Trees and graph packing

Ph.D. Thesis

Bálint Márk Vásárhelyi

Supervisor:

Dr. Béla Csaba

Docent

Doctoral School of Mathematics and Computer Science

University of Szeged

Faculty of Science and Informatics

Bolyai Institute

Szeged, 2018

Contents

Chapter 1

Introduction

This thesis is about trees and graph packing. In the first part, which is mainly based on [61], we deal with suffix trees. At this point, we present an overview of suffix trees, while at the beginning of Chapter 2, we will give the detailed definitions. A suffix tree is a powerful data structure which is used for a large number of combinatorial problems involving strings. Suffix tree is a structure for compact storage of the suffixes of a given string. The compact suffix tree is a modified version of the suffix tree, and it can be stored in linear space of the length of the string, while the non-compact suffix tree is quadratic (see [32, 51, 60, 64]).

The notion of suffix trees was first introduced by Weiner [64], though he used the name compacted bi-tree. Grossi and Italiano mention that in the scientific literature, suffix trees have been rediscovered many times, sometimes under different names, like compacted bi-tree, prefix tree, PAT tree, position tree, repetition finder, subword tree etc. [31]

Linear time and space algorithms for creating the compact suffix tree were given soon by Weiner [64], McCreight [51], Ukkonen [60], Chen and Sciferas [13] and others.

The statistical behavior of suffix trees has been also studied. Most of the studies consider improved versions of suffix trees. The average size of compact suffix trees was examined by Blumer, Ehrenfeucht and Haussler [6]. They proved that the average number of nodes in the compact suffix tree is asymptotically the sum of an oscillating function and a small linear function.

An important question is the height of suffix trees, which was answered by Devroye, Szpankowski and Rais [23], who proved that the expected height is logarithmic in the length of the string.

The application of suffix trees is very wide. We mention but only a few examples. Apostolico et al. [4] mention that these structures are used in text searching, indexing, statistics, compression. In computational biology, several algorithms are based on suffix trees. Just to refer a few of them, we mention the works of Höhl et al. [36], Adebiyi et al. [1] and Kaderali et al. [37]

Suffix trees are also used for detecting plagiarism [4], in cryptography [52, 54], in data compression [26, 28, 54] or in pattern recognition [59].

For the interested readers further details on suffix trees, their history and their applications can be found in [4], in [31] and in [32], which sources we also used for the overview of the history of suffix trees.

It is well-known that the non-compact suffix tree can be quadratic in space as we referred before. In Chapter 2 we are setting a lower bound on the average size, which is also quadratic.

Now, we turn to the other main field of the thesis, which is graph packing.

All graphs considered in this thesis are simple. We use standard graph theory notations (see for example [65]): $deg_G(v)$ (or briefly, if G is understood from the context, $deg(v)$) is the degree of v in G. The number of edges between X and Y for $X \cap Y = \emptyset$ is denoted by $e(X, Y)$. The number of neighbors of x in a subset $S \subseteq V(G)$ is denoted by $deg_G(x, S)$, and $\delta(G)$ and $\Delta(G)$ denote the minimum and maximum degree of G, respectively.

For any function f on V let $f(X) = \sum_{v \in X} f(v)$ for every $X \subseteq V$. $\pi(G)$ is the degree sequence of G. The number of vertices in G is denoted by $v(G)$, while the number of its edges is denoted by $e(G)$.

Given a bipartite graph $G(A, B)$ we call it *balanced* if $|A| = |B|$. This notion naturally generalizes for r-partite graphs with $r \in \mathbb{N}$, $r \geq 2$.

The complete graph on n vertices is denoted by K_n, the complete bipartite graph with vertex class sizes n and m is denoted by $K_{n,m}$.

A finite sequence of natural numbers $\pi = (d_1, \ldots, d_n)$ is a *graphic sequence* or *degree sequence* if there exists a graph G such that π is the (not necessarily) monotone degree sequence of G. Such a graph G *realizes* π. The largest value of π is denoted by $\Delta(\pi)$. We sometimes refer to the value of π at vertex v as $\pi(v)$. The degree sequence $\pi = (a_1, \ldots, a_k, b_1, \ldots, b_l)$ is a *bigraphic* sequence if there exists a simple bipartite graph $G = G(A, B)$ with $|A| = k$, $|B| = l$ realizing π such that the degrees of vertices in A are $a_1, \ldots, a_k$, and the degrees of the vertices of B are $b_1, \ldots, b_l$.

Let G and H be two graphs on n vertices. We say that H is a subgraph of G, if we can delete edges from G so that we obtain an isomorphic copy of H. We denote this relation by $H \subseteq G$. In the literature the equivalent complementary formulation can be found as well: we say that H and $\overline{G}$ *pack* if there exist edge-disjoint copies of H and $\overline{G}$ in K_n. Here $\overline{G}$ denotes the *complement* of G.

If $S \subseteq V$ for some graph $G = (V, E)$ then the subgraph spanned by S is denoted by $G[S]$. Moreover, let $Q \subseteq V$ so that $S \cap Q = \emptyset$, then $G[S, Q]$ denotes the bipartite subgraph of G on vertex classes S and Q, having every edge of G that connects a vertex of S with a vertex of Q.

It is an old an well-understood problem in graph theory to tell whether a given sequence of natural numbers is a degree sequence or not. We consider a generalization of it, which is remotely related to the so-called discrete tomography (or degree sequence packing) problem (see e.g. [24]) as well. In the discrete tomography problem we are given two degree sequences of length n, π_1 and π_2, and the question is whether there exists a graph G on n vertices with a red-blue edge coloring so that the following holds: for every vertex v the red degree of v is $\pi_1(v)$ and the blue degree of v is $\pi_2(v)$.

The question whether a sequence of n numbers π is a degree sequence can also be formulated as follows: Does K_n have a subgraph H such that the degree sequence of H is π? The question becomes more general if K_n is replaced by some (simple) graph G on n vertices. If the answer is yes, we say that π *can be embedded into* G, or equivalently, π packs with $\overline{G}$.

The graph packing problem is the following. Let G and H be two graphs on n vertices. We say that G and H *pack* if and only if K_n contains edge-disjoint copies of G and H as subgraphs.

The graph packing problem can be formulated as an embedding problem, too. G and H pack if and only if H is isomorphic to a subgraph of $\overline{G}$ ($H \subseteq \overline{G}$).

A classical result in this field is the following theorem of Sauer and Spencer.

Theorem 1 (Sauer, Spencer [57]). *Let G_1 and G_2 be graphs on n vertices with maximum degrees Δ_1 and Δ_2, respectively. If $\Delta_1\Delta_2 < \frac{n}{2}$, then G_1 and G_2 pack.*

Many seemingly unrelated problems can be translated to the language of embedding/packing, for a (non-complete) list see for example [38]. Therefore, it is not surprising that in general many embedding/packing problems are open. In order to prove meaningful results one usually imposes condition on the graphs in question.

In Chapter 3 we study the **bipartite packing problem** as it is formulated by Catlin [11], Hajnal and Szegedy [34] and was used by Hajnal for proving deep results in complexity theory of decision trees [33].

Let $G_1 = (A, B; E_1)$ and $G_2 = (S, T; E_2)$ be bipartite graphs with $|A| = |S| = m$ and $|B| = |T| = n$. Sometimes, we use only $G(A, B)$ if we want to say that G is a bipartite graph with classes A and B. Let $\Delta_A(G_1)$ be the maximal degree of G_1 in A. We use $\Delta_B(G_1)$ similarly.

The bipartite graphs G_1 and G_2 pack in the bipartite sense (i.e. they have a *bipartite packing*) if there are edge-disjoint copies of G_1 and G_2 in $K_{m,n}$.

The bipartite packing problem can be also formulated as a question of embedding. The bipartite graphs $G_1 = (A, B; E)$ and G_2 pack if and only if G_2 is isomorphic to a subgraph of $\widetilde{G_1}$, which is the bipartite complement of G_1, i.e. $\widetilde{G_1} = (A, B; (A \times B) - E)$.

Let us mention two classical results in extremal graph theory.

Theorem 2 (Dirac, [25]). *Every graph G with $n \geq 3$ vertices and minimum degree $\delta(G) \geq \frac{n}{2}$ has a Hamilton cycle.*

Theorem 3 (Corrádi-Hajnal, [16]). *Let $k \geq 1$, $n \geq 3k$, and let H be an n-vertex graph with $\delta(H) \geq 2k$. Then H contains k vertex-disjoint cycles.*

Observe, that Dirac's theorem implies that given a constant 2 degree sequence π of length n and any graph G on n vertices having minimum degree $\delta(G) \geq n/2$, π can be embedded into G. One can interpret the Corrádi-Hajnal theorem similarly, but here one may require more on the structure of the graph that realizes π and in exchange a larger minimum degree of G is needed.

In Chapter 5 we extend the results of [18]. We consider bounded degree bipartite graphs that have a small separator and large bandwidth, and prove that under reasonable conditions these are spanning subgraphs of n-vertex graphs that have minimum degree just slightly larger than $n/2$. We also show that using earlier methods such graphs cannot be embedded in general into host graphs with such small minimum degree.

Regularity Lemma

An important tool for our results is the Regularity Lemma, for which Endre Szemerédi received Abel Prize in 2012. At this point we give a short overview of it. For a more detailed discussion we refer to [44] and [45].

The *density* between disjoint sets X and Y is defined as:

$$d(X,Y) = \frac{e(X,Y)}{|X||Y|} \tag{1.1}$$

We will need the following definition to state the Regularity Lemma.

Definition 4 (Regularity condition)**.** Let $\varepsilon > 0$. A pair (A, B) of disjoint vertex-sets in G is ε-regular if for every $X \subseteq A$ and $Y \subseteq B$, satisfying

$$|X| > \varepsilon|A|, \ |Y| > \varepsilon|B| \tag{1.2}$$

we have

$$|d(X,Y) - d(A,B)| < \varepsilon. \tag{1.3}$$

This definition implies that regular pairs are highly uniform bipartite graphs; namely, the density of any reasonably large subgraph is almost the same as the density of the regular pair.

Definition 5. We say that a partition $\{W_0; W_1, \dots, W_k\}$ is ε-regular if there is an m such that for all $i > 0$ $|W_i| = m$; for all but at most εk^2 pairs (i,j) the pair (W_i, W_j) is ε -regular $(i,j > 0)$; and $|W_0| \leq \varepsilon m^2$.

The original form of Szemerédi's Regularity Lemma is the following:

Lemma 6. *[58] For every ε and t, there exist N and T such that for each $n \geq N$ every n-vertex graph G admits an ε-regular partition $W_0 \cup W_1 \cup \cdots \cup W_k$ satisfying $t \leq k \leq T$.*

We will also use the following form of the Regularity Lemma:

Lemma 7 (Degree Form, [18])**.** *For every $\varepsilon > 0$ there is an $M = M(\varepsilon)$ such that if $G = (W, E)$ is any graph and $d \in [0, 1]$ is any real number, then there is a partition of the vertex set V into $\ell + 1$ clusters $W_0, W_1, \dots, W_\ell$, and there is a subgraph G' of G with the following properties:*

- $\ell \leq M$,
- $|W_0| \leq \varepsilon|W|$,

- *all clusters* W_i*,* $i \geq 1$*, are of the same size* $m \leq \left\lfloor \frac{|W|}{\ell} \right\rfloor < \varepsilon|W|$*,*
- $deg_{G'}(v) > deg_G(v) - (d+\varepsilon)|W|$ *for all* $v \in W$*,*
- $G'|_{W_i} = \emptyset$ *(*W_i *is an independent set in* G'*) for all* $i \geq 1$*,*
- *all pairs* (W_i, W_j)*,* $1 \leq i < j \leq \ell$*, are* ε*-regular, each with density either 0 or greater than* d *in* G'*.*

We call W_0 the *exceptional cluster*, $W_1, \ldots, W_\ell$ are the *non-exceptional clusters.*

Definition 8 (Reduced graph, [18])**.** Apply Lemma 7 to the graph $G = (W, E)$ with parameters ε and d, and denote the clusters of the resulting partition by $W_0, W_1, \ldots, W_\ell$, W_0 being the exceptional cluster. We construct a new graph G_r, the reduced graph of G' in the following way: The non-exceptional clusters of G' are the vertices of the reduced graph G_r (hence $v(G_r) = \ell$). We connect two vertices of G_r by an edge if the corresponding two clusters form an ε-regular pair with density at least d.

The following corollary is immediate:

Corollary 9. *[18] Apply Lemma 7 with parameters* ε *and* d *to the graph* $G = (W, E)$ *satisfying* $\delta(G) \geq \gamma n \ \ (v(G) = n)$ *for some* $\gamma > 0$*. Denote* G_r *the reduced graph of* G'*. Then* $\delta(G_r) \geq (\gamma - \theta)\ell$*, where* $\theta = 2\varepsilon + d$*.*

The (fairly easy) proof of the lemma below can be found in [45].

Lemma 10. *Let* (A, B) *be an* ε*-regular–pair with density* d *for some* $\varepsilon > 0$*. Let* $c > 0$ *be a constant such that* $\varepsilon \ll c$*. We arbitrarily divide* A *and* B *into two parts, obtaining the non-empty subsets* A', A'' *and* B', B''*, respectively. Assume that* $|A'|, |A''| \geq c|A|$ *and* $|B'|, |B''| \geq c|B|$*. Then the pairs* (A', B')*,* (A', B'')*,* (A'', B') *and* (A'', B'') *are all* ε/c*–regular pairs with density at least* $d - \varepsilon/c$*.*

Blow-up Lemma

Let H and G be two graphs on n vertices. Assume that we want to find an isomorphic copy of H in G. In order to achieve this one can apply a very powerful tool, the Blow-up Lemma of Komlós, Sárközy and Szemerédi [42, 40]. For stating it we need a new notion, a stronger one-sided property of regular pairs.

Definition 11 (Super-Regularity condition)**.** Given a graph G and two disjoint subsets of its vertices A and B, the pair (A, B) is (ε, δ)-super-regular, if it is ε-regular and furthermore,

$$deg(a) > \delta|B|, \text{ for all } a \in A, \tag{1.4}$$

and

$$deg(b) > \delta|A|, \text{ for all } b \in B. \tag{1.5}$$

Theorem 12 (Blow-up Lemma [42, 40])**.** *Given a graph R of order r and positive integers δ, Δ, there exists a positive $\varepsilon = \varepsilon(\delta, \Delta, r)$ such that the following holds: Let $n_1, n_2, \ldots, n_r$ be arbitrary positive parameters and let us replace the vertices $v_1, v_2, \ldots, v_r$ of R with pairwise disjoint sets $W_1, W_2, \ldots, W_r$ of sizes $n_1, n_2, \ldots, n_r$ (blowing up R). We construct two graphs on the same vertex set $V = \cup_i W_i$. The first graph F is obtained by replacing each edge $v_iv_j \in E(R)$ with the complete bipartite graph between W_i and W_j. A sparser graph G is constructed by replacing each edge v_iv_j arbitrarily with an (ε, δ)-super-regular pair between W_i and W_j. If a graph H with $\Delta(H) \leq \Delta$ is embeddable into F then it is already embeddable into G.*

Chapter 2

Suffix trees

In this chapter, we will set up a lower bound on the average size of a suffix tree. Here, the results of [61] are presented. Before we give the exact definition of a suffix tree, we will need a few definitions.

Definition 13. An *alphabet* Σ is a set of different characters. The *size* of an alphabet is the size of this set, which we denote by $\sigma(\Sigma)$, or more simply σ. A string S is over the alphabet Σ if each character of S is in Σ. We will use \$ as a character not in Σ.

Definition 14. Let S be a string. $S[i]$ is its ith character, while $S[i,j]$ is a *substring* of S, from $S[i]$ to $S[j]$, if $j \geq i$, else $S[i,j]$ is the empty string. Usually $n(S)$ (or n if there is no danger of confusion) denotes the *length* of the string.

Now, we are at defining the suffix trees.

Definition 15. The *suffix tree* of the string S is a rooted directed tree with n leaves, where n is the length of S.

Its structure is the following:

> Each edge e has a *label* $\ell(e)$, and the edges from a node v have different labels (thus, the suffix tree of a string is unique). If we concatenate the edge labels along a path $\mathcal{P}$, we get the *path label* $\mathcal{L}(\mathcal{P})$.

We denote the path from the root to the leaf j by $\mathcal{P}(j)$. The edge labels are such that $\mathcal{L}(j) = \mathcal{L}(\mathcal{P}(j))$ is $S[j, n]$ and a \$ character (which is not in Σ) at the end. The definition becomes more clear if we check the example on Figure 2.1 and Algorithm 16.

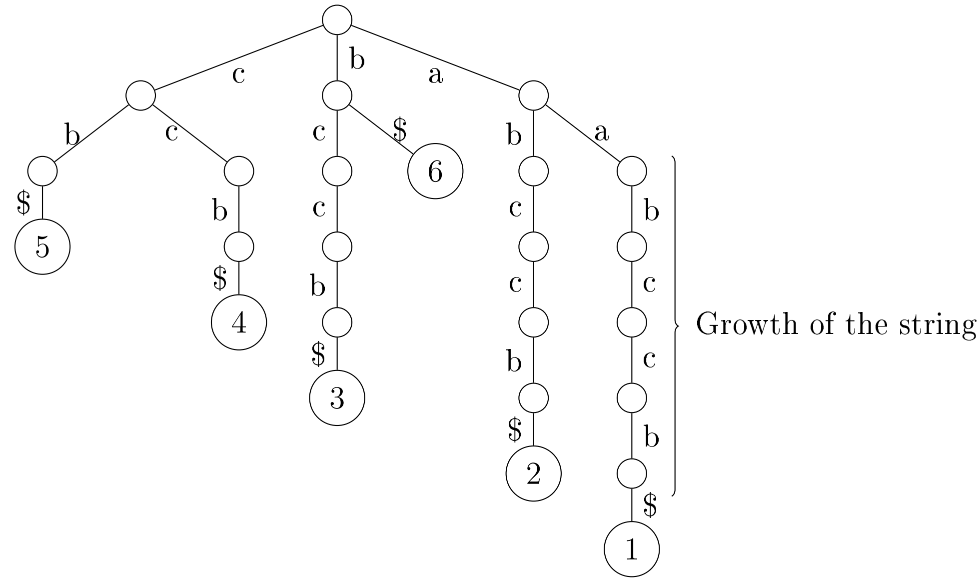

Figure 2.1: Suffix tree of string *aabccb*

A naive algorithm for constructing the suffix tree is the following:

Algorithm 16. Let S be a string of length n. Let $j = 1$ and T be a tree of one vertex r (the root of the suffix tree).

