2017.

[28] S. Brandt and E. Dobson. The Erdős–Sós conjecture for graphs of girth 5. *Discrete Math.*, 150:411–414, 1996.

[29] G. R. Brightwell and Y. Kohayakawa. Ramsey properties of orientations of graphs. *Random Structures Algorithms*, 4(4):413–428, 1993.

[30] B. Bukh and L. Zhou. Twins in words and long common subsequences in permutations. *Israel J. Math.*, 213(1):183–209, 2016.

[31] F. Chung and R. Graham. Quasi-random hypergraphs. *Proc. Natl. Acad. Sci.*, 86(21):8175–8177, 1989.

[32] F. Chung and R. Graham. Quasi-random subsets of $\mathbb{Z}_n$. *J. Combin. Theory, Series A*, 61(1):64–86, 1992.

[33] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[34] V. Chvátal and D. Sankoff. Longest common subsequences of two random sequences. *J. Appl. Probab.*, 12(2):306–315, 1975.

[35] D. Conlon. Combinatorial theorems relative to a random set. *Proceedings of the International Congress of Mathematicians (ICM 2014)*, 4:303–328, 2014.

[36] D. Conlon, J. Fox, and B. Sudakov. *Recent developments in graph Ramsey theory*, volume 424 of *Surveys in Combinatorics 2015 (London Mathematical Society Lecture Note Series)*, pages 49–118. Cambridge University Press, 2015.

[37] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Ann. of Math.*, pages 367–454, 2016.

[38] J. N. Cooper. Quasirandom permutations. *J. Combin. Theory, Series A*, 106(1):123–143, 2004.

[39] B. Csaba, I. Levitt, J. Nagy-György, and E. Szemerédi. Tight bounds for embedding bounded degree trees. In *Katona G.O.H., Schrijver A., Szenyi T., Sági G. (eds) Fête of Combinatorics and Computer Science*, volume 20, pages 95–137, 2010.

[40] N. G. De Bruijn. A combinatorial problem. In *Proc. Koninklijke Nederlandse Academie van Wetenschappen*, volume 49, pages 758–764, 1946.

[41] G. A. Dirac. Some theorems on abstract graphs. *Proc. Lond. Math. Soc.*, 2:69–81, 1952.

[42] P. Erdős. Extremal problems in graph theory. In *Theory of graphs and its applications, Proc. Sympos. Smolenice*, pages 29–36, 1964.

[43] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar.*, 10(3):337–356, 1959.

[44] P. Erdős and R. L. Graham. On partition theorems for finite graphs. In *Infinite and finite sets (Colloq. Keszthely, 1973; dedicated to P. Erdős on his 60th birthday)*, volume 10, pages 515–527. Colloq. Math. Soc. János Bolyai, 1975.

[45] P. Erdős, J. Pach, R. Pollack, and Z. Tuza. Radius, diameter, and minimum degree. *J. Combin. Theory Ser. B*, 47(1):73 – 79, 1989.

[46] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6(18):290–297, 1959.

[47] U. Feige, T. Koren, and M. Tennenholtz. Chasing ghosts: Competing with stateful policies. *SIAM J. Comput.*, 46(1):190–223, 2017.

[48] M. Fekete. Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Math. Z.*, 17:228–249, 1923.

[49] P. Frankl and V. Rödl. Large triangle-free subgraphs in graphs withoutk 4. *Graphs Combin.*, 2(1):135–144, 1986.

[50] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *arXiv preprint arXiv:1910.13433*, 2019.

[51] J. Friedman and N. Pippenger. Expanding graphs contain all small trees. *Combinatorica*, 7(1):71–76, 1987.

[52] S. Gerke and A. Steger. *The sparse regularity lemma and its applications*, page 227–258. Surveys in Combinatorics 2005 (London Mathematical Society Lecture Note Series). Cambridge University Press, 2005.

