Trees and graph packing

Bálint Márk Vásárhelyi

Outline of Ph.D. Thesis

Supervisor:

Dr. Béla Csaba

Docent

Doctoral School of Mathematics and Computer Science

University of Szeged

Faculty of Science and Informatics

Bolyai Institute

Szeged, 2018

1 Introduction

In this thesis we investigate two main topics, namely, suffix trees and graph packing problems. In Chapter 2, we present the suffix trees. The main result of this chapter is a lower bound on the size of simple suffix trees. In the rest of the thesis we deal with packing problems. In Chapter 3 we give almost tight conditions on a bipartite packing problem. In Chapter 4 we consider an embedding problem regarding degree sequences. In Chapter 5 we show the existence of bounded degree bipartite graphs with a small separator and large bandwidth and we prove that under certain conditions these graphs can be embedded into graphs with minimum degree slightly over $n/2$.

The thesis is mainly based on [3], [4], [5], [11], and [12]. Numbers between $\langle$ and $\rangle$ show the numbering in the dissertation.

2 Suffix trees

Let S be a string of length n. The ith character of S is $S[i]$. The substring of S from $S[i]$ to $S[j]$ is $S[i, j]$ $(j \geq i)$.

The *suffix tree* of S is a rooted directed tree with n leaves. Each edge has a label, and the edges from a node have different labels. By concatenating the edge labels along a path from the root to a leaf j, we obtain the suffix $S[j, n]$ with a \$ sign at the end, which is not contained in the alphabet.

We present a simple algorithm for constructing the suffix tree, by considering the suffices one by one, starting with $S[1, n]$.

The *growth of* S denoted by $\gamma(S)$ is one less than the shortest distance of leaf 1 from an internal node having at least two children. Let $\Omega(n, k, \sigma)$ be the number of strings of length n with growth k over an alphabet of size σ.

We can observe 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 suffix tree.

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

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

Theorem 1. *(23) For any $k \in \mathbb{N}$, on any alphabet of size σ for all $n \geq 2k$,*

$$\Omega(n, k, \sigma) \leq \varphi(k, \sigma) \tag{1}$$

for some function φ.

Theorem 2. *(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 3. *(25) There is a $d > 0$ that for any $n > n_0$ (where n_0 is the same as in Theorem 2) 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.*

Theorem 1 implies Theorem 2, which in turn implies Theorem 3.

For proving Theorem 1, we need a few lemmas on aperiodic strings.

Lemma 4. *(26) For all $j > 0$ integer and for all alphabets of size σ the number of aperiodic strings is*

$$\mu(j, \sigma) = \sigma^j - \sum_{\substack{d|j \\ d \neq j}} \mu(d, \sigma). \tag{2}$$

This implies the following corollary:

Corollary 5. *(27) If p is prime and $t \in \mathbb{N}$, then*

$$\mu\left(p^t, \sigma\right) = \sigma^{p^t} - \sigma^{p^{t-1}}. \tag{3}$$

for all alphabets of size σ.

An upper and a lower bound on the number of aperiodic strings are provided by the following lemmas, which we use in the proof of Theorem 1.

Lemma 6. *(28) For all $j > 1$ and for all alphabets of size σ,*

$$\mu(j, \sigma) \leq \sigma^j - \sigma. \tag{4}$$

Lemma 7. *(29) For all $j \geq 1$, and for all alphabets of size σ*

$$\mu(j, \sigma) \geq \sigma(\sigma - 1)^{j-1}. \tag{5}$$

3 Bipartite packing problem

Let H and G be two graphs on n vertices. We say that H is embeddable into G, if H is a subgraph of G, i.e. there is a copy of H in G. We denote this with $H \subseteq G$.

H can be packed with G, if H is embeddable into $\overline{G}$. An equivalent definition is that edge-disjoint copies of H and G can be found in the complete graph K_n.

Note that the embedding and packing are complementer problems.

The main result of this chapter is the following theorem.

Theorem 8. *(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 $y \in T$.*

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 $\frac{cn}{\log n}$. The graph H cannot be embedded into G, which follows from the example of Komlós et al. [8].

An important lemma which we use here is that of Gale and Ryser, in the formulation of [9].

Lemma 9. *(36) [6, 10] Let $G(A, B)$ be a bipartite graph and π a bigraphic sequence on (A, B). If for all $X \subseteq A$, $Y \subseteq B$*

$$\sum_{x \in X} \pi(x) \leq e_G(X, Y) + \sum_{y \in \overline{Y}} \pi(y), \tag{6}$$

then π can be embedded into G.

For proving Theorem 8 we first show the following crucial lemma.

Lemma 10. *(37) Let* $\varepsilon \in (0, 0.5)$ *and* c *as stated in Theorem 8. 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*

$$M \leq deg_H(x) \leq M(1 + \delta) \text{ for all } x \in Z', \tag{7}$$

and

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

Then there is an embedding of H *into* G*.*

Using martingales we finally prove Theorem 8, for which we also need the Azuma-Hoeffding Inequality.