Step 1: Consider $X = S[j, n] + \$$. Set $i = 0$, and $v = r$.

Step 2: If there is an edge vu labeled $X[i + 1]$, then set $v = u$ and $i = i + 1$.

Step 3: Repeat Step 2 while it is possible.

Step 4: If there is no such an edge, add a path of $n - j - i + 2$ edges from v, with labels corresponding to $S[j + i, n] + \$$, consecutively on the edges. At the end of the path, number the leaf with j.

Step 5: Set $j = j + 1$, and if $j \leq n$, go to Step 1. $\diamond$

Note that in Algorithm 16 a leaf always remain a leaf, as \$ (which is the last edge label before a leaf) is not a character in S.

Definition 17. The *compact suffix tree* is a modified version of the suffix tree. We get it from the suffix tree by compressing its long branches.

The structure of the compact suffix tree is basically similar to that of the suffix tree, but an edge label can be longer than one character, and each internal node (i.e. not leaf) must have at least two children. For an example see Figure 2.2.

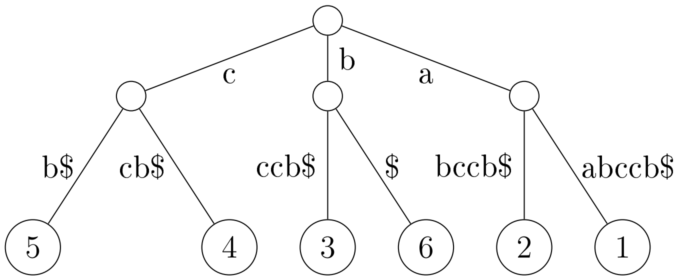

Figure 2.2: Compact tree of string *aabccb*

With a regard to suffix trees, we can define further notions for strings.

Definition 18. Let S be a string, and $\mathcal{T}$ be its (non-compact) suffix tree.

A *natural direction* of $\mathcal{T}$ is that all edges are directed from the root towards the leaves. If there is a directed path from u to v, then v is a *descendant* of u and u is an *ancestor* of v.

We say that *the growth of* S (denoted by $\gamma(S)$) is one less than the shortest distance of leaf 1 from an internal node v which has at least two children (including leaf 1), that is, we count the internal nodes on the path different from v. If leaf j is a descendant of v, then the common prefix of $S[j,n]$ and $S[1,n]$ is the longest among all j's.

If we consider the string $S = aabccb$, the *growth* of S is 5, as it can be seen on Figure 2.1.

An important notion is the following one.

Definition 19. Let $\Omega(n, k, \sigma)$ be the number of strings of length n with growth k over an alphabet of size σ.

Observe that the connection between the growth and the number of nodes in a suffix tree is the following:

Observation 20. *If we construct the suffix tree of S by using Algorithm 16, we get that the sum of the growths of $S[n-1, n], S[n-2, n], \ldots, S[1, n]$ is a lower bound to the number of nodes in the final suffix tree. In fact, there are only two more internal nodes, the root vertex, the only node on the path to leaf n, and we have the leaves.*

In the proofs we will need the notion of period and of aperiodic strings.

Definition 21. Let S be a string of length n. We say that S is *periodic* with period d, if there is a $d|n$ for which $S[i] = S[i+d]$ for all $i \leq n-d$. Otherwise, S is *aperiodic*.

The *minimal period* of S is the smallest d with the property above.

Definition 22. $\mu(j, \sigma)$ is the number of j-length aperiodic strings over an alphabet of size σ.

A few examples for the number of aperiodic strings are given in Table 2.1.

σ	$\mu(1,\sigma)$	$\mu(2,\sigma)$	$\mu(3,\sigma)$	$\mu(4,\sigma)$	$\mu(5,\sigma)$	$\mu(6,\sigma)$	$\mu(7,\sigma)$	$\mu(8,\sigma)$
2	2	6	12	30	54	126	240	504
3	3	6	24	72	240	696	2184	648
4	4	12	60	240	1020	4020	16380	65280
5	5	20	120	600	3120	15480	78120	390000

Table 2.1: Number of aperiodic strings for small alphabets. σ is the size of the alphabet, and $\mu(j, \sigma)$ is the number of aperiodic strings of length j

The following three theorems are the main results of this chapter.

Theorem 23. *For any $k \in \mathbb{N}$, on any alphabet of size σ for all $n \geq 2k$,*

$$\Omega(n, k, \sigma) \leq \varphi(k, \sigma) \tag{2.1}$$

for some function φ.

Theorem 24. *There is a $c > 0$ and an n_0 such that for any $n > n_0$ the following is true. Let S' be a string of length $n-1$, and S be a string obtained from S' by adding a character to its beginning chosen uniformly random from the alphabet. Then the expected growth of S is at least cn.*

Theorem 25. *There is a $d > 0$ that for any $n > n_0$ (where n_0 is the same as in Theorem 24) the following holds. On an alphabet of size σ the simple suffix tree of a random string S of length n has at least dn^2 nodes in expectation.*

The main goal is to prove Theorem 25. First, we show that Theorem 24 implies Theorem 25, then we show that Theorem 23 implies Theorem 24. Finally, we prove Theorem 23.

Proof. (Theorem 25)

Considering Observation 20 we have that the expected size of the simple suffix tree of a random string S is at least

$$\mathbb{E}\sum_{m=1}^{n}\gamma(S[n-m,n]) \geq \sum_{m=1}^{n}\mathbb{E}\big(\gamma(S[n-m,n])\big). \tag{2.2}$$

We can divide the sum into two parts:

$$\sum_{m=1}^{n}\mathbb{E}\big(\gamma(S[n-m,n])\big) = \sum_{m=1}^{n_0}\mathbb{E}\big(\gamma(S[n-m,n])\big) + \sum_{m=n_0+1}^{n}\mathbb{E}\big(\gamma(S[n-m,n])\big). \tag{2.3}$$

The first part of the sum is non-negative, while the second part can be estimated with Theorem 24:

$$\sum_{m=n_0+1}^{n}\mathbb{E}\big(\gamma(S[n-m,n])\big) \geq \sum_{m=n_0+1}^{n} cn = dn^2. \tag{2.4}$$

This proves Theorem 25.

□

Before turning to the proof of Theorem 23, we show a few lemmas about the number of aperiodic strings. Lemma 26 can be found in [30] or in [15], but we give a short proof also here.

Lemma 26. *For all $j > 0$ integer and for all alphabets of size σ the number of aperiodic strings is*

$$\mu(j,\sigma) = \sigma^j - \sum_{\substack{d|j \\ d\neq j}} \mu(d,\sigma). \tag{2.5}$$

Proof. $\mu(1,\sigma) = \sigma$ is trivial.

There are σ^j strings of length j. Suppose that a string is periodic with minimal period d. This implies that its first d characters form an aperiodic string of length d, and there are $\mu(d,\sigma)$ such strings. This finishes the proof. □

Specially, if p is prime, then $\mu(p,\sigma) = \sigma^p - \sigma$.

Corollary 27. *If p is prime and $t \in \mathbb{N}$, then*

$$\mu\left(p^t,\sigma\right) = \sigma^{p^t} - \sigma^{p^{t-1}} \tag{2.6}$$

for all alphabets of size σ.

Proof. We count the aperiodic strings of length p^t. There are σ^{p^t} strings. Consider the minimal period of the string, i.e. the period which is aperiodic. If we exclude all minimal periods of length k, we exclude $\mu(k,\sigma)$ strings. This yields the following equality:

$$\mu\left(p^t,\sigma\right) = \sigma^{p^t} - \sum_{1\leq s<t} \mu\left(p^s,\sigma\right). \tag{2.7}$$

With a few transformations and using Lemma 26, we have that (2.7) is equal to

$$\sigma^{p^t} - \mu\left(p^{t-1},\sigma\right) - \sum_{1\leq s<t-1} \mu\left(p^s,\sigma\right) = \sigma^{p^t} - \sigma^{p^{t-1}} + \\ \sum_{1\leq s<t-1} \mu\left(p^s,\sigma\right) - \sum_{1\leq s<t-1} \mu\left(p^s,\sigma\right), \tag{2.8}$$

which is

$$\sigma^{p^t} - \sigma^{p^{t-1}}. \tag{2.9}$$

□

Lemma 28. *For all $j > 1$ and for all alphabets of size σ we have*

$$\mu(j,\sigma) \leq \sigma^j - \sigma. \tag{2.10}$$

Proof. From Lemma 26 we have $\mu(j,\sigma) = \sigma^j - \sum_{\substack{d|j \\ d\neq j}} \mu(d,\sigma)$. Considering $\mu(d,\sigma) \geq 0$ and $\mu(1,\sigma) = \sigma$, we get the claim of the lemma. □

Lemma 29. *For all $j \geq 1$ and for all alphabets of size σ we have*

$$\mu(j,\sigma) \geq \sigma(\sigma-1)^{j-1}. \tag{2.11}$$

Proof. We prove by induction. For $j = 1$ the claim is obvious, as $\mu(1,\sigma) = \sigma$.

Suppose we know the claim for $j - 1$. Consider $\sigma(\sigma-1)^{j-2}$ aperiodic strings of length $j - 1$. Now, for any of these strings there is at most one character by appending that to the end of the string we receive a periodic string of length j. Therefore we can append at least $\sigma - 1$ characters to get an aperiodic string, which gives the desired result.

□

Observation 30. *Observe that if the growth of S is k, then there is a j such that $S[1, n-k] = S[j+1, j+n-k]$. For example, if the string is* abcdefabcdab *($n = 12$), one can check that the growth is 8 (the new branch in the suffix tree which ends in leaf* 1 *starts after* abcd*), and with $j = 6$ we have $S[1,4] = S[7,10] =$* abcd.

The reverse of this observation is that if there is a $j < n$ such that $S[1, n-k] = S[j+1, j+n-k]$, then the growth is at most k, as $S[j+1, n]$ and $S[1, n]$ shares a common prefix of length $n - k$, thus, the paths to the leaves $j + 1$ and n share $n - k$ internal nodes, and at most k new internal nodes are created.

Now, we turn to the proof of Theorem 23.

Proof. (Theorem 23) We count the number of strings with growth k for $n \geq 2k$.

First, we fix j, and then count the number of possible strings where the growth occurs such that $S[1, n-k] = S[j+1, j+n-k]$ for that fixed j. Note that by this way, we only have an upper bound for this number, as we might found an ℓ such that $S[1, n-k+1] = S[\ell+1, \ell+n-k+1]$.

We know that $j \leq k$, otherwise $S[j+1, j+n-k]$ does not exist.

If $j = k$, then we know $S[1, n-k] = S[k+1, n]$.

$S[1, k]$ must be aperiodic. Suppose the opposite and let $S[1, k] = p \dots p$, where p is the minimal period with length d. Then $S[k+1, n] = p \dots p$. Obviously, in this case $S[1, n-d] = S[d+1, n]$, which by Observation 30 means that the growth would be at most d. See also Figure 2.3.

Therefore this case gives us at most $\mu(k)$ strings of growth k.

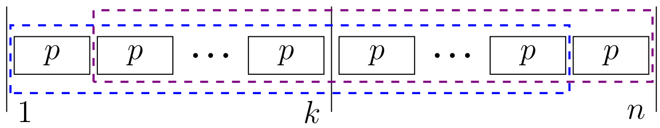

Figure 2.3: Proof of Theorem 23, case $j = k$

If $j < k$, then we have $S[1, n-k] = S[j+1, j+n-k]$.

First, we note that $S[1, j]$ must be aperiodic. Suppose the opposite and let $S[1, j] = p \dots p$, where p is the minimal period, and its length is d. Then

$$S[j+1, 2j] = S[2j+1, 3j] = \dots = p \dots p, \tag{2.12}$$

which means that

$$S\left[1, \left\lfloor \frac{k}{j} \right\rfloor \cdot j\right] = S\left[j+1, j + \left\lfloor \frac{k}{j} \right\rfloor \cdot j\right] = p \dots p. \tag{2.13}$$

This implies that $S[1, j+n-k] = p \dots pp'$, where p' is a prefix of p. However, $S[1, j+n-k-d] = S[d, j+n-k]$ is true, and using Observation 30, we have that $\gamma(S) \leq n - (j + n - k) + d = k - j + d < k$, which is a contradiction. See also Figure 2.4

Further, $S[j + n - k + 1]$ must not be the same as $S[k + 1]$, which means that this character can be chosen $\sigma - 1$ ways.

Therefore this case gives us at most $\mu(j)(\sigma - 1)\sigma^{k-j-1}$ strings of growth k for each j.

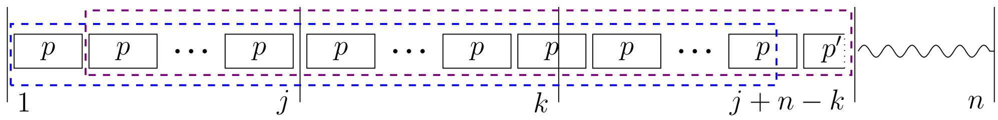

Figure 2.4: Proof of Theorem 23, case $j < k$

By summing up for each j, we have

$$\varphi(k, \sigma) = \sum_{j=1}^{k-1} \mu(j, \sigma)(\sigma - 1)\sigma^{k-j-1} + \mu(k, \sigma) \tag{2.14}$$

This completes the proof. □

Finally, we prove Theorem 24.

Proof. (Theorem 24)

According to Lemma 28, $\mu(j, \sigma) \leq \sigma^j - \sigma$ (if $j > 1$).

In the proof of Theorem 23 at (2.14) we saw for $k \geq 1$ and $n \geq 2k - 1$ that

$$\varphi(k, \sigma) = \mu(k, \sigma) + \sum_{j=1}^{k-1} \mu(j, \sigma)(\sigma - 1)\sigma^{k-j-1}. \tag{2.15}$$

We can bound the right hand side of (2.15) from above as it follows:

$$\mu(k, \sigma) + \sum_{j=1}^{k-1} \mu(j, \sigma)(\sigma - 1)\sigma^{k-j-1} = \mu(k, \sigma) + \mu(1, \sigma)(\sigma - 1)\sigma^{k-2} + \sum_{j=2}^{k-1} \mu(j, \sigma)(\sigma - 1)\sigma^{k-j-1}, \tag{2.16}$$

which is by Lemma 28 at most

$$\sigma^k - \sigma + \sigma(\sigma-1)\sigma^{k-2} + \sum_{j=2}^{k-1}(\sigma^j - \sigma)(\sigma-1)\sigma^{k-j-1} \leq$$
$$\sigma^k + \sigma^k + \sum_{j=2}^{k-1}\sigma^j\sigma\sigma^{k-j-1} \leq k\sigma^k. \tag{2.17}$$

Thus, $\varphi(k,\sigma) \leq k\sigma^k$, which means

$$\sum_{k=1}^{m}\varphi(k,\sigma) \leq \sum_{k=1}^{m}k\sigma^k \leq (m+1)\sigma^{m+1}. \tag{2.18}$$

The left hand side of (2.18) is an upper bound for the strings of growth at most m.

Let $m = \left\lfloor \frac{n}{2} \right\rfloor$.

As $\sigma^n \gg \frac{n}{2}\sigma^{\frac{n}{2}}$, this implies that in most cases the suffix tree of S has at least $\frac{n}{2}$ more nodes than the suffix tree of $S[1, n-1]$.

Thus, a lower bound on the expectation of the growth of S is

$$\mathbb{E}\big(\gamma(S)\big) \geq \frac{1}{\sigma^n}\left(\frac{n}{2}\sigma^{\frac{n}{2}} + \left(\sigma^n - \frac{n}{2}\sigma^{\frac{n}{2}}\right)\left(\frac{n}{2}+1\right)\right), \tag{2.19}$$

which is

$$\frac{1}{\sigma^n}\left(\frac{n+2}{2}\sigma^n + \left(\frac{n}{2} - \frac{n(n+2)}{4}\right)\sigma^{\frac{n}{2}}\right) = cn, \tag{2.20}$$

with some c, if n is large enough.

□

With this, we have finished the proof and gave a quadratic lower bound on the average size of suffix trees.

Chapter 3

Bipartite packing problem

In this and the following chapters of the thesis we will deal with graph packing problems. This chapter presents the results of [62]. First, we present a related result of Wojda and Vaderlind. For this, we need to introduce three families of graph pairs which they use in [66].

Let Γ_1 be the family of pairs $\{G(L,R), G'(L',R')\}$ of bipartite graphs such that G contains a star (i.e. one vertex in L is connected to all vertices of R), and in $\delta_{L'}(G') \geq 1$.

Let Γ_2 be the family of pairs $\{G(L,R), G'(L',R')\}$ of bipartite graphs such that $L = \{a_1, a_2\}$, and $deg_G(a_1) = deg_G(a_2) = 2$; and $L' = \{a'_1, a'_2\}$, $deg_{G'}(a'_1) = |R| - 1$, $deg_{G'}(a'_2) = 0$, finally, $\Delta_R(G) = \Delta_R(G') = 1$.

The family Γ_3 is the pair $\{G, G'\}$, where $G = K_{2,2} \cup \overline{K_{1,1}}$, and G' is a one-factor.

Theorem 31. *[66] Let $G = (L, R; E)$ and $G' = (L', R'; E')$ be two bipartite graphs with $|L| = |L'| = p \geq 2$ and $|R| = |R'| = q \geq 2$, such that*

$$e(G) + e(G') \leq p + q + \varepsilon(G, G'), \tag{3.1}$$

where $\varepsilon(G, G') = \min\{p - \Delta_R(G), p - \Delta_{R'}(G'), q - \Delta_L(G), q - \Delta_{L'}(G')\}$.

Then G and G' pack unless either

1. $\varepsilon(G, G') = 0$ *and* $\{G, G'\} \in \Gamma_1$*, or*

2. $\varepsilon(G, G') = 1$ *and* $\{G, G'\} \in \Gamma_2 \cup \Gamma_3$*.*

Another theorem in this field is by Wang.

