A better approximation for longest noncrossing spanning trees*

Sergio Cabello[1], Aruni Choudhary[†2], Michael Hoffmann[‡3], Katharina Klost[2], Meghana M. Reddy[‡3], Wolfgang Mulzer[2], Felix Schröder[4], and Josef Tkadlec[5]

1 **University of Ljubljana, Slovenia**
`sergio.cabello@fmf.uni-lj.si`
2 **Institut für Informatik, Freie Universität Berlin, Germany**
`{arunich,kathklost,mulzer}@inf.fu-berlin.de`
3 **Department of Computer Science, ETH Zürich, Switzerland**
`{hoffmann,meghana.mreddy}@inf.ethz.ch`
4 **Technische Universität Berlin, Germany**
`fschroed@math.tu-berlin.de`
5 **Institute of Science and Technology, Austria**
`jtkadlec@ist.ac.at`

Abstract

Let P be a finite set of points in the plane. For any spanning tree T on P, we denote by $|T|$ the Euclidean length of T. Let T_{OPT} be a noncrossing spanning tree of maximum length for P. We show how to construct a noncrossing spanning tree T_{ALG} with $|T_{\text{ALG}}| \geq \delta \cdot |T_{\text{OPT}}|$ with $\delta = 0.512$. We also show how to improve this bound when the points lie in a thin rectangle.

1 Introduction

In this paper we address the problem of finding a longest noncrossing spanning tree. The closely related problems of finding both a shortest (noncrossing) and a longest (possibly crossing) spanning tree are computationally easy. The minimization version is simply the classical minimum spanning tree problem, and the noncrossing property follows from the triangle inequality. Similarly, the longest spanning tree can be computed in a greedy fashion. In contrast, finding the longest noncrossing spanning tree is conjectured to be NP-hard [1].

As obtaining an efficient exact algorithm seems to be difficult, we focus on polynomial-time approximation algorithms for the longest noncrossing spanning tree. One of the first results is due to Alon et al. [1] who gave an 0.5-approximation. Dumitrescu and Tóth [3] refined this algorithm and achieved an approximation factor of 0.502. In their analysis, they compare the output of their algorithm to a longest, possibly crossing, spanning tree. With a modification of this algorithm, Biniaz et al. improved this factor slightly to 0.503 [2]. They also compare their result to the longest crossing spanning tree. While such a tree provides a safe upper bound, it is not a valid solution for the problem and may be up to $\pi/2 > 1.5$ times longer than a longest noncrossing spanning tree [1].

* This research was started at the 3rd DACH Workshop on Arrangements and Drawings, August 19–23, 2019, in Wergenstein (GR), Switzerland, and continued at the 16th European Research Week on Geometric Graphs, November 18–22, 2019, in Strobl, Austria. We thank all the participants of the workshop for valuable discussions and for creating a conducive research atmosphere.
† Supported in part by ERC StG 757609.
‡ Supported by the Swiss National Science Foundation within the collaborative DACH project *Arrangements and Drawings* as SNSF Project 200021E-171681.

36th European Workshop on Computational Geometry, Würzburg, Germany, March 16–18, 2020.
This is an extended abstract of a presentation given at EuroCG'20. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

In this paper, we aim to design a better approximation algorithm by making use of the noncrossing property. In this way, we obtain a significant improvement on the approximation factor to 0.512. Our algorithm uses similar ideas and constructions as the previous algorithms.

Moreover, we can show an even better approximation for "thin" point sets. In particular, we show that when the point set lies in a thin rectangular strip, then there is always a noncrossing spanning tree of length at least 2/3 the length of the longest (possibly crossing) spanning tree, and that this bound is tight.

2 Preliminaries

Let $P \subset \mathbb{R}^2$ be the given point set. Without loss of generality we assume that $\operatorname{diam}(P) = 1$. Similar to the existing algorithms [1, 2, 3], we make extensive use of stars. The *star* S_p rooted at some point $p \in P$ is the tree that connects p to all other points of P (see Figure 1).

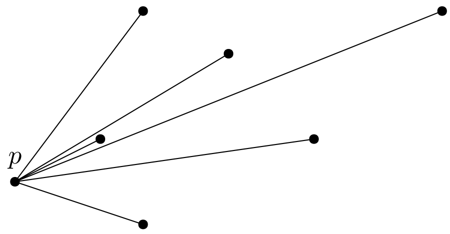

Figure 1 A star S_p.

The following slight generalization of Lemma 3 in Dumitrescu and Tóth [3] will be very useful throughout the paper.