Lemma 11. *(38) [1] If* $\mathcal{Z}$ *is a martingale with martingale differences at most* 1*, then for any* j *and* t *the following holds:*

$$\mathbb{P}\left(\mathcal{Z}_j \geq \mathbb{E}\mathcal{Z}_j - t\right) \geq 1 - e^{-\frac{t^2}{2j}}. \tag{8}$$

4 Embedding degree sequences

In this chapter we consider embedding questions of degree sequences and graphs. These problems are related to the previous results. The first theorem which we proved is the following.

Theorem 12. *⟨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 12 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.

Theorem 12 is proved by finding a suitable realization H of π, then embedding H into G. In the proof, we use the Regularity Lemma of Szemerédi.

The other main result uses the following definition.

Definition 13. ⟨41⟩ Let $q \geq 1$ be an integer. A bipartite graph H with vertex classes S and T is q-unbalanced, if $q|S| \leq |T|$. The degree sequence π is q-unbalanced, if it can be realized by a q-unbalanced bipartite graph.

Theorem 14. *(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 $n \geq n_0$, π is a q-unbalanced degree sequence of length $n - M$ 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 π can be embedded into G.*

The proof of Theorem 14 consists of first showing the existence of an appropriate q-unbalanced bipartite graph H realizing π, then with the help of the Regularity Lemma and the Blow-up Lemma we embed H into G.

For finding the proper H graph, we prove the following lemma.

Lemma 15. *(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, $v(H_i) \leq 4D^2$.*

A consequence of the Gale-Ryser Lemma (Lemma 9) is given as it follows.

Lemma 16. *(49) [6, 10] Let $G = (A, B; E(G))$ be a bipartite graph and f be a nonnegative 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*

$$f(X) \leq e(X, Y) + f(\overline{Y}) \tag{9}$$

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

With its help, we prove the following lemma:

Lemma 17. *(50) If $f = (a_1, \ldots, a_s; b_1, \ldots, b_t)$ 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 $A = \{a_1, \ldots, a_s\}$ and $B = \{b_1, \ldots, b_t\}$ then f is bigraphic.*

Applying Lemma 17, we show Lemma 15.

After we found H, with the help of the Regularity Lemma, we decompose the graph G, and finally, we embed H into G.

A corollary of Theorem 14 is formulated in the following theorem.

Theorem 18. *(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'.*

5 On the relation of separability and bandwidth

In this chapter, we consider another embedding question, extending the result of [2].

First, we define the bandwidth of a graph.

Definition 19. ⟨65⟩ Let $H(V,E)$ be a graph. Let $\mathcal{F} = \{f : V \to \{1,\ldots,n\}\}$ be a family of bijective functions on V. The bandwidth of H is

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

Definition 20. ⟨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.

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 21. *⟨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, $\Delta(H) = \mathcal{O}(1/\gamma)$.*

The construction for Theorem 21 is based on Ramanujan graphs. First, we create an $r+1$-regular bipartite graph 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 $E(F)$.

Two claims help us to finish the construction.

Claim 22. *⟨74⟩ Let $A \subseteq V_1$ and $B \subseteq V_2$ be arbitrary such that $|A| = |B| = k/3$. Then $e(A, B) \geq 1$.*

Claim 23. *⟨75⟩ For every $A \subseteq V_1$ we have $|N(A)| \geq |A|$. Analogous statement holds for any subset $B \subseteq V_2$.*

Now, we turn to the construction.

Definition 24. ⟨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.

In the following, we prove a few properties of the family $\mathcal{H}_{r,t}$.

Lemma 25. *⟨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 $2t + 4$.*

Corollary 26. *(78) Let H be an element of $\mathcal{H}_{r,t}$ on n vertices. Then the bandwidth of H is at least $\frac{0.3n}{2t+4}$.*

Definition 27. ⟨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 [7].

Let us construct a robust expander G on n vertices. Let $V = A_0 \dot{\cup} A_1 \dot{\cup} \cdots \dot{\cup} A_{400}$, where $|A_i| = (1+\alpha)^i \frac{n}{1000}$ for every $0 \leq i < 400$ 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 $0 \leq i < 400$, 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.

Lemma 28. *(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 $t \leq 47$.*

The other main result of this chapter is that such graphs can be embedded into graphs with minimum degree slightly over $n/2$.

Theorem 29. *(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 29 is very similar to that of the main result of [2]. There are but a few differences between the two. First, we apply the Regularity Lemma to G, then construct the reduced graph G_r, and find a maximum matching M in it. We make M super-regular, and distribute the vertices of the exceptional cluster while maintaining super-regularity. Finally, we assign the vertices H to the clusters of G_r, and we finish by applying the Blow-up Lemma.

References

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

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

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

[4] 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.

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

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

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

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

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

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

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

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

The dissertation is based on the following papers of the author:

[3], [4], [5], [11] and [12].