Theorem 32. *[63] Let $G(A, B)$ and $H(S, T)$ be two C_4-free bipartite graphs of order n with $|A| = |B| = |S| = |T| = n$, and $e(G) + e(H) \leq 2n - 2$. Then there is a packing of G and H in $K_{n+1,n+1}$ (i.e. an edge-disjoint embedding of G and H into $K_{n+1,n+1}$), unless one is a union of vertex-disjoint cycles and the other is a union of two-disjoint stars.*

For more results in this field, we refer the interested reader to the monograph on factor theory of Yu and Liu [67].

Let us formulate the result of this chapter of the thesis in the following theorem as an embedding problem.

Theorem 33. *For every $\varepsilon \in \left(0, \frac{1}{2}\right)$ there is an $n_0 = n_0(\varepsilon)$ such that if $n > n_0$, and $G(A, B)$ and $H(S, T)$ are bipartite graphs with $|A| = |B| = |S| = |T| = n$ and the following conditions hold, then $H \subseteq G$.*

Condition 1: $deg_G(x) > \left(\frac{1}{2} + \varepsilon\right) n$ *holds for all* $x \in A \cup B$

Condition 2: $deg_H(x) < \frac{\varepsilon^4}{100} \frac{n}{\log n}$ *holds for all* $x \in S$*,*

Condition 3: $deg_H(y) = 1$ *holds for all* $y \in T$*.*

In the following remarks, we show cases in which our main theorem can guarantee packings that were beyond reach by the previous techniques.

Remark 34. There are graphs which can be packed using Theorem 33, though Theorem 31 does not imply that they pack.

For instance, let $G(A, B)$ and $H(S, T)$ be bipartite graphs with $|A| = |B| = |S| = |T| = n$. Choose H to be a 1-factor, and G to be a graph such that all vertices in A have degree $\left(\frac{1}{2} + \frac{1}{100}\right) n$. This pair of graphs obviously satisfies the conditions of Theorem 33, thus, H can be embedded into G, which means that H can be packed with the bipartite complement of $\widetilde{G}$.

Now, we check the conditions of Theorem 31 for the graphs $\widetilde{G}$ and H. We know that $e(H) = n$, as H is a 1-factor. Furthermore, in $\widetilde{G}$ each vertex in A has degree $\left(\frac{1}{2} - \frac{1}{100}\right) n$, which means that the number of edges is approximately $\frac{n^2}{4}$. As $\varepsilon(H, \widetilde{G}) \leq n$, the condition of Theorem 31 is obviously not satisfied.

Remark 35. There are graphs which can be packed using Theorem 33, though Theorem 32 does not imply that they pack. Let G be the union of $\frac{n}{3}$ disjoint copies of C_6's and H be a 1-factor. Obviously, H is C_4-free, but the condition of Theorem 32 is not satisfied for G and H, as $e(G) + e(H) = 3n$.

However, our theorem can give an embedding of H into $\widetilde{G}$, as all conditions of Theorem 33 are satisfied with these graphs. This provides a packing of H and G.

The following two examples show that it is necessary to make an assumption on $\delta(G)$ (see Condition 1) and on $\Delta_S(H)$ (see Condition 2).

First, let $G = K_{\frac{n}{2}+1, \frac{n}{2}-1} \cup K_{\frac{n}{2}-1, \frac{n}{2}+1}$. Clearly, G has no perfect matching. This shows that the bound in Condition 1 is close to being best possible.

For the second example, we choose $G = G(n, n, 0.6)$ to be a random bipartite graph. Standard probability reasoning shows that with high probability, G satisfies Condition 1. However, H cannot be embedded into G, where $H(S, T)$ is the following bipartite graph: each vertex in T has degree 1. In S all vertices have degree 0, except $\frac{\log n}{c}$ vertices with degree $\frac{cn}{\log n}$ with a sufficiently large constant c. The graph H cannot be embedded into G, what follows from the example of Komlós et al. [41]. The graph H is also shown on Figure 3.1

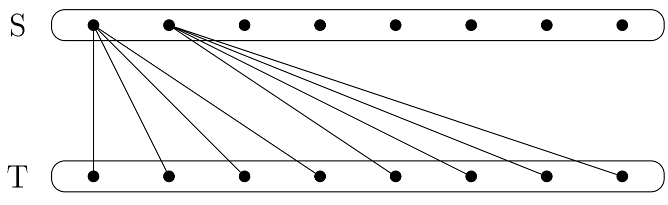

Figure 3.1: The graph H of the example for the necessity of Condition 2

3.1 The proof of Theorem 33

We will use the following lemma by Gale [29] and Ryser [55] in the form as discussed by Lovász [46].

First, we need a definition. We say that a sequence $\pi = (a_1, \ldots, a_k; b_1, \ldots, b_l)$ is *bigraphic*, if and only if there is a bipartite graph $G(A, B)$ with $|A| = k$ and $|B| = l$ realizing π such that the degrees of vertices in A are $a_1, \ldots, a_k$, and the degrees of the vertices of B are $b_1, \ldots, b_l$ [65]. In this case, we say that π is a fixed order realization of the bipartite degree sequence deg_G. Note that this notion is different from the usual degree sequence notion, which contains only an ordered list of the degrees, which are not connected to specific vertices.

Lemma 36. *[29, 55] Let $G(A, B)$ be a bipartite graph and π a bigraphic sequence on (A, B). If for all $X \subseteq A$, $Y \subseteq B$*

$$\sum_{x \in X} \pi(x) \leq e_G(X, Y) + \sum_{y \in \overline{Y}} \pi(y), \tag{3.2}$$

then π can be embedded into G.

We formulate the key technical result for the proof of Theorem 33 in the following lemma.

Lemma 37. *Let $\varepsilon \in (0, 0.5)$ and c as stated in Theorem 33. Let $G(Z, W)$ and $H(Z', W')$ be bipartite graphs with $|Z| = |Z'| = z$ and $|W| = |W'| = n$, respectively, with $z > \frac{2}{\varepsilon}$.*

Suppose that

Condition 1a: $deg_G(x) > \left(\frac{1}{2} + \varepsilon\right) n$ *for all* $x \in Z$,

Condition 1b: $deg_G(y) > \left(\frac{1}{2} + \frac{\varepsilon}{2}\right) z$ *for all* $y \in W$,

Condition 2: *There is an $M \in \mathbb{N}$ and a $0 < \delta \leq \frac{\varepsilon}{10} < \frac{1}{20}$ such that*

$$M \leq deg_H(x) \leq M(1 + \delta) \text{ for all } x \in Z', \tag{3.3}$$

and

Condition 3: *$deg_H(y) = 1$ for all $y \in W'$.*

Then there is an embedding of H into G.

Proof. We show that the conditions of Lemma 36 are satisfied.

First, assign the vertices of Z and W to the vertices and Z' and W', respectively. Then, let $\emptyset \neq X \subseteq Z$ and $\emptyset \neq Y \subseteq W$. Let $\overline{X} = Z - X$ and $\overline{Y} = W - Y$. We distinguish five cases depending on the sizes of X and Y.

In all cases we will use the obvious inequality $Mz \leq n$, as $deg_H(Z) = deg_H(W)$.

Case (a) $|X| \leq \frac{z}{2(1+\delta)}$ and $|Y| \leq \frac{n}{2}$.

We have

$$deg_H(X) \leq M(1+\delta)|X| \leq M(1+\delta)\frac{z}{2(1+\delta)} = \frac{Mz}{2} \leq \frac{n}{2}, \tag{3.4}$$

and

$$\frac{n}{2} \leq |\overline{Y}| = deg_H\left(\overline{Y}\right). \tag{3.5}$$

Therefore, $deg_H(X) \leq deg_H(\overline{Y}) + e_G(X,Y)$.

Case (b) $|X| \leq \frac{z}{2(1+\delta)}$ and $|Y| > \frac{n}{2}$.

Let $\varphi = \frac{|Y|}{n} - \frac{1}{2}$, so $|Y| = \left(\frac{1}{2} + \varphi\right) n$. Obviously, $0 \leq \varphi \leq \frac{1}{2}$.

Therefore, $deg_H(\overline{Y}) = |\overline{Y}| = \left(\frac{1}{2} - \varphi\right) n$.

We have $deg_H(X) \leq \frac{n}{2}$, as we have seen in *Case (a)*.

Using Condition 1a, we know that $deg_G(X) > \left(\frac{1}{2} + \varepsilon\right) n|X|$.

As $|\overline{Y}| = \left(\frac{1}{2} - \varphi\right) n$, we have

$$e_G(X,Y) \geq deg_G(X) - |\overline{Y}||X| > \left(\frac{1}{2} + \varepsilon\right) n|X| - \left(\frac{1}{2} - \varphi\right) n, \tag{3.6}$$

Thus,

$$e_G(X,Y) > (\varepsilon + \varphi)n|X| \geq (\varepsilon + \varphi)n, \tag{3.7}$$

we obtain $deg_H(X) \leq deg_H(\overline{Y}) + e_G(X,Y)$.

Case (c) $\frac{z}{2} \geq |X| > \frac{z}{2(1+\delta)}$ and $|Y| \leq \frac{n}{2}$.

Let $\psi = \frac{|X|}{z} - \frac{1}{2(1+\delta)}$, hence, $|X| = \left(\frac{1}{2(1+\delta)} + \psi\right) z$.

Let $\psi_0 = \frac{\delta}{2(1+\delta)} = \frac{1}{2} - \frac{1}{2(1+\delta)}$, so $\psi \leq \psi_0$. This means that $|X| = \left(\frac{1}{2} - \psi_0 + \psi\right) z$.

As $0 < \delta \leq \frac{\varepsilon}{10}$, we have $\psi_0 < \frac{\delta}{2} \leq \frac{\varepsilon}{20}$.

Let $\varphi = \frac{1}{2} - \frac{|Y|}{n}$, so $|Y| = \left(\frac{1}{2} - \varphi\right) n$. As $|Y| \leq \frac{n}{2}$, this gives $0 \leq \varphi \leq \frac{1}{2}$.

We have the following bounds:

(1) $deg_H(\overline{Y}) = |\overline{Y}| = n\left(\frac{1}{2} + \varphi\right)$

(2) As above,

$$deg_H(X) \leq M(1+\delta)|X| = Mz(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right) \leq n(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right). \tag{3.8}$$

(3) We claim that $e_G(X,Y) \geq |Y|\left(\frac{\varepsilon}{2} - \psi_0 + \psi\right) z$. Indeed, the number of neighbors of a vertex $y \in Y$ in X is at least $\left(\frac{\varepsilon}{2} + \psi - \psi_0\right) z$, considering the degree bounds of W in G.

We have to show that $deg_H(X) \leq e_G(X,Y) + deg_H(\overline{Y})$. We estimated each term, hence it is enough to prove the following:

$$n(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right) \leq n\left(\frac{1}{2} - \varphi\right)\left(\frac{\varepsilon}{2} - \psi_0 + \psi\right) z + n\left(\frac{1}{2} + \varphi\right). \tag{3.9}$$

This is equivalent to

$$\psi + \delta\psi \leq z\left(\frac{1}{2} - \varphi\right)\left(\frac{\varepsilon}{2} + \psi - \psi_0\right) + \varphi. \tag{3.10}$$

The left hand side of (3.10) is at most $\psi_0 + \delta\psi_0 \leq \frac{\delta}{2} + \frac{\delta^2}{2} \leq \delta$, as $\delta \leq \varepsilon \leq \frac{1}{2}$.

If $\varphi > \delta$, (3.10) holds, since $\frac{\varepsilon}{2} + \psi - \psi_0 \geq 0$, using $\psi_0 \leq \frac{\varepsilon}{20}$.

Otherwise, if $\varphi \leq \delta$, the right hand side of (3.10) is

$$z\left(\frac{1}{2} - \varphi\right)\left(\frac{\varepsilon}{2} + \psi - \psi_0\right) \geq \left(\frac{1}{2} - \delta\right)\left(\frac{\varepsilon}{2} - \frac{\delta}{2}\right) z. \tag{3.11}$$

We can bound each factor: $\frac{1}{2} - \delta > \frac{1}{2} - \frac{1}{20}$, $\frac{\varepsilon}{2} - \frac{\delta}{2} > \frac{\varepsilon}{2} - \frac{\varepsilon}{20}$, and $z > \frac{2}{\varepsilon}$.

Using these bounds for (3.11), we have

$$\left(\frac{1}{2} - \delta\right)\left(\frac{\varepsilon}{2} - \frac{\delta}{2}\right) z > \left(\frac{1}{2} - \frac{1}{20}\right)\left(\frac{\varepsilon}{2} - \frac{\varepsilon}{20}\right)\frac{2}{\varepsilon} = \frac{81}{200} > \frac{1}{20} > \delta. \tag{3.12}$$

This completes the proof of this case.

Case (d) $|X| > \frac{z}{2}$ and $|Y| \leq \frac{n}{2}$.

We have

(1) $deg_H(X) = deg_H(Z) - deg_H(\overline{X}) = n - deg_H(\overline{X}) \leq n - M|\overline{X}|$,

(2) $deg_H(\overline{Y}) = n - |Y|$ and

(3) $e_G(X, Y) \geq |Y|\left(|X| - \frac{z}{2} + \frac{\varepsilon z}{2}\right)$, using the degree bound on Y.

We have to show that $deg_H(X) \leq e_G(X, Y) + deg_H(\overline{Y})$. Using the estimations of the terms, all we have to check is whether

$$n - M|\overline{X}| \leq n - |Y| + |Y|\left(|X| - \frac{z}{2} + \frac{\varepsilon z}{2}\right). \tag{3.13}$$

It is equivalent to

$$0 \leq |Y|\left(|X| - \frac{z}{2} + \frac{\varepsilon z}{2} - 1\right) + M\left(z - |X|\right). \tag{3.14}$$

We know that $|X| > \frac{z}{2}$, and $\frac{\varepsilon z}{2} - 1 > 0$, and $z - |X| > 0$, which gives that (36) is true. This case is also finished.

Case (e) $|X| > \frac{z}{2(1+\delta)}$ and $|Y| > \frac{n}{2}$.

Let $\psi = \frac{|X|}{z} - \frac{1}{2(1+\delta)}$, hence, $|X| = z\left(\frac{1}{2(1+\delta)} + \psi\right)$. Let $\psi_0 = \frac{\delta}{2(1+\delta)}$, as it was defined in *Case (c)*. Again, $\psi_0 \leq \frac{\delta}{2}$. We have $0 \leq \psi \leq \frac{1}{2} + \psi_0 \leq \frac{1+\delta}{2}$.

Let $\varphi = \frac{|Y|}{n} - \frac{1}{2}$, hence, $|Y| = n\left(\frac{1}{2} + \varphi\right)$.

We have

(1) $deg_H(X) \leq zM(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right) \leq n(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right)$,

(2) $deg_H(\overline{Y}) = n\left(\frac{1}{2} - \varphi\right)$ and

(3) $e_G(X,Y) \geq z\left(\frac{1}{2(1+\delta)} + \psi\right)(\varphi+\varepsilon)n$.

We have to show again that $deg_H(X) \leq e_G(X,Y) + deg_H(\overline{Y})$. Using the estimation of the terms it is sufficient to show that

$$n(1+\delta)\left(\frac{1}{2(1+\delta)} + \psi\right) \leq n\left(\frac{1}{2} - \varphi\right) + \\ z\left(\frac{1}{2(1+\delta)+\psi}\right)(\varphi+\varepsilon)n. \tag{3.15}$$

It is equivalent to

$$\psi(1+\delta) \leq -\varphi + z\left(\frac{1}{2(1+\delta)} + \psi\right)(\varphi+\varepsilon). \tag{3.16}$$

Using $\psi \leq \frac{1+\delta}{2}$ and $\delta \leq \frac{\varepsilon}{10}$, the left hand side of (3.16) is at most

$$\frac{1+\delta}{2}(1+\delta) \leq \frac{1+\frac{1}{20}}{2}\left(1+\frac{1}{20}\right) < \frac{3}{5}, \tag{3.17}$$

as $\varepsilon \leq \frac{1}{2}$.

The right hand side of (3.16) is

$$\varphi\frac{z-2(1+\delta)}{2(1+\delta)} + \frac{z}{2(1+\delta)}\varepsilon + z\psi(\varphi+\varepsilon) \tag{3.18}$$

The first and the last term of (3.18) is always positive.

The middle term can be easily bounded since $z\varepsilon > 2$, and $\frac{1}{1+\delta} > \frac{1}{1+1/20} = \frac{20}{21}$.

This means that (3.18) is at least $\frac{20}{21}$, which is more than $\frac{3}{5}$. This finishes the proof of this case.

□

Proof. (Theorem 33) First, form a partition $C_0, C_1, \ldots, C_k$ of S in the graph H. For $i > 0$ let $u \in C_i$ if and only if

$$\frac{\varepsilon^4}{100}\frac{n}{\log n} \cdot \frac{1}{(1+\delta)^{i-1}} \geq deg_H(u) > \frac{\varepsilon^4}{100}\frac{n}{\log n} \cdot \frac{1}{(1+\delta)^{i}} \tag{3.19}$$

with $\delta = \frac{\varepsilon}{10}$. Let C_0 be the class of the isolated points in S. Note that the number of partition classes, k is $\log_{1+\delta} n = \log_{1+\frac{\varepsilon}{10}} n = \frac{\log n}{\log\left(1+\frac{\varepsilon}{10}\right)} = c \log n$.

Now, we embed the partition of S into A. Take a random ordering of the vertices of A. Say this is $(v_1, \ldots, v_n)$. The first $|C_1|$ vertices of A form A_1, the vertices $|C_1| + 1^{st}, \ldots, |C_1| + |C_2|^{th}$ form A_2 etc., while C_0 maps to the last $|C_0|$ vertices. Obviously, C_0 can be always embedded, as it contains only isolated vertices.

We say that a partition class C_i is *small* if $|C_i| \leq \frac{16}{\varepsilon^2} \log n$.

We claim that the total size of the neighborhood in B of small classes is at most $\frac{\varepsilon n}{4}$.