▶ **Lemma 2.1.** *Let $p, q \in P$. Then* $\max\{|S_p|, |S_q|\} \geq \frac{n}{2}\|pq\|$.

Proof. First we note that $\max\{|S_p|, |S_q|\} \geq \frac{1}{2}(|S_p| + |S_q|)$. The triangle inequality yields:

$$|S_p| + |S_q| = \sum_{r\in P} \|pr\| + \|rq\| \geq \sum_{r\in P} \|pq\| = n \cdot \|pq\|. \qquad ◀$$

▶ **Observation 2.2.** *Let ab be a longest edge of T_{OPT}. As $\|ab\| \leq 1$ by assumption, we have*

$$|T_{\text{OPT}}| \leq \|ab\|(n-1) < \|ab\|n \leq n.$$

3 The 0.512-approximation

We show how to compute a spanning tree T_{ALG} with $|T_{\text{ALG}}| \geq \delta \cdot |T_{\text{OPT}}| = 0.512 \cdot |T_{\text{OPT}}|$. Our approach is the following: we guess a longest edge ab of T_{OPT}. If $\|ab\| < d := \frac{1}{2\delta}$ then it is straightforward to give a good approximation, as shown below in Lemma 3.1. Otherwise, we describe six different noncrossing spanning trees for the set P and show that at least one of them gives an approximation ratio of at least δ.

We use the noncrossing property of the optimal tree T_{OPT} in Lemma 3.2, which also is the bottleneck case in our construction.

From now on, we assume that ab is a longest edge in T_{OPT} and that p, q is a pair of vertices that realizes the diameter, that is, $\|pq\| = 1 \geq \|ab\|$.

▶ **Lemma 3.1.** *Let T_{DIAM} the longer of S_p and S_q. If $\|ab\| < d$, then* $|T_{\text{DIAM}}| \geq \delta \cdot |T_{\text{OPT}}|$.

Proof. From Lemma 2.1 it follows that $\max\{|S_p|, |S_q|\} \geq \frac{n}{2}$. As we observed above, we have $|T_{\text{OPT}}| \leq \|ab\|n < dn$. Thus, we get an approximation ratio of

$$\frac{|T_{\text{DIAM}}|}{|T_{\text{OPT}}|} \geq \frac{n/2}{dn} = \frac{1}{2d} = \delta. \qquad ◀$$

Now we only consider the case where $\|ab\| \geq d$. Additionally for ease of presentation, we will assume that $a = (0, 0)$ and $b = (\|ab\|, 0)$ without loss of generality.

First, we define $F = D(a, 1) \cap D(b, 1)$ to be the region with distance at most 1 from a and b. Since the diameter of the point set is 1, we can be sure that $P \subset F$. Let $\hat{\alpha}$ be a constant to be determined later. Set $\gamma = \frac{2 \cdot \delta - 1 + \hat{\alpha}}{\hat{\alpha}}$ and let $E = \{x \in \mathbb{R}^2 \mid \|ax\| + \|xb\| \leq \gamma\}$.

Lastly, we subdivide $E \cap F$ into three vertical strips. We fix a parameter $\omega = 0.1$. Let ℓ_1, ℓ_2 be the vertical lines at $\omega\|ab\|$ and $(1 - \omega)\|ab\|$, respectively. Let L be the part of $E \cap F$ to the left of ℓ_1, let M be the part between ℓ_1 and ℓ_2, and let R be the part to the right of ℓ_2. See Figure 2 for a schematic.

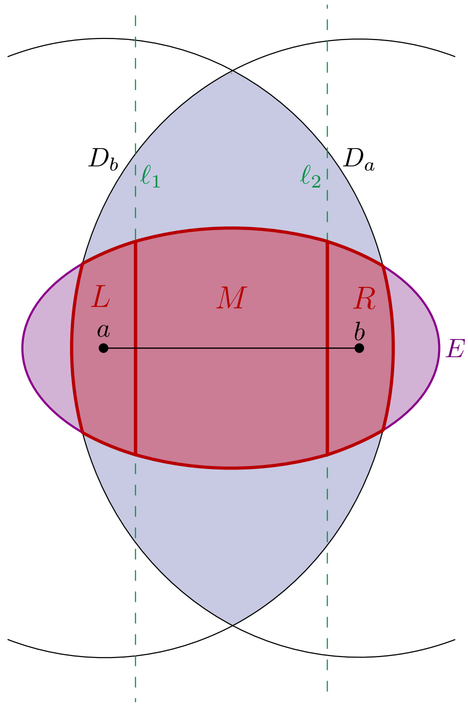

Figure 2 Subdivision of the plane into regions with respect to a longest edge ab of T_{OPT}.