[53] R. Glebov, A. Grzesik, T. Klimošová, and D. Král'. Finitely forcible graphons and permutons. *J. Combin. Theory B*, 110:112 – 135, 2015.

[54] A. Goerlich and A. Zak. On Erdős–Sós Conjecture for Trees of Large Size. *Electron. J. Combin.*, 23(1):P1–52, 2016.

[55] O. Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

[56] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[57] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[58] W. T. Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combin. Probab. Comput.*, 15(1-2):143–184, 2006.

[59] W. T. Gowers. Quasirandom groups. *Combin. Probab. Comput.*, 17(3):363–387, 2008.

[60] H. Hàn, M. Kiwi, and M. Pavez-Signé. Quasi-random words and limits of word sequences. *arXiv preprint arXiv:2003.03664*, 2020.

[61] F. Havet, B. Reed, M. Stein, and D. R. Wood. A variant of the Erdős–Sós conjecture. *J. Graph Theory*, 94(1):131–158, 2020.

[62] P. E. Haxell. Tree embeddings. *J. Graph Theory*, 36(3):121–130, 2001.

[63] X. He and M. Kwan. Universality of random permutations. *Bull. Lond. Math. Soc.*, 52(3):515–529, 2020.

[64] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. M. Sampaio. Limits of permutation sequences. *J. Combin. Theory, Series B*, 103(1):93–113, 2013.

[65] C. Hoppen, Y. Kohayakawa, C. G. Moreira, and R. M. Sampaio. Testing permutation properties through subpermutations. *Theoret. Comput. Sci.*, 412(29):3555–3567, 2011.

[66] S. Janson, T. Łuczak, and A. Rucinski. *Random Graphs.* John Wiley & Sons, Ltd, 2011.

[67] J. Kahn and G. Kalai. Thresholds and expectation thresholds. *Combin. Probab. Comput.*, 16(3):495–502, 2007.

[68] M. Kiwi, M. Loebl, and J. Matoušek. Expected length of the longest common subsequence for large alphabets. *Adv. Math.*, 197(2):480–498, 2005.

[69] T. Klimošová and D. Král'. Hereditary properties of permutations are strongly testable. In *Proc. of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'14, pages 1164–1173. SIAM, 2014.

[70] Y. Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In *Found. Comput. Math.*, pages 216–230. Springer, 1997.

[71] Y. Kohayakawa and V. Rödl. Szemerédi's regularity lemma and quasi-randomness. In *Recent advances in algorithms and combinatorics*, pages 289–351. Springer, 2003.

[72] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of a Packing Conjecture of Bollobás. *Combin. Probab. Comput.*, 4(3):241–255, 1995.

[73] J. Komlós, G. N. Sárközy, and E. Szemerédi. Spanning trees in dense graphs. *Combin. Probab. Comput.*, 10(5):397–416, 2001.

[74] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. In *Theoretical Aspects of Computer Science, Advanced Lectures (First Summer School on Theoretical Aspects of Computer Science, Tehran, Iran, July 2000)*, pages 84–112, 2000.

[75] D. Král' and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geom. Funct. Anal.*, 23(2):570–579, 2013.

[76] M. Krivelevich. Embedding spanning trees in random graphs. *SIAM J. Discrete Math.*, 24(4):1495–1500, 2010.

[77] D. Kühn and D. Osthus. *Embedding large subgraphs into dense graphs*, pages 137–167. Surveys in Combinatorics 2009 (London Mathematical Society Lecture Note Series). Cambridge University Press, 2009.

[78] S. Letzter. Path ramsey number for random graphs. *Combin. Probab. Comput.*, 25(4):612–622, 2016.

[79] N. Linial and Z. Luria. An upper bound on the number of high-dimensional permutations. *Combinatorica*, 34(4):471–486, 2014.

[80] N. Linial and M. Simkin. Monotone subsequences in high-dimensional permutations. *Combin. Probab. Comput.*, 27(1):69–83, 2018.