The size of the neighborhood of C_i is at most

$$\frac{\varepsilon^4}{100}\frac{n}{\log n} \cdot \frac{1}{(1+\delta)^{i-1}} \cdot \frac{16}{\varepsilon^2} \log n. \tag{3.20}$$

If we sum up, we have that the total size of the neighborhood of small classes is at most

$$\begin{aligned} \sum_{i=1}^{k} \frac{\varepsilon^4}{100}\frac{n}{\log n} \cdot \frac{1}{(1+\delta)^{i-1}} \cdot \frac{16}{\varepsilon^2} \log n = \frac{4}{25}\varepsilon^2 n \sum_{i=0}^{k-1} \frac{1}{(1+\delta)^{i}} \leq \\ \leq \frac{4}{25}\varepsilon^2 n \frac{1+\delta}{\delta} \leq \frac{4}{25}\varepsilon^2 n \frac{3/20}{\varepsilon/10} \leq \frac{\varepsilon n}{4}. \end{aligned} \tag{3.21}$$

The vertices of the small classes can be dealt with using a greedy method: if v_i is in a small class, choose randomly $deg_H(v_i)$ of its neighbors, and fix these edges. After we finished fixing these edges, the degrees of the vertices of B are still larger than $\left(\frac{1}{2} + \frac{\varepsilon}{2}\right) n$.

Continue with the large classes $C_{i_1}, \dots, C_{i_\ell}$ and form a random partition $E_1, \dots, E_\ell$ of the unused vertices in B such that $|E_j| = \sum\limits_{u \in C_{i_j}} deg_H(u)$. We will consider the pairs (C_{i_j}, E_j).

We will show that the conditions of Lemma 37 are satisfied for (C_{i_j}, E_j), then we apply Lemma 37 with $\frac{\varepsilon}{2}$ instead of ε, and we get an embedding in each pair (C_{i_j}, E_j), which gives an embedding of H into G.

Conditions 2 and 3 are immediate.

For Conditions 1a and 1b we have to show that for any j every vertex $y \in E_j$ has at least $\left(\frac{1}{2} + \frac{\varepsilon}{4}\right) z$ neighbors in D_i and every vertex $x \in C_{i_j}$ has at least $\left(\frac{1}{2} + \frac{\varepsilon}{2}\right) z$ in E_j.

For this, we will use the martingale technique (see [3]).

Let $|C_{i_j}| = z$. We know $z > \frac{16}{\varepsilon^2} \log n$, as C_{i_j} is large.

Let $y \in E_j$ be fixed. Consider the random variable $\mathcal{X} = |N(y) \cap C_{i_j}|$.

Define the following chain: $\mathcal{Z}_0 = \mathbb{E}\mathcal{X}, \mathcal{Z}_1 = \mathbb{E}[\mathcal{X}|v_1], \mathcal{Z}_2 = \mathbb{E}[\mathcal{X}|v_1, v_2]$; in general, $\mathcal{Z}_k = \mathbb{E}[\mathcal{X}|v_1, \dots, v_k]$ for $1 \leq k \leq n$. In other words, $\mathcal{Z}_k$ is the expectation of $\mathcal{X}$ with the condition that we already know $v_1, \dots, v_k$. This chain of random variables define a martingale (see Chapter 8.3 of the book of Matoušek and Vondrák [50]) with martingale differences $\mathcal{Z}_k - \mathcal{Z}_{k-1} \leq 1$.

According to the Azuma–Hoeffding inequality [5, 35] we have the following lemma:

Lemma 38 (Azuma [5]). *If $\mathcal{Z}$ is a martingale with martingale differences at most 1, then for any j and t the following holds:*

$$\mathbb{P}\left(\mathcal{Z}_j \geq \mathbb{E}\mathcal{Z}_j - t\right) \geq 1 - e^{-\frac{t^2}{2j}}. \tag{3.22}$$

The conditional expected value $\mathbb{E}(\mathcal{Z}_z|\mathcal{Z}_0)$ is $\mathbb{E}\mathcal{Z}_z = \left(\frac{1}{2} + \frac{3\varepsilon}{4}\right) z$.

Lemma 38 shows that

$$\mathbb{P}\left(\mathcal{Z}_z \geq \left(\frac{1}{2}+\frac{\varepsilon}{2}\right)z\right) \geq 1-e^{-\frac{\varepsilon^2 z^2/4}{2z}} = 1-e^{-\varepsilon^2 z/8}. \tag{3.23}$$

We say that a vertex $v \in E_j$ is *bad*, if it has less than $\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)z$ neighbors in C_{i_j}. Lemma 38 means that a vertex v is bad with probability at most $e^{-\varepsilon^2 z/8}$. As we have n vertices in B, the probability of the event that any vertex is bad is less than

$$n \cdot e^{-\varepsilon^2 z/8} < \frac{1}{n}, \tag{3.24}$$

as $z > \frac{16}{\varepsilon^2}\log n$.

Then we have that with probability $1-\frac{1}{n}$ no vertex in E_j is bad. Thus, Condition (ii) of Lemma 37 is satisfied with probability 1 for any pair (C_{i_j}, E_j).

Using Lemma 38, we can also show that each $x \in C_{i_j}$ has at least $\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)|E_j|$ neighbors in E_j with probability 1.

Thus, the conditions of Lemma 37 are satisfied, and we can embed H into G. The proof of Theorem 33 is finished. $\square$

Remark

In the bipartite discrete tomography problem we are given two bigraphic sequences π_1 and π_2 on the vertex set (A, B), where $|A| = |B| = n$. The goal is to color the edges of $K(A, B)$ by red, blue and grey such that for each $v \in A \cup B$ the blue degree of v is $\pi_1(v)$, its red degree is $\pi_2(v)$, and its grey degree is $n - \pi_1(v) - \pi_2(v)$.

A previous result in this field is the following theorem.

Theorem 39 (Diemunsch et al. [24]). *Let π_1 and π_2 be bigraphic sequences with parts of sizes r and s, and $\Delta_i = \Delta(\pi_i)$ and $\delta_i = \delta(\pi_i)$ for $i = 1, 2$ such that $\Delta_1 \leq \Delta_2$ and $\delta_1 \geq 1$. If*

$$\Delta_1\Delta_2 \leq \delta_1\frac{r+s}{8}, \tag{3.25}$$

then π_1 *and* π_2 *pack.*

In Theorem 33, we study an "ordinary" packing problem. However, inspecting the proof one obtains the following result in discrete tomography.

Assume the conditions of Theorem 33. Let π_1 be the bipartite degree sequence of $\widetilde{G}$, and π_2 be the bipartite degree sequence of H. Consider a fixed order realization $\widetilde{\pi}_1$ and $\widetilde{\pi}_2$ of them, where $\widetilde{\pi}_1$ is an arbitrary, and $\widetilde{\pi}_2$ is a random realization. Then, with probability tending to 1, $\widetilde{\pi}_1$ and $\widetilde{\pi}_2$ pack.

Hence, in certain cases for most orderings we can improve the bounds of Theorem 39.

Chapter 4

Embedding degree sequences

In this chapter, we deal with an embedding question of degree sequences and graphs, presenting the results of [21, 22]. The main result is formulated in the following theorem.

Theorem 40. *For every $\eta > 0$ and $D \in \mathbb{N}$ there exists an $n_0 = n_0(\eta, D)$ such that for all $n > n_0$ if G is a graph on n vertices with $\delta(G) \geq \left(\frac{1}{2} + \eta\right) n$ and π is a degree sequence of length n with $\Delta(\pi) \leq D$, then π is embeddable into G.*

It is easy to see that Theorem 40 is sharp up to the ηn additive term. For that let n be an even number, and suppose that every element of π is 1. Then the only graph that realizes π is the union of $n/2$ vertex disjoint edges. Let $G = K_{n/2-1,n/2+1}$ be the complete bipartite graph with vertex class sizes $n/2 - 1$ and $n/2 + 1$. Clearly G does not have $n/2$ vertex disjoint edges.

In order to state the other main result of this chapter we introduce a new notion.

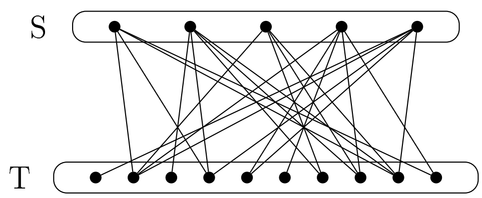

Figure 4.1: A 2-unbalanced bipartite graph

Definition 41. Let $q \geq 1$ be an integer. A bipartite graph H with vertex classes S and T is q-unbalanced, if $q|S| \leq |T|$. See also Figure 4.1. The degree sequence π is q-unbalanced, if it can be realized by a q-unbalanced bipartite graph.

Theorem 42. *Let $q \geq 1$ be an integer. For every $\eta > 0$ and $D \in \mathbb{N}$ there exist an $n_0 = n_0(\eta, q)$ and an $M = M(\eta, D, q)$ such that if $n \geq n_0$, and π is a q-unbalanced degree sequence of length $n - M$ with $\Delta(\pi) \leq D$, and G is a graph on n vertices with $\delta(G) \geq \left(\frac{1}{q+1} + \eta\right) n$, then π can be embedded into G.*

Hence, if π is unbalanced, the minimum degree requirement of Theorem 40 can be substantially decreased. What we pay for this is that π has to be slightly smaller than the number of vertices in the host graph.

4.1 Proof of Theorem 40

Proof. First, we find a suitable realization H of π. Our H will consist of components of bounded size. Second, we embed H into G using a theorem by Chvátal and Szemerédi and a result on embedding so-called well-separable graphs. The details are given in the following.

We construct H in several steps. At the beginning, let H be the empty graph and let all degrees in π be *active*.

While we can find $2i$ active degrees of π with value i (for some $1 \leq i \leq \Delta(\pi)$), we realize them with a $K_{i,i}$, i.e. we add this complete bipartite graph to H, and "inactivate" the $2i$ degrees. When we stop we have at most $\sum_{i=1}^{\Delta(\pi)} (2i - 1)$ active degrees.

This way we obtain several components, each being a balanced complete bipartite graph. These are the *type 1 gadgets*. Observe that if a vertex v belongs to some type 1 gadget, then its degree is exactly $\pi(v)$. Observe further that if there are no active degrees in π at this point then the graph H we have just found is a realization of π.

Assume that there are active degrees left in π. Let $R = R_{odd} \cup R_{even}$ be the vertex set that is identified with the active vertices ($v \in R_{odd}$ if and only if the assigned active degree is odd). Since $\sum_{v \in R} d(v)$ must be an even number we have that $|R_{odd}|$ is even. Add a perfect matching on R_{odd} to H. With this we achieved that every vertex of R misses an even number of edges.

Next, we construct the *type 2 gadgets* using the following algorithm. In the beginning every type 1 gadget is *unmarked*.

Suppose that $v \in R$ is an active vertex. Take a type 1 gadget K, *mark* it, and let M_K denote an arbitrarily chosen perfect matching in K (M_K exists since K is a balanced complete bipartite graph). Let xy be an arbitrary edge in M_K. Delete the xy edge and add the new edges vx and vy. While v is missing edges repeat the above procedure with edges of M_K, until M_K becomes empty. If M_K becomes empty, take a new unmarked type 1 gadget L, and repeat the method with L. It is easy to see that in $\pi(v)/2$ steps v reaches its desired degree and gets inactivated. Clearly, the degrees of vertices in the marked type 1 gadgets have not changed.

Figure 4.2 shows examples of type 2 gadgets. In the upper one, two vertices of R_{odd} were first connected by an edge and then two type 1 gadgets were used so that they could reach their desired degree, while in the lower one, we used three type 1 gadgets for a vertex of R. The numbers at the vertices indicate the colors in the 3-coloring of H.

Let $F \subseteq H$ denote the set of vertices containing the union of all type 2 gadgets. Observe that type 2 gadgets of F are 3-chromatic, and all have less than $5\Delta^2(\pi)$ vertices.

Let us summarize our knowledge about H for later reference.

Claim 43. *(1)* $|F| \leq 5\Delta^3(\pi)$,

(2) the components of $H[V - F]$ *are balanced complete bipartite graphs, each having size at most* $2\Delta(\pi)$,

(3) $\chi(H[F]) \leq 3$, *and*

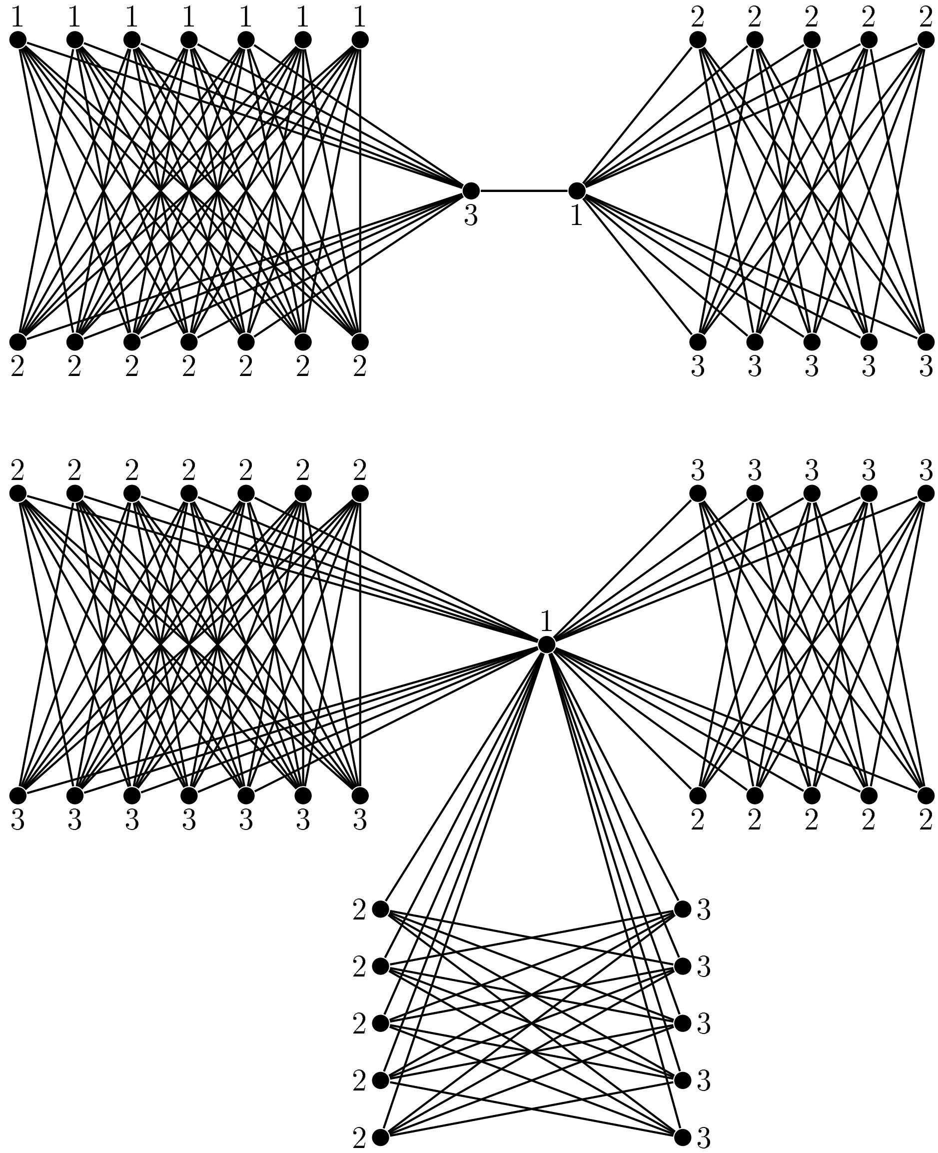

Figure 4.2: Type 2 gadgets of H with a 3-coloring

(4) $e(H[F, V - F]) = 0$.

We are going to show that $H \subseteq G$. For that we first embed the 3-chromatic part $H[F]$ using the following strengthening of the Erdős–Stone theorem proved by Chvátal and Szemerédi [14].

Theorem 44. *Let $\varphi > 0$ and assume that G is a graph on n vertices where n is sufficiently large. Let $r \in \mathbb{N}$, $r \geq 2$. If*

$$e(G) \geq \left(\frac{r-2}{2(r-1)} + \varphi\right) n^2, \tag{4.1}$$

then G contains a $K_r(t)$, i.e. a complete r-partite graph with t vertices in each class, such that

$$t > \frac{\log n}{500 \log \frac{1}{\varphi}}. \tag{4.2}$$

Since $\delta(G) \geq n/2 + \eta n$, the conditions of Theorem 44 are satisfied with $r = 3$ and $\varphi = \eta/2$, hence, G contains a balanced complete tripartite subgraph T on $\Omega(\log n)$ vertices. Using Claim 43 and the 3-colorability of F this implies that $H[F] \subseteq T$.

Observe that after embedding $H[F]$ into G every uncovered vertex of G still has at least $\delta(G) - v(F) > \left(\frac{1}{2} + \frac{\eta}{2}\right) n$ uncovered neighbors. Denoting the subgraph of the uncovered vertices of G by G' we obtain that $\delta(G') > \left(\frac{1}{2} + \frac{\eta}{2}\right) n$.

In order to prove that $H[V - F] \subseteq G'$ we first need a definition.

Definition 45. A graph L on n vertices is *well-separable*, if it has a subset $S \subseteq V(L)$ of size $o(n)$ such that all components of $L - S$ are of size $o(n)$. See also Figure 4.3

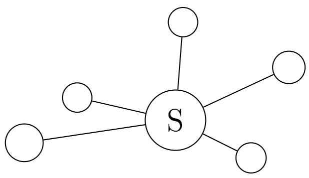

Figure 4.3: Separator set in a well-separable graph

We need the following theorem.

Theorem 46. *For every $\gamma > 0$ and positive integer D there exists an n_0 such that for all $n > n_0$ if F is a bipartite well-separable graph on n vertices, $\Delta(F) \leq D$ and $\delta(G) \geq \left(\frac{1}{2} + \gamma\right) n$ for a graph G of order n, then $F \subseteq G$.*

Since $H[V - F]$ has bounded size components by Claim 43, we can apply Theorem 46 for $H[V - F]$ and G', with parameter $\gamma = \eta/2$. With this we finished proving what was desired. $\square$

4.2 Proof of Theorem 42

When proving Theorem 40, we used the Regularity Lemma of Szemerédi (see Page 9), but implicitly, via the result on embedding well-separable graphs. When proving Theorem 42, we will apply this very powerful result explicitly. Let us give a brief sketch first.