We denote by α the fraction of points in $F \setminus E$, and by β_L, β_M and β_R the fraction of points in L, M and R respectively. Note that $\alpha + \beta_L + \beta_M + \beta_R = 1$. Now we are equipped to consider the next two cases:

▶ **Lemma 3.2.** *Assume*

$$\beta_M \geq \hat{\beta} = \frac{\delta - 0.5}{\delta \cdot \left(1 - \sqrt{1 - d^2(\omega - \omega^2)}\right)}$$

and recall that T_{DIAM} is the larger of the stars at the diameter. Then $|T_{\text{DIAM}}| \geq \delta \cdot |T_{\text{OPT}}|$.

Proof. The main insight in this case is that we can find a tighter bound on T_{OPT} by exploiting that ab is an edge of T_{OPT} and so no other edge of T_{OPT} can cross ab. Let $\overline{M}$ be the region of

F between ℓ_1 and ℓ_2 and above ab. Refer to Figure 3 for illustration. We will argue that every edge with an endpoint in M has length at most $\mathrm{diam}(\overline{M})$.

Let $c_1 = \ell_1 \cap \partial(E) \cap \overline{M}$ and $c_2 = \ell_1 \cap ab$. Disregarding symmetry, it follows from convexity that the longest possible edge starting in M has either c_1 or c_2 as an endpoint. If the endpoint is c_1, then the edge may reach below the line through ab. A maximum length edge starting from c_1 ends at the intersection z of the line through c_1 and b with the boundary of F. If the endpoint is c_2, the length of this edge is $\mathrm{diam}(\overline{M})$. Both cases are shown in Figure 3.

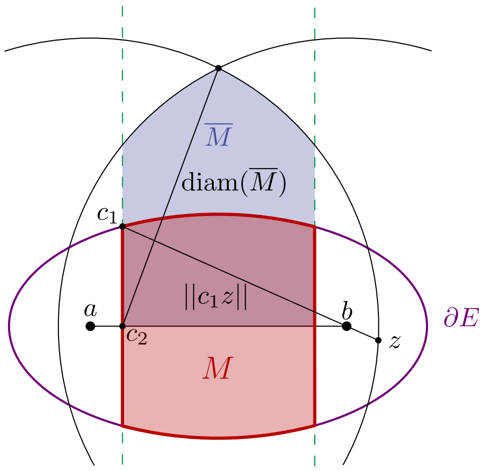

Figure 3 The starting points c_1 and c_2 of longest edges in T_{OPT}.

Now we consider how these lengths change for $d \leq \|ab\| \leq 1$. By basic trigonometry, we can give expressions for $\mathrm{diam}(\overline{M})$ and $\|c_1 z\|$ that only depend on $\|ab\|$:

$$\mathrm{diam}(\overline{M}) = \sqrt{1 - \|ab\|^2(\omega - \omega^2)}$$

$$\|c_1 b\| = \sqrt{((1-\omega)\|ab\|)^2 + \left(\frac{\sqrt{\gamma^2 - \|ab\|^2} \cdot \sqrt{(\gamma/2)^2 - (\|ab\|/2 - \omega\|ab\|)^2}}{\gamma}\right)^2}$$

$$\|c_1 z\| \leq \|c_1 b\| + \frac{\|c_1 b\|(1 - \|ab\|)}{(1-\omega)\|ab\|}$$

The last bound is tight for $\|ab\| = 1$.

When considering $\mathrm{diam}(\overline{M})$ and $\|c_1 z\|$ as functions of $\|ab\|$, by considering the plots (Figure 4) it follows that

$$\begin{aligned} \sqrt{1 - d^2(\omega - \omega^2)} = \mathrm{diam}(\overline{M})_d \geq \quad & \mathrm{diam}(\overline{M}) \quad \geq \mathrm{diam}(\overline{M})_1 \quad \text{and} \\ \|c_1 z\|_d \geq \quad & \|c_1 z\| \quad \geq \|c_1 z\|_1 \,, \end{aligned} \tag{1}$$

where the subscripted versions denote the values at $\|ab\| = d$ and $\|ab\| = 1$, respectively.

With the chosen constants we get $\mathrm{diam}(\overline{M})_1 \geq \|c_1 z\|_d$ (again refer to Figure 4). Thus, $\mathrm{diam}(\overline{M})$ is a valid upper bound for the length of the edge starting in M.