[81] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, 2 edition, 1997.

[82] L. Lovász. *Large Networks and Graph Limits.*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.

[83] L. Lovász and V. T. Sós. Generalized quasirandom graphs. *J. Combin. Theory, Series B*, 98(1):146–163, 2008.

[84] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory, Series B*, 96(6):933–957, 2006.

[85] L. Lovász and B. Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178(1):113–156, 2010.

[86] L. Lovász and B. Szegedy. Finitely forcible graphons. *J. Combin. Theory, Series B*, 101(5):269–301, 2011.

[87] T. Łuczak, A. Ruciński, and B. Voigt. Ramsey properties of random graphs. *J. Combin. Theory Ser. B*, 56(1):55–68, 1992.

[88] R. Montgomery. Spanning trees in random graphs. *Adv. Math.*, 356:106793, 2019.

[89] J. W. Moon. On minimal n-universal graphs. *Proceedings of the Glasgow Mathematical Association*, 7(1):32–33, 1965.

[90] R. Morris. Personal communication.

[91] M. Pavez-Signé, D. A. Quiroz, and N. Sanhueza-Matamala. Universal arrays. *arXiv preprint arXiv:2001.05767*, 2020.

[92] M. Pinsky. *Introduction to Fourier Analysis and Wavelets*. Graduate studies in mathematics. American Mathematical Society, 2008.

[93] R. Rado. Universal graphs and universal functions. *Acta Arithmetica*, 9(4):331–340, 1964.

[94] B. Reed and M. Stein. Spanning trees in graphs of high minimum degree with a universal vertex I: An approximate asymptotic result. Preprint 2019, arXiv 1905.09801.

[95] B. Reed and M. Stein. Spanning trees in graphs of high minimum degree with a universal vertex II: A tight result. Preprint 2019, arXiv 1905.09806.

[96] V. Rödl. On universality of graphs with uniformly distributed edges. *Discrete Math.*, 59(1-2):125–134, 1986.

[97] V. Rödl and A. Rucinski. Lower bounds on probability thresholds for Ramsey properties. *Combinatorics, Paul Erdős is eighty*, 1:317–346, 1993.

[98] V. Rödl and A. Ruciński. Threshold functions for Ramsey properties. *J. Amer. Math. Soc.*, 8(4):917–942, 1995.

[99] V. Rozhoň. A local approach to the Erdős–Sós conjecture. *SIAM J. Discrete Math.*, 33(2):643–664, 2019.

[100] J.-F. Saclé and M. Woźniak. A note on the Erdős–Sós conjecture for graphs without C_4. *J. Combin. Theory Ser. B*, 70(2):229–234, 1997.

[101] D. Saxton and A. Thomason. Hypergraph containers. *Invent. Math.*, 201(3):925–992, 2015.

[102] M. Schacht. Extremal results for random discrete structures. *Ann. of Math.*, pages 333–365, 2016.

[103] M. Simonovits and E. Szemerédi. Embedding graphs into larger graphs: results, methods, and problems. In *Building Bridges II*, pages 445–592. Springer, 2019.

[104] B. Szegedy. From graph limits to higher order Fourier analysis. In *Proc. of the International Congress of Mathematicians*, volume 3, pages 3197–3218. World Scientific, 2018.

[105] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[106] A. Thomason. Pseudo-random graphs. In A. Barlotti, M. Biliotti, A. Cossu, G. Korchmaros, and G. Tallini, editors, *Annals of Discrete Mathematics (33)*, volume 144 of *North-Holland Mathematics Studies*, pages 307–331. North-Holland, 1987.

[107] A. Thue. Über unendliche zeichenreihen. *Norske Vid. Selsk. Skr. I Mat.-Nat. Kl*, 7:1–22, 1906.

[108] H. Towsner. σ-algebras for quasirandom hypergraphs. *Random Structures Algorithms*, 50(1):114–139, 2017.