Recall that π is a q-unbalanced and bounded degree sequence with $\Delta(\pi) \leq D$. In the proof we first show that there exists a q-unbalanced bipartite graph H realizing π such that H is the vertex disjoint union of the graphs $H_1, \ldots, H_k$, where each H_i graph is a bipartite q-unbalanced graph having bounded size. We will apply the Regularity Lemma to G and find a special substructure (a decomposition into vertex-disjoint stars) in the reduced graph of G. This substructure can then be used to embed the union of the H_i graphs, for the majority of them we use the Blow-up Lemma.

4.2.1 Finding H

The goal of this subsection is to prove the lemma below.

Lemma 47. *Let π be a q-unbalanced bipartite degree sequence of positive integers with $\Delta(\pi) \leq D$. Then π can be realized by a q-unbalanced bipartite graph H which is the vertex disjoint union of the graphs $H_1, \ldots, H_k$, such that for every i we have that H_i is q-unbalanced, moreover, $v(H_i) \leq 4D^2$.*

Before starting the proof of Lemma 47, we list a few necessary notions and results.

We call a finite sequence of integers a *zero-sum sequence* if the sum of its elements is zero. The following result of Sahs, Sissokho and Torf plays an important role in the proof of Lemma 47.

Proposition 48. *[56] Assume that K is a positive integer. Then any zero-sum sequence on $\{-K, \dots, K\}$ having length at least $2K$ contains a proper nonempty zero-sum subsequence.*

The following result formulated by Gale [29] and Ryser [55] is a consequence of Lemma 36. We present it in the form as discussed by Lovász [46].

Lemma 49. *[29, 55] Let $G = (A, B; E(G))$ be a bipartite graph and f be a non-negative integer function on $A \cup B$ with $f(A) = f(B)$. Then G has a subgraph $F = (A, B; E(F))$ such that $deg_F(x) = f(x)$ for all $x \in A \cup B$ if and only if*

$$f(X) \leq e(X, Y) + f(\overline{Y}) \tag{4.3}$$

for any $X \subseteq A$ and $Y \subseteq B$, where $\overline{Y} = B - Y$.

We remark that such a subgraph F is also called an f-factor of G.

Lemma 50. *If $f = (a_1, \dots, a_s; b_1, \dots, b_t)$ is a sequence of positive integers with $s, t \geq 2\Delta^2$, where Δ is the maximum of f, and $f(A) = f(B)$ with $A = \{a_1, \dots, a_s\}$ and $B = \{b_1, \dots, b_t\}$ then f is bigraphic.*

Proof. We only have to check whether the conditions of Lemma 49 are met if $G = K_{s,t}$.

Suppose indirectly that there is an (X, Y) pair for which (4.3) does not hold. Choose such a pair with minimal $|X| + |Y|$. Then $X = \emptyset$ or $Y = \emptyset$ are impossible, as in those cases (4.3) trivially holds. Hence, $|X|, |Y| \geq 1$. Assuming that (4.3) does not hold, we have that

$$f(X) \geq e(X, Y) + f(\overline{Y}) + 1, \tag{4.4}$$

which is equivalent to

$$f(X) \geq |X||Y| + f(\overline{Y}) + 1, \tag{4.5}$$

as G is a complete bipartite graph. Furthermore, using the minimality of $|X|+|Y|$, we know that

$$f(X - a) \leq |X - a||Y| + f(\overline{Y}) \tag{4.6}$$

for any $a \in X$. (4.6) is equivalent to

$$f(X) - f(a) \leq |X||Y| - |Y| + f(\overline{Y}). \tag{4.7}$$

From (4.5) and (4.7) we have

$$f(a) - 1 \geq |Y| \tag{4.8}$$

for any $a \in X$, which implies

$$\Delta > |Y|. \tag{4.9}$$

The same reasoning also implies that $\Delta > |X|$ whenever (X, Y) is a counterexample. Therefore we only have to verify that (4.3) holds in case $|X| < \Delta$ and $|Y| < \Delta$. Recall that $f(B) \geq t$, as all elements of f are positive. Hence, $f(X) \leq \Delta|X| \leq \Delta^2$, and $f(\overline{Y}) \geq f(B) - f(Y) \geq t - \Delta^2$, and we get that

$$f(X) \leq \Delta^2 \leq t - \Delta^2 \leq f(\overline{Y}) \leq f(\overline{Y}) + e_G(X, Y) \tag{4.10}$$

holds, since $t \geq 2\Delta^2$. □

Proof. (Lemma 47) Assume that $J = \big(S, T; E(J)\big)$ is a q-unbalanced bipartite graph realizing π. Hence, $q|S| \leq |T|$. Moreover, $|T| \leq D|S|$, since $\Delta(\pi) \leq D$. We form vertex disjoint tuples of the form $(s; t_1, \ldots, t_h)$, such that $s \in S$, $t_i \in T$, $q \leq h \leq D$, and the collection of these tuples contains every vertex of $S \cup T$ exactly once. We define the *bias* of the tuple as

$$\zeta = \pi(t_1) + \cdots + \pi(t_h) - \pi(s). \tag{4.11}$$

Obviously, $-D \leq \zeta \leq D^2$. The conditions of Proposition 48 are clearly met with $K = D^2$. Hence, we can form groups of size at most $2D^2$ in which the sums of biases are zero. This way we obtain a partition of (S, T) into q-unbalanced set pairs which have zero bias. While these sets may be small, we can combine them so that each combined set is of size at least $2D^2$ and has zero bias. By Lemma 50 these are bigraphic sequences. The realizations of these small sequences give the graphs $H_1, \ldots, H_k$. It is easy to see that $v(H_i) \leq 4D^2$ for every $1 \leq i \leq k$. Finally, we let $H = \cup_i H_i$. □

4.2.2 Decomposing G_r

Let us apply the Regularity Lemma (Lemma 7) with parameters $0 < \varepsilon \ll d \ll \eta$. By Corollary 9 we have that $\delta(G_r) \geq \ell/(q+1) + \eta\ell/2$.

Let $h \geq 1$ be an integer. An h-star is a $K_{1,h}$. The *center* of an h-star is the vertex of degree h, the other vertices are the *leaves*. In case $h = 1$ we pick one of the vertices of the 1-star arbitrarily to be the center.

Lemma 51. *The reduced graph G_r has a decomposition $\mathcal{S}$ into vertex disjoint stars such that each star has at most q leaves.*

Proof. Take a partial star-decomposition of G_r as large as possible. Assume that there are uncovered vertices in G_r. Let U denote the set of covered vertices (we assume that U has maximal cardinality), and let v be an uncovered vertex. See Figure 4.4 for the possible neighbors of v. Observe that v has neighbors only in U, otherwise, if $uv \in E(G_r)$ with $u \notin U$, then we can simply add uv to the star-decomposition, contradicting to the maximality of U.

a) If v is connected to a 1-star, then we can replace it with a 2-star.

b) If v is connected to the center u of an h-star, where $h < q$, then we can replace this star with an $h+1$-star by adding the edge uv to the h-star.

c) If v is connected to a leaf u of an h-star, where $2 \leq h \leq q$, then replace the star with the edge uv and an $(h-1)$-star (i.e., delete u from it).

We have not yet considered one remaining case: when v is connected to the center of a q-star (case d)). However, simple calculation shows that for every vertex v at least one of the above three cases must hold, using the minimum degree condition of G_r. Hence we can increase the number of covered vertices. We arrived at a contradiction, G_r has the desired star-decomposition. □

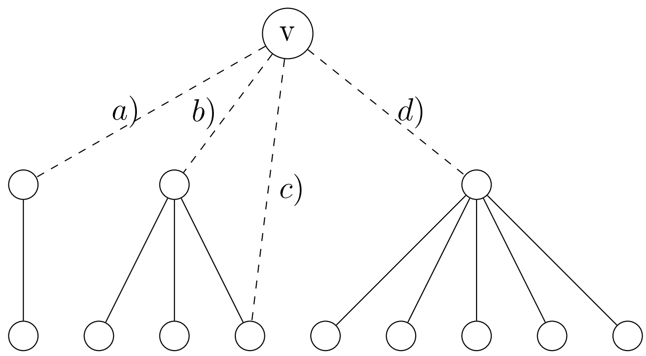

Figure 4.4: An illustration for Lemma 51

4.2.3 Preparing G for the embedding

Consider the q-star-decomposition $\mathcal{S}$ of G_r as in Lemma 51. Let ℓ_i denote the number of $(i-1)$-stars in the decomposition for every $2 \leq i \leq q+1$. It is easy to see that

$$\sum_{i=2}^{q+1} i\ell_i = \ell. \tag{4.12}$$

First we will make every ε-regular pair in $\mathcal{S}$ super-regular by discarding a few vertices from the non-exceptional clusters. Let for example $\mathcal{C}$ be a star in the decomposition of G_r with center cluster A and leaves $B_1, \ldots, B_k$, where $1 \leq k \leq q$. Recall that the (A, B_i) pairs has density at least d. We repeat the following for every $1 \leq i \leq k$: if $v \in A$ such that v has at most $2dm/3$ neighbors in B_i then discard v from A, put it into W_0. Similarly, if $w \in B_i$ has at most $2dm/3$ neighbors in A, then discard w from B_i, put it into W_0. Repeat this process for every star in $\mathcal{S}$. We have the following:

Claim 52. *We do not discard more than $q\varepsilon m$ vertices from any non-exceptional cluster.*

Proof. Given a star $\mathcal{C}$ in the decomposition $\mathcal{S}$ assume that its center cluster is A and let B be one of its leaves. Since the pair (A, B) is ε-regular with density at least d, neither A, nor B can have more than εm vertices that have at most $2dm/3$ neighbors in the opposite cluster. Hence, during the above process we may discard up to $q\varepsilon m$ vertices from A. Next, we may discard vertices from the leaves, but since no leaf B had more than εm vertices with less than $(d-\varepsilon)m$ neighbors in A, even after discarding at most $q\varepsilon m$ vertices of A, there can be at most εm vertices in B that have less than $(d-(q+1)\varepsilon)m$ neighbors in A. Using that $\varepsilon \ll d$, we have that $(d-(q+1)\varepsilon) > 2d/3$. We obtained what was desired. □

By the above claim we can make every ε-regular pair in $\mathcal{S}$ a $(2\varepsilon, 2d/3)$-super-regular pair so that we discard only relatively few vertices. Notice that we only have an upper bound for the number of discarded vertices, there can be clusters from which we have not put any points into W_0. We repeat the following for every non-exceptional cluster: if s vertices were discarded from it with $s < q\varepsilon m$ then we take $q\varepsilon m - s$ arbitrary vertices of it, and place them into W_0. This way every non-exceptional cluster will have the same number of points, precisely $m - q\varepsilon m$. For simpler notation, we will use the letter m for this new cluster size. Observe that W_0 has increased by $q\varepsilon m\ell$ vertices, but we still have $|W_0| \leq 3dn$ since $\varepsilon \ll d$ and $\ell m \leq n$. Since $q\varepsilon m \ll d$, in the resulting pairs the minimum degree will be at least $dm/2$.

Summarizing, we obtained the following:

Lemma 53. *By discarding a total of at most $q\varepsilon n$ vertices from the non-exceptional clusters we get that every edge in $\mathcal{S}$ represents a $(2\varepsilon, d/2)$-super-regular pair, and* *all non-exceptional clusters have the same cardinality, which is denoted by* m*. Moreover,* $|W_0| \leq 3dn$*.*

Since $v(G) - v(H)$ is bounded above by a constant, when embedding H we need almost every vertex of G, in particular those in the exceptional cluster W_0. For this reason we will assign the vertices of W_0 to the stars in $\mathcal{S}$. This is not done in an arbitrary way.

Definition 54. Let $v \in W_0$ be a vertex and (Q, T) be an ε-regular pair. We say that $v \in T$ *has large degree to* Q if v has at least $\eta|Q|/4$ neighbors in Q. Let $S = (A, B_1, \ldots, B_k)$ be a star in $\mathcal{S}$ where A is the center of S and $B_1, \ldots, B_k$ are the leaves, here $1 \leq k \leq q$. If v has large degree to any of $B_1, \ldots, B_k$, then v *can be assigned to* A. If $k < q$ and v has large degree to A, then v *can be assigned to* any of the B_i leaves.

Observation 55. *If we assign new vertices to a* q*-star, then we necessarily assign them to the center. Since before assigning, the number of vertices in the leaf-clusters is exactly* q *times the number of vertices in the center cluster, after assigning new vertices to the star,* q *times the cardinality of the center will be larger than the total number of vertices in the leaf-clusters. If* $S \in \mathcal{S}$ *is a* k*-star with* $1 \leq k < q$*, and we assign up to* cm *vertices to any of its clusters, where* $0 < c \ll 1$*, then even after assigning new vertices we will have that* q *times the cardinality of the center is larger than the total number of vertices in the leaf-clusters.*

The following lemma plays a crucial role in the embedding algorithm.

Lemma 56. *Every vertex of* W_0 *can be assigned to at least* $\eta\ell/4$ *non-exceptional clusters.*

Proof. Suppose that there exists a vertex $w \in W_0$ that can be assigned to less than $\eta\ell/4$ clusters. If w cannot be assigned to any cluster of some k-star S_k with $k < q$, then the total degree of w into the clusters of S_k is at most $k\eta m/4$. If w cannot be assigned to any cluster of some q-star S_q, then the total degree of w into the clusters of S_q is at most $m + q\eta m/4$, since every vertex of the center cluster could be adjacent to w. Considering that w can be assigned to at most $\eta\ell/4 - 1$ clusters and that $deg(w, W - W_0) \geq n/(q+1) + \eta n/2$, we obtain the following inequality:

$$\frac{n}{q+1} + \frac{\eta n}{2} \leq deg(v, W - W_0) \leq \eta\frac{\ell m}{4} + \sum_{k=1}^{q-1}(k+1)\eta\frac{\ell_{k+1}m}{4} + q\eta\frac{\ell_{q+1}m}{4} + \ell_{q+1}m. \tag{4.13}$$

Using $m\ell \leq n$ and $\sum_{k=1}^{q}(k+1)\ell_{k+1} = \ell$, we get

$$\frac{m\ell}{q+1} + \frac{\eta m\ell}{2} \leq \eta\frac{\ell m}{4} + (\ell - \ell_{q+1})\frac{\eta m}{4} + q\eta\frac{\ell_{q+1}m}{4} + \ell_{q+1}m. \tag{4.14}$$

Dividing both sides by m and cancellations give

$$\frac{\ell}{q+1} \leq q\frac{\eta\ell_{q+1}}{4} + \left(1 - \frac{\eta}{4}\right)\ell_{q+1}. \tag{4.15}$$

Noting that $(q+1)\ell_{q+1} \leq \ell$, one can easily see that we arrived at a contradiction. Hence every vertex of W_0 can be assigned to several non-exceptional clusters. □

Lemma 56 implies the following:

Lemma 57. *One can assign the vertices of W_0 so that at most $\sqrt{d}m$ vertices are assigned to non-exceptional clusters.*

Proof. Since we have at least $\eta\ell/4$ choices for every vertex, the bound follows from the inequality $\frac{4|W_0|}{\eta\ell} \leq \sqrt{d}m$, where we used that $d \ll \eta$ and that $|W_0| \leq 3dn$. □

Observation 58. *A key fact is that the number of newly assigned vertices to a cluster is much smaller than their degree into the opposite cluster of the regular pair since $\sqrt{d}m \ll \eta m/4$.*

4.2.4 The embedding algorithm

The embedding is done in two phases. In the first phase we cover every vertex that belonged to W_0, together with some other vertices of the non-exceptional clusters. In the second phase we are left with super-regular pairs into which we embed what is left from H using the Blow-up Lemma.

The first phase

Let (A, B) be an ε-regular cluster-edge in the h-star $\mathcal{C} \in \mathcal{S}$. We begin with partitioning A and B randomly, obtaining $A = A' \cup A''$ and $B = B' \cup B''$ with $A' \cap A'' = B' \cap B'' = \emptyset$. For every $w \in A$ (except those that came from W_0) flip a coin. If it is heads, we put w into A', otherwise we put it into A''. Similarly, we flip a coin for every $w \in B$ (except those that came from W_0) and depending on the outcome, we either put the vertex into B' or into B''. The proof of the following lemma is standard, uses Chernoff's bound (see in [3]).

Lemma 59. *With high probability, that is, with probability at least* $1 - 1/n$*, we have the following:*

- $\big||A'| - |A''|\big| = o(n)$ *and* $\big||B'| - |B''|\big| = o(n)$
- $deg(w, A'), deg(w, A'') > deg(w, A)/3$ *for every* $w \in B$
- $deg(w, B'), deg(w, B'') > deg(w, B)/3$ *for every* $w \in A$
- *the density* $d(A', B') \geq d/2$

It is easy to see that Lemma 59 implies that (A', B') is a $(5\varepsilon, d/6)$-super-regular pair having density at least $d/2$ with high probability.

Assume that v was an element of W_0 before we assigned it to the cluster A, and assume further that $deg(v, B) \geq \eta m/4$. Since (A, B) is an edge of the star-decomposition, either A or B must be the center of $\mathcal{C}$.

Let H_i be one of the q-unbalanced bipartite subgraphs of H that has not been embedded yet. We will use H_i to cover v. Denote S_i and T_i the vertex classes of H_i, where $|S_i| \geq q|T_i|$. Let $S_i = \{x_1, \ldots, x_s\}$ and $T_i = \{y_1, \ldots, y_t\}$.

If A is the center of $\mathcal{C}$ then the vertices of T_i will cover vertices of A', and the vertices of S_i will cover vertices of B'. If B is the center, S_i and T_i will switch roles. The embedding of H_i is essentially identical in both cases, so we will only discuss the case when A is the center. (Recall that if $h < q$ then we may assigned v to a leaf, so in such a case B could be the center.)

In order to cover v we will essentially use a well-known method called Key Lemma in [45]. We will heavily use the fact that

$$0 < \varepsilon \ll d \ll \eta. \tag{4.16}$$

The details are as follows. We construct an edge-preserving injective mapping $\varphi : S_i \cup T_i \longrightarrow A' \cup B'$. In particular, we will have $\varphi(S_i) \subseteq B'$ and $\varphi(T_i) - v \subseteq A'$. First we let $\varphi(y_1) = v$. Set $N_1 = N(v) \cap B'$. Using Lemma 59, we have that $|N_1| \geq \eta m/12 \gg \varepsilon m$.

Next we find $\varphi(y_2)$. Since $|N_1| \gg \varepsilon m$, by 5ε-regularity the majority of the vacant vertices of A' will have at least $d|N_1|/3$ neighbors in N_1. Pick any of these, denote it by v_2 and let $\varphi(y_2) = v_2$. Also, set $N_2 = N_1 \cap N(v_2)$.

In general, assume that we have already found the vertices $v_2, v_3, \ldots, v_i$, their common neighborhood in B' is N_i, and

$$|N_i| \geq \frac{\eta d^{i-1}}{3^{i-2} \cdot 36} m \gg \varepsilon m. \tag{4.17}$$

By 5ε-regularity, this implies that the majority of the vacant vertices of A' has large degree into N_i, at least $d|N_i|/3$, and this, as above, can be used to find v_{i+1}. Then we set $\varphi(y_{i+1}) = v_{i+1}$. Since η and d is large compared to ε, even into the last set N_{t-1} many vacant vertices will have large degrees.

As soon as we have $\varphi(y_1), \ldots, \varphi(y_t)$, it is easy to find the images for $x_1, \ldots, x_t$. Since $|N_t| \gg \varepsilon m \gg s = |S_i|$, we can arbitrarily choose s vacant points from N_t for the $\varphi(x_j)$ images.

Note that we use less than $v(H_i) \le 4D^2$ vertices from A' and B' during this process. We can repeat it for every vertex that were assigned to A, and still at most $\sqrt{d}2D^2m$ vertices will be covered from A' and from B'.

Another observation is that every h-star in the decomposition before this embedding phase was h-unbalanced, now, since we were careful, these have become h'-balanced with $h' \le h$.

Of course, the above method will be repeated for every (A, B) edge of the decomposition to which we have assigned vertices of W_0.

The second phase

In the second phase we first unite all the randomly partitioned clusters. For example, assume that after covering the vertices coming from W_0 the set of vacant vertices of A' is denoted by A'_v. Then we let $A_v = A'_v \cup A''$, and using analogous notation, let $B_v = B'_v \cup B''$.

Claim 60. *All the (A_v, B_v) pairs are $(3\varepsilon, d/6)$-super-regular with density at least $d/2$.*

Proof. The 3ε-regularity of these pairs is easy to see, like the lower bound for the density, since we have only covered relatively few vertices of the clusters. For the large minimum degrees note that by Lemma 59 every vertex of A had at least $dm/6$ neighbors in B'', hence, in B_v as well, and analogous bound holds for vertices of B. □

At this point we want to apply the Blow-up Lemma for every star of $\mathcal{S}$ individually. For that, we first have to assign those subgraphs of H to stars that were not embedded yet. We need a lemma.

Lemma 61. *Let $K_{a,b}$ be a complete bipartite graph with vertex classes A and B, where $|A| = a$ and $|B| = b$. Assume that $a \le b = ha$, where $1 \le h \le q$. Let H' be* *the vertex disjoint union of q-unbalanced bipartite graphs:*

$$H' = \bigcup_{j=1}^{t} H_j, \tag{4.18}$$

such that $v(H_j) \le 2D^2$ *for every* j*. If* $v(H') \le a+b-4(2q+1)D^2$*, then* $H' \subseteq K_{a,b}$*.*

Observe that if we have Lemma 61, we can distribute the H_i subgraphs among the stars of $\mathcal{S}$, and then apply the Blow-up Lemma. Hence, we are done with proving Theorem 42, if we prove Lemma 61 above.

Proof. The proof is an assigning algorithm and its analysis. We assign the vertex classes of the H_j subgraphs to A and B, one-by-one. Before assigning the jth subgraph H_j, the number of vacant vertices of A is denoted by a_j and the number of vacant vertices of B is denoted by b_j.

Assume that we want to assign H_k. If $ha_k - b_k > 0$, then the larger vertex class of H_k is assigned to A, the smaller is assigned to B. Otherwise, if $ha_k - b_k \le 0$, then we assign the larger vertex class to B and the smaller one to A. Then we update the number of vacant vertices of A and B. Observe that using this assigning method we always have $a_k \le b_k$.

The question is whether we have enough room for H_k. If $ha \ge 4hD^2$, then we must have enough room, since $b_k \ge a_k$ and every H_j has at most $2D^2$ vertices. Hence, if the algorithm stops, we must have $a_k < 4D^2$. Since $b_k - ha_k \le 2D^2$ must hold, we have $b_k < (2h+1)2D^2 < (2q+1)2D^2$. From this the lemma follows. □

4.3 Remarks

One can prove a very similar result to Theorem 42, in fact the result below follows easily from it. For stating it we need the notion of graph edit distance which is defined in [49] as it follows: the edit distance between two graphs on the same labeled vertex set is the size of the symmetric difference of the edge sets.

Theorem 62. *Let $q \geq 1$ be an integer. For every $\eta > 0$ and $D \in \mathbb{N}$ there exist an $n_0 = n_0(\eta, q)$ and a $K = K(\eta, D, q)$ such that if $n \geq n_0$, π is a q-unbalanced degree sequence of length n with $\Delta(\pi) \leq D$, G is a graph on n vertices with $\delta(G) \geq \left(\frac{1}{q+1} + \eta\right) n$, then there exists a graph G' on n vertices such that the edit distance of G and G' is at most K, and π can be embedded into G'.*

Here is an example showing that Theorem 42 and 62 are essentially best possible.

Example 63. *Assume that π has only odd numbers and G has at least one odd sized component. The embedding is impossible. Indeed, any realization of π has only even sized components, hence G cannot contain it as a spanning subgraph.*

Note that this example does not work in case G is connected. In Theorem 40 the minimum degree $\delta(G) \geq n/2 + \eta n$, hence, G is connected, and in this case we can embed π into G.

Chapter 5

On the relation of separability and bandwidth

In this chapter we consider a third embedding/packing problem, and we present the results of [20].

The famous Bollobás-Eldridge-Catlin (BEC) Conjecture [7, 12] below is among the most important conjectures in the area:

Conjecture 64 (Bollobás, Eldridge; Catlin). *If G_1 and G_2 are graphs on n vertices with maximum degree Δ_1 and Δ_2, respectively, and*

$$(\Delta_1 + 1)(\Delta_2 + 1) \leq n + 1,$$

then G_1 and G_2 pack.

Since the above conjecture is open in general, we impose further conditions for H and G in order to be able to solve special cases of the problem. One possibility is to consider only bounded degree H graphs to be embedded. The BEC Conjecture was solved in case $\Delta(H) = 2$ [2], $\Delta(H) = 3$ [19], and when $\Delta(H)$ is bounded and H is bipartite [17]. There is an approximation result in which $(\Delta_1 + 1)(\Delta_2 + 1) \leq 0.6n$ [38].

One may impose other restrictions on H still obtaining hard (but somewhat easier) problems. For example one may upper bound the so-called bandwidth of H, this guarantees that H is "far from being an expander".

Definition 65. Let $H(V, E)$ be a graph. Let $\mathcal{F} = \big\{f : V \to \{1, \ldots, n\}\big\}$ be a family of bijective functions on V. The *bandwidth* of H is

$$\varphi(H) = \min_{f \in \mathcal{F}} \max_{v_i v_j \in E} \{|f(v_i) - f(v_j)|\}.$$

See also Figure 5.1.

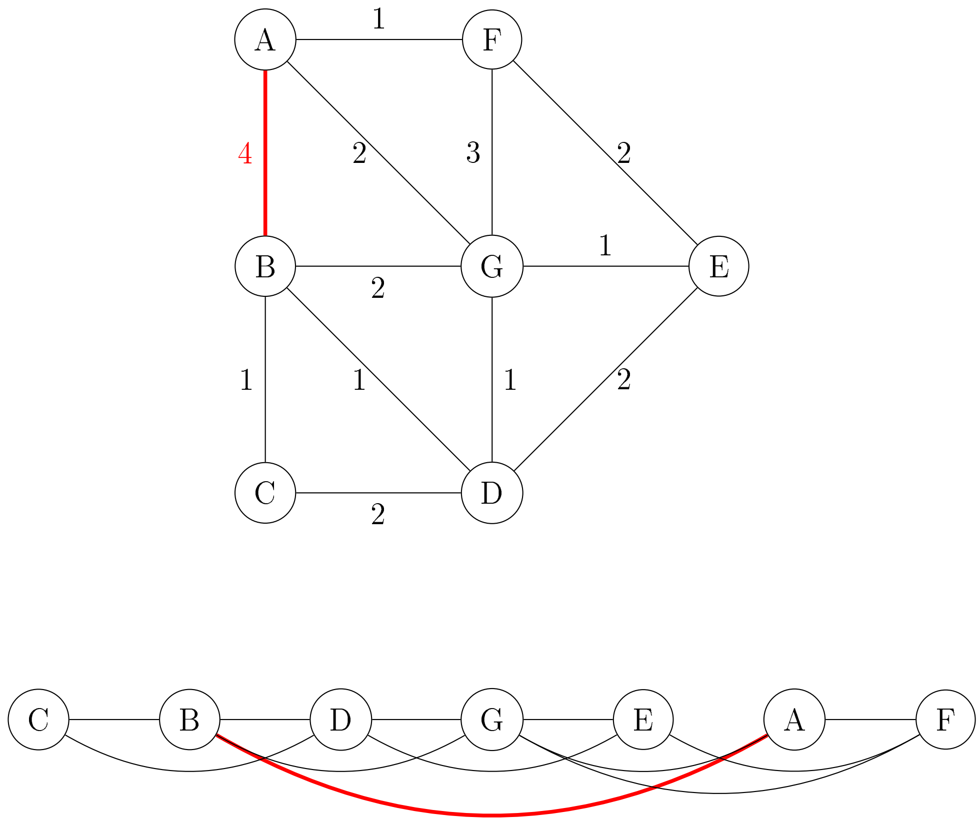

Figure 5.1: Bandwidth. Figure based on [53]

Note that a Hamilton path has bandwidth 1, a Hamilton cycle has bandwidth 2. Expander graphs have large, linear bandwidth, a star on n vertices has bandwidth

$n/2$, a complete graph has bandwidth $n-1$. One of the important open problems of the area was the following conjecture by Bollobás and Komlós.

Conjecture 66 (Bollobás, Komlós)**.** *For every D, k, ε there exists β such that the following holds. Every n-vertex graph G of minimum degree at least $(1-1/k+\varepsilon)n$ contains all k-chromatic n-vertex graphs of maximum degree at most D and bandwidth at most βn as subgraphs.*

This conjecture was proved by Böttcher, Schacht and Taraz [10] using deep tools, in particular the proof of the celebrated Pósa-Seymour conjecture by Komlós, Sárközy and Szemerédi [43].

In [8] and [9] Böttcher and Böttcher et al. go further and explore relations of bandwidth with other notions, like separability. Separability plays an important role in this chapter.

Definition 67. We say that an n-vertex graph H is γ-separable if there exists a separator set $S \subseteq V(H)$ with $|S| \leq \gamma n$ such that every component of $H-S$ has at most $o(n)$ vertices.

Böttcher et al. [9] observed that bandwidth and separability are closely related: they proved the Sublinear Equivalence Theorem. This states that, roughly speaking, sublinear bandwidth implies the existence of a sublinear sized separating set and vice versa.

One of our main results shows that when the separating set has linear (small, but not very small) size, the bandwidth can be very large even for bounded degree graphs.

Theorem 68. *Let $r \geq 35$ and $t \geq 2$ be integers and set $\gamma = \gamma(r) = 1/(8r2^r)$. Then there exists an infinite class of graphs $\mathcal{H}_{r,t}$ such that every element H of $\mathcal{H}_{r,t}$ has a separator set of size at most $\gamma v(H)$, has bandwidth at least $0.3v(H)/(2t+4)$, moreover, $\Delta(H) = \mathcal{O}(1/\gamma)$.*

It is easy to see that there are bounded degree graphs having linearly large bandwidth since it is well-known that a random l-regular graph with $l \geq 3$ has large bandwidth with positive probability. However, such random graphs do not have small separators.

Knox and Treglown [39] embedded bounded degree graphs with sublinear bandwidth into so-called robust expanders.

Definition 69. Let $0 < \nu \leq \mu < 1$. Assume that G is a graph of order n and $S \subseteq V(G)$. The *ν-robust neighborhood* $RN_{\nu,G}(S)$ of S is the set of vertices $v \in V(G)$ such that $|N(v) \cap S| \geq \nu n$. We say that G is a *robust* (ν, μ)*-expander* if $|RN_{\nu,G}(S)| \geq |S| + \nu n$ for every $S \subseteq V(G)$ such that $\mu n \leq |S| \leq (1-\mu)n$.

See also Figure 5.2.

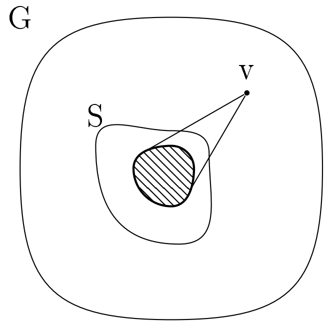

Figure 5.2: ν-robust neighborhood

We will also show that elements of $\mathcal{H}_{r,t}$ cannot be embedded into arbitrary robust expanders. However, if an n-vertex graph G has minimum degree slightly larger than $n/2$, then it contains the elements of $\mathcal{H}_{r,t}$ as spanning subgraphs. We will prove the following.

Theorem 70. *Let $r \geq 35$ and $t \geq 2$ be integers and set $\gamma = \gamma(r) = 1/(8r2^r)$. Then there exists an $n_0 = n_0(\gamma)$ such that the following holds. Assume that $n \geq n_0$ and G is an n-vertex graph having minimum degree $\delta(G) \geq (1/2 + 2\gamma^{1/3})n$. If $H \in \mathcal{H}_{r,t}$ is a graph on n vertices, then $H \subseteq G$.*

The proof of Theorem 70 will rely heavily on the proof method of [18]. Let us remark that in [18] the size of the separator set was $o(n)$, and therefore the bandwidth was also $o(n)$. This time the separator set is quite large compared to previous results. This is why the minimum degree bound for G contains γ, unlike in the main result of [18], although we still need only slightly larger $\delta(G)$, than $n/2$. The latter is the bound, for example, when we want to tile G with vertex disjoint copies of a fixed bipartite graph.

5.1 Construction of $\mathcal{H}_{r,t}$ and proof of Theorem 68

In order to exhibit the infinite family of graphs $\mathcal{H}_{r,t}$ we first need to construct certain kind of bipartite expander graphs. We begin with defining a bipartite graph F with vertex classes V_1 and V_2 such that $|V_1| = |V_2| = k$ and F has relatively good expansion properties. Our construction of F relies on the existence of so-called Ramanujan graphs.

Definition 71. An r-regular (nonbipartite) graph U is a Ramanujan graph if $\lambda \le 2\sqrt{r-1}$, where λ is the second largest in absolute value of the eigenvalues of U (since U is r-regular, the largest eigenvalue is r).

Lubotzky, Phillips and Sarnak [47], and independently Margulis [48], constructed for every $r = p + 1$ where $p \equiv 1 \mod 4$ infinite families of r-regular graphs with second largest eigenvalues at most $2\sqrt{r-1}$. We need a fact about these graphs, a lower bound for the number of edges between subsets of U.

Lemma 72. *Let U be a graph as above. Then for every two subsets $A, B \subseteq V(U)$ where $|A| = ak$ and $|B| = bk$ we have*

$$|e(A,B) - abrk| \le 2\sqrt{r-1}\sqrt{ab}k.$$

The proof of Lemma 72 can be found for example in [3].

Corollary 73. *Let U be an r-regular Ramanujan graph on k vertices with $r \ge 35$. Let us assume that $A, B \subseteq V(U)$ with $|A| = |B| = k/3$ and $A \cap B = \emptyset$. Then $e(A,B) \ge 1$.*

Proof. It is easy to see that the expression of Lemma 72 gives a lower bound for $e(A, B)$ which is monotone increasing in r. Hence it is sufficient to apply Lemma 72 with $r = 35$ and $a = b = 1/3$. Straightforward computation gives what was desired. $\square$

We are ready to discuss the details of the construction of F. Given an r-regular Ramanujan graph U with $r \geq 35$, the vertex classes of F will be copies of $V(U)$: for every $x \in V(U)$ we have two copies of it, $x_1 \in V_1$ and $x_2 \in V_2$. For every $xy \in E(U)$ we include the edges x_1y_2 and x_2y_1 in $E(F)$. Finally, for every $x \in V(U)$ we will also have the edge x_1x_2 in $E(F)$. Observe that F is an $(r+1)$-regular bipartite graph. The following claims are crucial for the construction of $\mathcal{H}_{r,t}$.

Claim 74. *Let $A \subseteq V_1$ and $B \subseteq V_2$ be arbitrary such that $|A| = |B| = k/3$. Then $e(A, B) \geq 1$.*

Proof. If there exists $x \in V(U)$ such that $x_1 \in A$ and $x_2 \in B$, then we are done, since every x_1x_2 edge is present in F. If there is no such $x \in V(U)$, then we can apply Corollary 73 and obtain what is desired. $\square$

Claim 75. *For every $A \subseteq V_1$ we have $|N(A)| \geq |A|$. Analogous statement holds for any subset $B \subseteq V_2$.*

Proof. The claim easily follows from the fact that we included a perfect matching in F when we added every x_1x_2 edge to $E(F)$. $\square$

Observe that we have a bipartite graph F with $v(F) = 2k$ whenever there exists a Ramanujan graph U with $v(U) = k$, for the latter we also assume that $r \geq 35$. Thus, there exists an infinite sequence of $\{F_i\}_{i=1}^{\infty}$ graphs on increasing number of vertices, say, F_i has $2k_i$ vertices.

We are ready to define $\mathcal{H}_{r,t}$. Each graph from this class is γ-separable where $\gamma = \gamma(r)$ can be relatively small as we will see soon. Still, the bandwidth of each of them is very large. Hence, $\mathcal{H}_{r,t}$ demonstrates that in spite of sublinear equivalence of separability and bandwidth, there is no *linear* equivalence.

The construction of $\mathcal{H}_{r,t}$ is somewhat specific, we do it with foresight as our goal is not only to further explore the relation of separability and bandwidth but also to be able to embed the elements of $\mathcal{H}_{r,t}$ later.

Definition 76. Let $n, m \in \mathbb{N}$ and set $\gamma = \gamma(r) = 1/(8r2^r)$. Let F_i be the bipartite graph as above on $2k_i$ vertices which is $(r+1)$-regular such that k_i is the largest for which $\gamma n \geq 2k_i$. The elements of $\mathcal{H}_{r,t}$ are constructed as follows. Given n we let $H = (A, B; E) \in \mathcal{H}_{r,t}$ to be the following bipartite graph.

1. $\big||A| - |B|\big| \leq 1$, and $|V| = |A \cup B| = n$,
2. let $S = S_A \cup S_B$ such that $|S_A| = |S_B| = k_i$,
3. $H[S] = F$ and $E(H[S_A]) = E(H[S_B]) = \emptyset$,
4. $D = \Delta(H) = O(r2^r)$,
5. for every point $x \in S$ we have a unique path P_x of length t starting at x and ending at z, and z has D neighbors such that each has degree 1 except one that precedes z in P_x.

See also Figure 5.3.

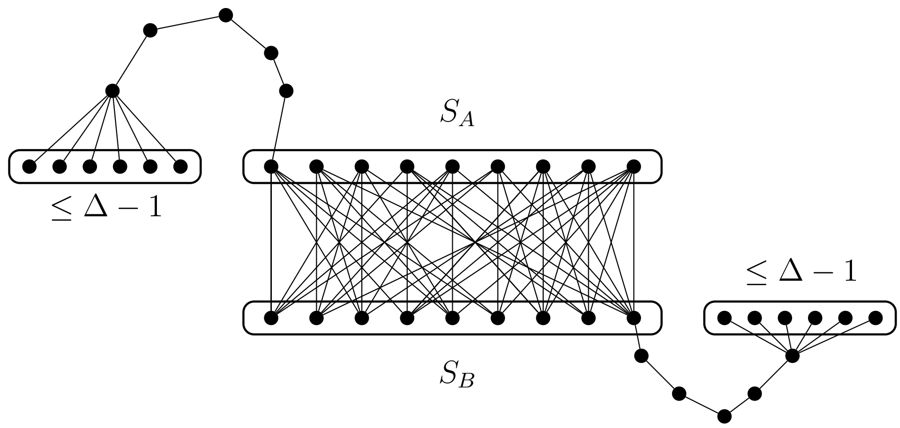

Figure 5.3: Sketch of an element of $\mathcal{H}_{r,t}$

Note that S is a separator set of H with $|S| = 2k_i \approx \gamma n$, every component of $H - S$ has less than $t + D$ vertices. From this one can easily obtain the bound $D < 3n/k_i$. The following lemma is crucial for bounding the bandwidth of $H \in \mathcal{H}_{r,t}$.

Lemma 77. *Let H be an element of $\mathcal{H}_{r,t}$ on n vertices. Assume that $X, Y \subseteq V(H)$ with $|X|, |Y| \geq 0.35n$ and $X \cap Y = \emptyset$. Then there exist an $x \in X$ and a $y \in Y$ such that the distance of x and y is at most $2t + 4$.*

Proof. Denote the vertices of $H - S$ closer to S_A by A^*, and analogously, the vertices of $H - S$ closer to S_B by B^*. By the construction of H we have $|A^*| = |B^*| = (1-\gamma)n/2$. Note that $\gamma < 0.01$ since $r \geq 35$. Hence we have that $|X - S| \geq 0.34n$ and $|Y - S| \geq 0.34n$. Thus, either $|X \cap A^*| \geq |A^*|/3$ or $|X \cap B^*| \geq |B^*|/3$. Without loss of generality, suppose the former. This also implies that at least $1/3$ of the components of A^* have vertices in X.

It is useful to introduce the notations X_A for $N_{S_A}(X)$, Y_A for $N_{S_A}(Y)$ and Y_B for $N_{S_B}(Y)$. Using these notations we have that $|X_A| \geq k/3$ and either $|Y_A| \geq k/3$, or $|Y_B| \geq k/3$.

If $|Y_B| \geq k/3$, then by Claim 74 there is an edge sq between X_A and Y_B, and therefore we have a path $xv_r \dots v_1 squ_1 \dots u_r y$ of length $2t + 3$, where $x \in X, y \in Y, v_i \in A^*$, $s \in X_A$, $q \in Y_B$, $u_i \in B^*$.

If $Y_B < \frac{k}{3}$, then $|Y_A| \geq k/3$. Let $Y_B' = N_{S_B}(Y_A)$. Claim 75 implies that $|Y_B'| \geq |Y_A| \geq k/3$, so by Claim 74 H has an edge between Y_B' and X_A. Thus, we have a path $xv_r \dots v_1 s_1 q s_2 u_1 \dots u_r y$ of length $2t + 4$, where $x \in X, y \in Y, v_i \in A^*, u_i \in B^*, s_1 \in X_A$, $s_2 \in Y_A$ and $q \in Y_B'$. □

Corollary 78. *Let H be an element of $\mathcal{H}_{r,t}$ on n vertices. Then the bandwidth of H is at least $\frac{0.3n}{2t+4}$.*

Proof. Take an arbitrary ordering $\mathcal{P}$ of the vertices of H. Let X be the first $0.35n$ vertices, while Y be the last $0.35n$ vertices of $\mathcal{P}$. Using Lemma 77 there is an $x \in X$ and an $y \in Y$ such that the distance of x and y is at most $2t + 4$. Their distance in $\mathcal{P}$ is at least $0.3n$. Thus at least one of the edges of the shortest path between x and y must have "length" at least $\frac{0.3n}{2t+4}$, from which the bound for the bandwidth follows immediately. $\square$

With this we proved Theorem 68. Note that choosing $t = 2$ results in graphs having bandwidth at least $3n/80$ while being γ-separable.

As we mentioned in the introduction Knox and Treglown [39] embedded spanning subgraphs of sublinear bandwidth into robust expanders. Recall the notion of robust expanders. The following example shows that graphs of $\mathcal{H}_{r,t}$ not only have very large bandwidth, these graphs are not necessarily subgraphs of robust expanders. Hence, in the theorem of Knox and Treglown one cannot replace small bandwidth by γ-separability, unless γ is very small.

Let us construct a robust expander. Set $\alpha = 0.002$ (we remark that we did not look for optimal constants here) and let $G = (V, E)$ be the following graph on n vertices. The vertex set of G is $V = A_0 \dot\cup A_1 \dot\cup \cdots \dot\cup A_{400}$, where $|A_i| = (1+\alpha)^i \frac{n}{1000}$ for every $0 \le i < 400$ and A_{400} contains the remainder of the vertices. The edges of G are defined as follows: $E(G)$ contains the edges $v_i v_{i+1}$ for every $v_i \in A_i$ and $v_{i+1} \in A_{i+1}$ for $0 \le i < 400$, and $G[A_{400}]$ is the complete graph on $|A_{400}|$ vertices. It is easy to see that G is a $(1/1000, 1/1000)$-robust expander. For the structure of G, see also Figure 5.4.

Lemma 79. *Let H be a graph from $\mathcal{H}_{r,t}$ on n vertices and let G be as above. Then $H \not\subseteq G$ if $t \le 47$.*

Proof. First we give an upper bound for $n - |A_{400}|$:

$$n - |A_{400}| = \left| \bigcup_{i=0}^{399} A_i \right| = \sum_{i=0}^{399} (1+\alpha)^i \frac{n}{1000} = \frac{n}{1000} \cdot \frac{1.002^{400} - 1}{0.002} < 0.62n,$$

so $|A_{400}| > 0.35n$. Let $B = \bigcup_{i=0}^{300} A_i$. We have $|B| = \frac{n}{1000} \cdot \frac{1.002^{301} - 1}{0.002} > 0.35n$.

The shortest path between B and A_{400} is of length 100. This means that H cannot be packed into G, as in H there is a path of length $2t + 4 \le 98$ between any two disjoint sets of size at least $0.35n$. $\square$

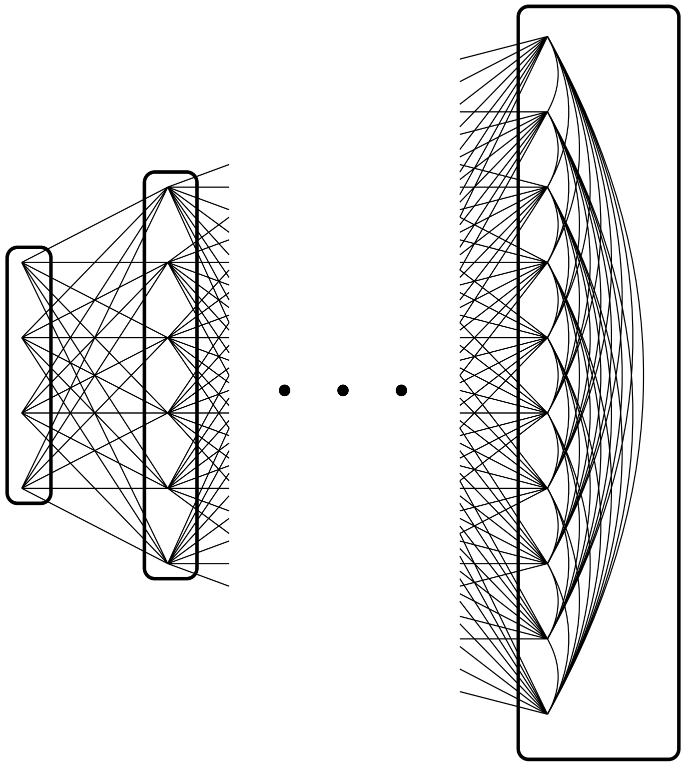

Figure 5.4: The robust expander G

5.2 Proof of Theorem 70

The proof of Theorem 70 is very similar to the proof of the main result of [18]. Hence, we will first sketch the proof of the latter one in an itemized list, and then discuss the differences we will make when proving our main result. First let us state a special case of the main theorem of [18] for embedding bipartite graphs with small separators.

Theorem 80. *For every $\varepsilon > 0$ and positive integer D there exists an $n_0 = n_0(\varepsilon, D)$ such that the following holds. Assume that H is a bipartite graph on $n \geq n_0$ vertices which has a separator set S such that $|S| = o(n)$, and every component of $H - S$ has $o(n)$ vertices. Assume further that $\Delta(H) \leq D$. Let G be an n-vertex graph such that $\delta(G) \geq (1/2 + \varepsilon)n$. Then $H \subseteq G$.*

One can observe the similarities with Theorem 70. The main difference is that in Theorem 70 the separator set can be very large compared to the separator set in the above result.

Besides the Regularity Lemma (Lemma 7) and the Blow-Up Lemma (Theorem 12), another very important tool for us is the following result by Fox and Sudakov [27].

Theorem 81. *Let H be a bipartite graph with n vertices and maximum degree $\Delta \geq 1$. If $\rho > 0$ and G is a graph with $N \geq 8\Delta\rho^{-\Delta}n$ vertices and at least $\rho\binom{N}{2}$ edges, then $H \subseteq G$.*

We are going to apply Theorem 81 in the special case $\rho = 1/2$.

5.2.1 Sketch of the proof of Theorem 80

The Regularity Lemma of Szemerédi [58] and the Blow-up Lemma [42] plays a very important role in the proof. On these important tools we gave a short overview in the Introduction. The interested reader may consult with the survey paper by Komlós and Simonovits [40] also.

Step 1: Apply the Degree Form of the Regularity Lemma with parameters $0 < \varepsilon \ll d \ll 1$ in order to obtain a partition of $V(G)$ into the clusters $W_0, W_1, \dots, W_\ell$, where W_0 is the exceptional cluster.

Step 2: Construct the reduced graph G_r on the non-exceptional clusters, in which two clusters are adjacent if and only if they form an ε-regular pair with density at least d.

Step 3: Find a maximum matching M in G_r. Using the minimum degree condition, the vertex set of M may not contain at most one cluster – its vertices are put into W_0.

Step 4: Make the edges of M super-regular. At most $2\varepsilon n$ vertices are put into W_0 at this point.

Step 5: Distribute the vertices of W_0 among the non-exceptional clusters while keeping super-regularity.

Step 6: Assign the vertices of H to clusters of G_r so that the following holds: whenever $xy \in E(H)$ for $x, y \in V(H)$, then $C(x)C(y) \in E(G_r)$, where $C(x)$, (respectively, $C(y)$) denotes the cluster to which x (respectively, y) is assigned. This is done in two steps: first randomly distribute the components of $H - S$ and S, then in the second step a few vertices may get reassigned in order to satisfy the above requirement for every edge of H.

Step 7: At this point it is possible that there are more (or less) vertices assigned to a cluster than its size. A procedure very similar to the one used in Step 5 helps in finding the balance.

Step 8: Applying the Blow-up Lemma finishes the proof.

Readers familiar with the Regularity Lemma – Blow-up Lemma method may observe that the first seven steps are essentially a preparation for being able to apply the Blow-up Lemma.

5.2.2 Proof of Theorem 70

As we indicated above, the proof of Theorem 70 is very similar to the proof of Theorem 80. Hence, below we will concentrate on the differences of the two.

Assume that $H \in \mathcal{H}_{r,t}$ has n vertices. Denote the separator set of H by S. Then we have $|S| \leq \gamma n$. Observe that we can apply the deep result of Fox and Sudakov, Theorem 81 for finding a copy of $H[S]$ in G, since $\delta(G) > n/2$. Let us denote the uncovered part of G by $\tilde{G}$ after embedding $H[S]$. Note that $\delta(\tilde{G}) \geq (1/2 + \gamma^{1/3})n$.

Next we apply the Degree Form of the Regularity Lemma for $\tilde{G}$ with parameters ε and $d = \sqrt{\gamma}$. We form the reduced graph $\tilde{G}_r$, and then find an (almost) perfect matching M in $\tilde{G}_r$. Then we make the edges of M super-regular, and then distribute the vertices of W_0 among the non-exceptional clusters so that the pairs in M remain super-regular. These steps are in fact identical to the corresponding ones (Step 1 – Step 5) in the proof of Theorem 80.

Next we assign the components of $H - S$ to the non-exceptional clusters. We do it using a random procedure (randomness is not necessary here, but a simple choice), the components are assigned randomly to edges of M. This immediately implies that whenever x, y belong to the same component and are adjacent, then $C(x)C(y)$ is an edge of M. Still, there could be vertices $x, y \in H$ such that $xy \in E(H)$, but $C(x)C(y) \notin E(\tilde{G}_r)$. This can happen only in case $x \in S$ and $y \in V(H) - S$. In such a case we repeat the procedure from the proof of Theorem 80 mentioned in Step 6.

There is an important difference of the two proofs at this point, so we provide more details here. Assume that $x \in S$ is mapped onto $v \in V(G)$ in the beginning. Let L denote those clusters in which v has at least $\sqrt{m}$ neighbors. Let $C(y)W_i$ denote the edge of M to which the component of y was assigned. Then we locate a cluster $W_j \in L$ such that W_j is adjacent to W_i in $\tilde{G}_r$. Then we reassign y to W_j. This way v will have many neighbors in the cluster of y and the cluster of y will be adjacent to the cluster of the neighbors of y in its component. Observe that if we locate the W_j clusters as evenly as possible then we can achieve that at most about $\gamma^{2/3}m$ vertices are reassigned to a particular cluster. Here we used that the set of vertices to be reassigned are neighbors of S, and there are less than n/D such vertices.

Next we repeat the procedure of Step 7. The method we use for balancing is essentially the same we discussed above. Say, that W_s has more vertices assigned to it than $|W_s|$. Then there must be a cluster W_i to which we assigned less than $|W_i|$ vertices of H. Let W_j denote the neighbor of W_s in the matching M. If W_jW_i is an edge in $\tilde{G}_r$, then we pick a vertex x such that $C(x) = W_j$ and $d(x) = D - 1$ (using the random distribution there are many choices for x). We reassign some of the leaves that are adjacent to x, the right number will be assigned to W_i.

If W_jW_i is not an edge, then there exists a cluster W_q such that W_qW_i and W_jW_p are edges in $\tilde{G}_r$, and W_pW_q is an edge in M. Then the above procedure is done in two steps: first we reassign some vertices from W_s to W_p and then from W_p to W_i. Note that the same computation works as above: at most $\gamma^{2/3}m$ vertices are reassigned at every cluster.

Since the density of the ε-regular pairs is at least $\sqrt{\gamma}$, and at most $\gamma^{2/3}m$ vertices are reassigned at every cluster, we are able to apply the Blow-up Lemma. This finishes the proof of Theorem 70.

Summary

This thesis consists of four main parts. In Chapter 2 we deal with a lower estimation on the size of simple suffix trees. First, we present a simple algorithm for constructing the suffix tree of a string. Then, with the help of aperiodic strings and with some simple combinatorial reasoning, we give a quadratic lower bound on the size of the suffix tree of a random string. The main results are formulated in Theorem 23, Theorem 24 and Theorem 25.

In Chapter 3 we consider bipartite graph packing and embedding problems. In Theorem 33 we give almost tight conditions on embedding of a bipartite graph into another, where the former has bounded degrees, while the latter has large degrees. For the proof, we use a well-known result of Lovász, Lemma 36, and with its help we prove the key technical lemma, and we also use martingales.

In Chapter 4 we consider another embedding question, that of degree sequences and graphs. In Theorem 40 we prove that a degree sequence π bounded with a constant from above can be embedded into a graph G with sufficiently large degrees. For achieving the proof, we find a suitable realization of π, then we embed it into G. Through the proof, we use the Erdős-Stone theorem formulated in Theorem 44 [14] and a result on well-separable graphs, formulated in Theorem 46 [18], which is based on the Regularity Lemma of Szemerédi.

In Theorem 42 we show that a q-unbalanced degree sequence π can be embedded into a graph G with sufficiently large degrees and on slightly more vertices than the length of π. First, we find again a suitable realization of π, i.e. a union

of smaller q-unbalanced bipartite graphs, then with the help of the Regularity Lemma we find a a special substructure in G, and finally, we finish the embedding. In finding H, which is done in Subsection 4.2.1 we use a result from number theory, which is formulated in Proposition 48 [56]. Using Lemma 49, which is a result of Lovász [46], we prove Lemma 50, which implies Lemma 47. The decomposition of G is done in Subsection 4.2.2, in Lemma 51. In Subsection 4.2.3 we prepare the graph for embedding by distributing the exceptional class of the reduced graph. Finally, we describe the embedding algorithm in Subsection 4.2.4.

In Chapter 5, in Theorem 68 we show the existence of bounded degree bipartite graphs with a small separator and large bandwidth. For the construction we use Ramanujan graphs. Then in Theorem 70 we prove that under certain conditions these graphs can be embedded into graphs with minimum degree only slightly over $n/2$. This proof is similar to that of the main result of [18], thus we give a sketch for the latter, and note the differences.

Összefoglalás

Ez a disszertáció négy fő részből áll. A 2. fejezetben alsó becslést adunk az egyszerű szuffixfák méretére. Először bemutatunk egy egyszerű algoritmust, mellyel előállíthatjuk egy szöveg szuffixfáját. Ezután az aperiodikus szövegek segítségével és alapvető kombinatorikai gondolatokkal négyzetes alsó korlátot adunk egy véletlen szöveg szuffixfájának méretére. E rész fő eredményeit a 23., a 24. és a 25. tételekben mondjuk ki.

A 3. fejezetben a páros gráfpakolást illetve beágyazási problémát vizsgáljuk. A 33. tételben majdnem pontos feltételeket adunk egy korlátos fokú páros gráf nagyfokú páros gráfba való beágyazására. A bizonyítás során használjuk Lovász jól ismert eredményét, a 36. lemmát. Ennek segítségével bizonyítjuk fő technikai lemmánkat. A bizonyítás során martingálokat is felhasználunk.

A 4. fejezetben egy újabb beágyazási problémával foglalkozunk, mégpedig a fokszámsorozatok gráfokba való beágyazásával. A 40. tételben megmutatjuk, hogy egy konstans fokkal korlátozott π fokszámsorozat beágyazható egy elég nagy fokú G gráfba. Ehhez megkeressük a π egy alkalmas megvalósítását, amit beágyazunk G-be. A bizonyítás során használjuk az Erdős–Stone-tételt, amit a 44. tételben mondunk ki [14], valamint a 46. tétel eredményét a jól-szeparálható gráfokról [18]. Ebben a bizonyításban Szemerédi Regularitási Lemmáját is használjuk.

A 42. tételben bebizonyítjuk, hogy egy q-kiegyensúlyozatlan π fokszámsorozat beágyazható egy elég nagy fokú gráfba, amelynek kicsit több csúcsa van, mint π hossza. Először megint egy alkalmas H realizációt keresünk π-hez, több kisebb q-

kiegyensúlyozatlan páros gráfból, majd a Regularitási Lemma segítségével találunk G-ben egy speciális struktúrát, végül befejezzük a beágyazást. A H gráf megtalálásában, amely a 4.2.1. szakaszban történik, egy számelméleti eredmény, a 48. lemma [56] segítségével és Lovász eredményével, a 49. lemmával [46] megmutatjuk az 50. lemmát, amelyből következik a 47. lemma. G felbontását a 4.2.2. szakaszban mutatjuk meg, az 51. lemmában. A 4.2.3. szakaszban a redukált gráf kivételes osztályának szétosztásával előkészítjük a G gráfot a beágyazásra, végül a 4.2.4. szakaszban befejezzük a beágyazást.

Az 5. fejezetben, a 68. tételben megmutatjuk, hogy léteznek olyan korlátos fokú páros gráfok, melyeknek szeparáló halmaza kicsi, miközben sávszélessége nagy. Ezek megkonstruálásához Ramanujan-gráfokat használunk. A 70. tételben megmutatjuk, hogy bizonyos feltételek mellett ezeket a gráfokat be lehet ágyazni olyan gráfokba, melyek minimális foka csak kicsit nagyobb $n/2$-nél. Ez a bizonyítás hasonló Csaba [18] fő eredményének bizonyításához, így az utóbbit vázoljuk, miközben kiemeljük az eltéréseket.

Acknowledgement

I would like to express my gratitude and my appreciation to my supervisor, Dr. Béla Csaba for the support of my PhD studies, for his help, guidance and immense knowledge. My sincere thanks also goes to Dr. Péter Hajnal, for all the encouragement and motivation with my studies, especially around the final exams and the dissertation. I appreciate the help of the Bolyai Institute, in providing the necessary conditions for finishing this dissertation.

Last but not the least, I would like to thank my family: to my wife, Kati, who thoroughly proofread my dissertation; to my parents and to my brothers and sisters for supporting me spiritually throughout writing this thesis and my life in general.

In all my works, I would like to glorify the Triune God, to whom be glory for ever and ever. Soli Deo Gloria.

Bibliography

[1] E.F. Adebiyi, T. Jiang, and M. Kaufmann. An efficient algorithm for finding short approximate non-tandem repeats. *Bioinformatics*, 17(Suppl 1):5S–12S, 2001.

[2] M. Aigner and S. Brandt. Embedding arbitrary graphs of maximum degree two. *Journal of the London Mathematical Society*, s2-48(1):39–51, 1993.

[3] N. Alon and J. H. Spencer. *The Probabilistic Method.* John Wiley and Sons, Inc., second edition, 2008.

[4] A. Apostolico, M. Crochemore, M. Farach-Colton, Z. Galil, and S. Muthukrishnan. 40 years of suffix trees. *Communications of the ACM*, 59(4):66–73, 2016.

[5] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

[6] A. Blumer, A. Ehrenfeucht, and D. Haussler. Average sizes of suffix trees and DAWGs. *Discrete Applied Mathematics*, 24(1-3):37–45, 1989.

[7] B. Bollobás and S.E. Eldridge. Packing of graphs and applications to computational complexity. *Journal of Combinatorial Theory, Series B*, 25(2):105–124, 1978.

[8] J. Böttcher. *Embedding large graphs: The Bollobás-Komlós conjecture and beyond.* PhD thesis, Technische Universität München, 2009.

[9] J. Böttcher, K.P. Pruessmann, A. Taraz, and A. Würfl. Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs. *European Journal of Combinatorics*, 31(5):1217–1227, 2010.

[10] J. Böttcher, M. Schacht, and A. Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen*, 343(1):175–205, 2009.

[11] P.A. Catlin. Subgraphs of graphs. *Discrete Mathematics*, 10(2):225–233, 1974.

[12] P.A. Catlin. *Embedding subgraphs and coloring graphs under extremal degree conditions*. PhD thesis, Ohio State University, Columbus, 1976.

[13] M.T. Chen and J. Sciferas. Efficient and elegant subword tree construction. In *Combinatorial algorithms on words*, pages 97–107. Springer-Verlag, 1985.

[14] V. Chvátal and E. Szemerédi. On the Erdős–Stone Theorem. *Journal of the London Mathematical Society*, s2-23(2):207–214, 1981.

[15] J.D. Cook. Counting primitve bit strings. `http://www.johndcook.com/blog/2014/12/23/counting-primitive-bit-strings/`, 2014. [Online; accessed 05-May-2018].

[16] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Mathematica Academiae Scientiarum Hungaricae*, 14(3-4):423–439, 1963.

[17] B. Csaba. On the Bollobás-Eldridge conjecture for bipartite graphs. *Combinatorics, Probability and Computing*, 16(5):661–691, 2007.

[18] B. Csaba. On embedding well-separable graphs. *Discrete Mathematics*, 308(19):4322–4331, 2008.

[19] B. Csaba, A. Shokoufandeh, and E. Szemerédi. Proof of a conjecture of Bollobás and Eldridge for graphs of maximum degree three. *Combinatorica*, 23(1):35–72, 2003.

[20] B. Csaba and B.M. Vásárhelyi. On the relation of separability and bandwidth. Submitted for publication, 2018.

[21] B. Csaba and B.M. Vásárhelyi. On embedding degree sequences. In *Proceedings of the Middle European Conference on Applied Theoretical Computer Science (MATCOS 2016)*, pages 68–71, 2016.

[22] B. Csaba and B.M. Vásárhelyi. On embedding degree sequences. *Informatica (SI)*, Accepted for publication, 2018+.

[23] L. Devroye, W. Szpankowski, and B. Rais. A note on the height of suffix trees. *SIAM Journal on Computing*, 21(1):48–53, 1992.

[24] J. Diemunsch, M.J. Ferrara, S. Jahanbekam, and J.M. Shook. Extremal theorems for degree sequence packing and the 2-color discrete tomography problem. *SIAM Journal of Discrete Mathematics*, 29(4):2088–2099, 2015.

[25] G.A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, s3-2(1):69–81, 1952.

[26] E. R. Fiala and D. H. Greene. Data compression with finite windows. *Communications of the ACM*, 32(4):490–505, 1989.

[27] J. Fox and B. Sudakov. Density theorems for bipartite graphs and related Ramsey-type results. *Combinatorica*, 29(2):153–196, 2009.

[28] C. Fraser, A. Wendt, and E.W. Myers. Analyzing and compressing assembly code. In *Proceedings SIGPLAN Symposium on Compiler Construction*, pages 117–121, 1984.

[29] D. Gale. A theorem on flows in networks. *Pacific Journal of Mathematics*, 7(2):1073–1082, 1957.

[30] E.N. Gilbert and J. Riordan. Symmetry types of periodic sequences. *Illinois Journal of Mathematics*, 5(4):657–665, 1961.

[31] R. Grossi and G.F. Italiano. Suffix trees and their applications in string algorithms. In *Proceedings of the 1st South American Workshop on String Processing*, pages 57–76, 1993.

[32] D. Gusfield. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, 1997.

[33] P. Hajnal. An $\Omega(n^{4/3})$ lower bound on the randomized complexity of graph properties. *Combinatorica*, 11(2):131–143, 1991.

[34] P. Hajnal and M. Szegedy. On packing bipartite graphs. *Combinatorica*, 12(3):295–301, 1992.

[35] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[36] M. Höhl, S. Kurtz, and E. Ohlebusch. Efficient multiple genome alignment. *Bioinformatics*, 18(Suppl 1):312S–320S, 2002.

[37] L. Kaderali and A. Schliep. Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics*, 18(10):1340–1348, 2002.

[38] H. A. Kierstead, A. V. Kostochka, and G. Yu. Extremal graph packing problems: Ore-type versus Dirac-type. In S. Huczynska, J.D. Mitchell, and C.M. Roney-Dougal, editors, *Surveys in Combinatorics, Proc. 22nd British Combinatorial Conference*, number 365 in London Mathematical Society Lecture Note Series, pages 113–135. Cambridge University Press, Cambridge, 2009.

[39] F. Knox and A. Treglown. Embedding spanning bipartite graphs of small bandwidth. *Combinatorics, Probability and Computing*, 22(1):71–96, 2013.

[40] J. Komlós, G. Sárközy, and E. Szemerédi. An algorithmic version of the blow-up lemma. *Random Structures and Algorithms*, 12(3):297–312, 1998.

[41] J. Komlós, G. N. Sárközy, and E. Szemerédi. Spanning trees in dense graphs. *Combinatorics, Probability and Computing*, 10(5):397–416, 2001.

[42] J. Komlós, G.N. Sárközy, and E.Szemerédi. Blow-up lemma. *Combinatorica*, 17(1):109–123, 1997.

[43] J. Komlós, G.N. Sárközy, and E. Szemerédi. On the Pósa-Seymour conjecture. *Journal of Graph Theory*, 29(3):167–176, 1998.

[44] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. In G. B. Khosrovshahi, A. Shokoufandeh, and A. Shokrollahi, editors, *Theoretical Aspects of Computer Science. Lecture Notes in Computer Science, vol. 2292*, pages 84–112. Springer, Berlin, Heidelberg, 2002.

[45] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. *Combinatorics: Paul Erdős is eighty*, 2:295–352, 1993.

[46] L. Lovász. *Combinatorial Problems and Exercises.* American Mathematical Society, second edition, 2007.

[47] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[48] G.A. Margulis. Explicit constructions of graphs without short cycles and low density codes. *Combinatorica*, 2(1):71–78, 1982.

[49] R.R. Martin. The edit distance in graphs: Methods, results, and generalizations. In A. Beveridge, J.R. Griggs, L. Hogben, G. Musiker, and P. Tetali, editors, *Recent Trends in Combinatorics*, pages 31–62. Springer International Publishing, Cham, 2016.

[50] J. Matoušek and J. Vondrák. The probabilistic method. Lecture notes, online: `https://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f09/www/handouts/matousek-vondrak-prob-ln.pdf`, last accessed: 05-May-2018.

[51] E. M. McCreight. A space-economical suffix tree construction algorithm. *Journal of the ACM*, 23(2):262–272, 1976.

[52] L. O'Connor and T. Snider. Suffix trees and string complexity. In *Advances in Cryptology: Proceedings of EUROCRYPT, LNCS 658*, pages 138–152. Springer-Verlag, 1992.

[53] C.H. Papadimitriou. The NP-completeness of the bandwidth minimization problem. *Computing*, 16(3):263–270, 1976.

[54] M. Rodeh. A fast test for unique decipherability based on suffix trees. *IEEE Transactions on Information Theory*, 28(4):648–651, 1982.

[55] H.J. Ryser. Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics*, 9:371–377, 1957.

[56] M.L. Sahs, P.A. Sissokho, and J.N. Torf. A zero-sum theorem over $\mathbb{Z}$. *Integers*, 13:#A70, 2013.

[57] N. Sauer and J. Spencer. Edge disjoint placement of graphs. *Journal of Combinatorial Theory*, 25(3):295–302, 1978.

[58] E. Szemerédi. Regular partitions of graph. *Colloques Internationaux C.N.R.S - Problèmes Combinatoires et Théorie des Graphes*, 260:399–401, 1976.

[59] S.L. Tanimoto. A method for detecting structure in polygons. *Pattern Recognition*, 13(6):389–494, 1981.

[60] E. Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249–260, 1995.

[61] B.M. Vásárhelyi. An estimation of the size of non-compact suffix trees. *Acta Cybernetica*, 22(4):823–832, 2016.

[62] B.M. Vásárhelyi. On the bipartite graph packing problem. *Discrete Applied Mathematics*, 227:149–155, 2017.

[63] H. Wang. Packing two bipartite graphs into a complete bipartite graph. *Journal of Graph Theory*, 26(2):95–104, 1997.

[64] P. Weiner. Linear pattern matching algorithms. In *Proceedings of the 14th IEEE Symposium on Switching and Automata Theory*, pages 1–11, 1973.

[65] D. B. West. *Introduction to Graph Theory*. Pearson Education Inc., Delhi, India, second edition, 2002.

[66] A. P. Wojda and P. Vaderlind. Packing bipartite graphs. *Discrete Mathematics*, 164(1-3):303–311, 1997.

[67] Q. R. Yu and G. Z. Liu. *Graph factors and matching extensions.* Higher Education Press, Beijing, China, 2009.

The publications of the author

Publications related to the thesis:

1. Béla Csaba, **Bálint Márk Vásárhelyi**, *On embedding degree sequences*, Informatica (SI), 2018+
2. **Bálint Márk Vásárhelyi**, *On the bipartite graph packing problem*, Discrete Applied Mathematics, 227:149–155, 2017.
3. Béla Csaba, **Bálint Márk Vásárhelyi**, *On embedding degree sequences*, Middle-European Conference on Applied Theoretical Computer Science (MATCOS 2016), Proceedings, p.68–71, 2016.
4. **Bálint Márk Vásárhelyi**, *An estimation of the size of non-compact suffix trees*, Acta Cybernetica, 22(4):823–832, 2016.
5. Béla Csaba, **Bálint Márk Vásárhelyi**,*On the relation of separability and bandwidth*, Submitted for publication, 2018.

Other publications:

6. Ákos Nyerges, Bálint Csörgő, Gábor Draskovits, Bálint Kintses, Petra Szili, Györgyi Ferenc, Tamás Révész, Eszter Ari, István Nagy, Balázs Bálint, **Bálint Márk Vásárhelyi**, Péter Bihari, Mónika Számel, Dávid Balogh, Henrietta Papp, Dorottya Kalapis, Balázs Papp, Csaba Pál, *Predicting the*

evolution of antibiotic resistance by directed mutagenesis at multiple loci, Proceedings of the National Academy of Sciences of the United States of America, 2018+

7. Bálint Kintses, Orsolya Méhi, Eszter Ari, Mónika Számel, Ádám Györkei, Pramod K. Jangir, István Nagy, Ferenc Pál, Gergő Fekete, Roland Tengölics, Ákos Nyerges, István Likó, Balázs Bálint, **Bálint Márk Vásárhelyi**, Balázs Papp, Csaba Pál, *Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiome*, Submitted for publication, 2018.

8. Marianna Nagymihály, **Bálint Márk Vásárhelyi**, Quentin Barrière, Teik-Min Chong, Balázs Bálint, Péter Bihari, Kar-Wai Hong, Balázs Horváth, Jamal Ibijbijen, Mohammed Amar, Attila Farkas, Éva Kondorosi, Kok-Gan Chan, Véronique Gruber, Pascal Ratet, Peter Mergaert, Attila Kereszt, *The complete genome sequence of Ensifer (Sinorhizobium) meliloti strain CCMM B554 (FSM-MA), a highly effective nitrogen-fixing microsymbiont of Medicago truncatula*, Standards in Genomic Sciences, 12:75, 2017.

9. Domonkos Sváb, Balázs Bálint, **Bálint Márk Vásárhelyi**, Gergely Maróti, István Tóth, *Comparative genomic and phylogenetic analysis of a Shiga toxin producing Shigella sonnei (STSS) strain*, Frontiers in Cellular and Infection Microbiology, 7:229, 2017.

10. **Bálint Márk Vásárhelyi**, Book review for L. S. Pitsoulis, Topics in Matroid Theory, xiv+127 pages, Springer, New York, 2014., Acta Mathematica, 80(3-4):702, 2014.