Using (1) and the definition of $\hat{\beta}$ we can bound the size of T_{OPT} and the approximation ratio:

$$\begin{aligned} |T_{\text{OPT}}| &\leq n \cdot (\beta_M \cdot \mathrm{diam}(\overline{M}) + (1 - \beta_M)) \leq n \cdot (\beta_M \cdot \mathrm{diam}(\overline{M})_d + (1 - \beta_M)) \\ &= n \cdot (1 - \beta_M \cdot (1 - \mathrm{diam}(\overline{M})_d)) \end{aligned}$$

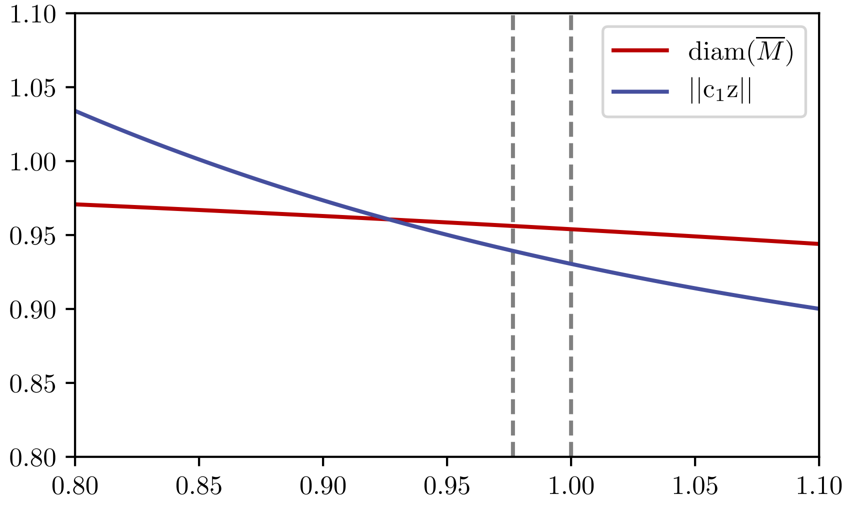

Figure 4 Plot of diam($\overline{M}$) and $\|c_1 z\|$ over the length of ab. The vertical lines are at d and 1.

$$\frac{|T_{\text{DIAM}}|}{|T_{\text{OPT}}|} \geq \frac{0.5n}{(1-\beta_M \cdot (1-\text{diam}(\overline{M})_d))n} \geq \frac{0.5}{1-\hat{\beta}\cdot(1-\text{diam}(\overline{M})_d)} = \delta. \quad \blacktriangleleft$$

In the next case, we assume that $\alpha \geq \hat{\alpha}$ and also show that there is a good star.

▶ **Lemma 3.3.** *If*

$$\alpha \geq \hat{\alpha} = 1 - \frac{2\delta + \hat{\beta}(1-\omega)}{2-3\omega},$$

then $\max\{|S_a|,|S_b|\} \geq \delta \cdot |T_{\text{OPT}}|$.

Proof. As before we bound $\max\{|S_a|,|S_b|\} \geq \frac{1}{2}(|S_a|+|S_b|)$. This time we get:

$$\begin{aligned}|S_a|+|S_b| &\geq n(\alpha\cdot\gamma + (1-\alpha)\|ab\|)\\ &= n(\|ab\| + \alpha(\gamma - \|ab\|))\\ &\geq n\cdot(\|ab\| + \alpha(\gamma-1)).\end{aligned}$$

With Observation 2.2 ($|T_{\text{OPT}}| \leq \|ab\|n$) we get

$$\frac{\max\{|S_a|,|S_b|\}}{|T_{\text{OPT}}|} \geq \frac{n(\|ab\|+\alpha(\gamma-1))}{2\|ab\|n} \geq \frac{1}{2} + \frac{\alpha}{2}(\gamma-1) \geq \frac{1}{2} + \frac{\hat{\alpha}}{2}(\gamma-1) = \delta. \quad \blacktriangleleft$$

Last but not least we consider the case where α and β_M are both small. Intuitively, this means that almost all points are located left or right in E.

▶ **Lemma 3.4.** *If $\alpha < \hat{\alpha}$ and $\beta_M < \hat{\beta}$, then there is a tree which gives a δ-approximation.*

Proof. In this case we do not use a star but trees B_{ab}, B_{ba} of diameter at most five. We will describe the structure B_{ab} with regard to a. The structure B_{ba} with regard to b is symmetric. See Figure 5 for an example of the construction.

We start by connecting all points in R to a (blue edges). This gives a star with length at least $\beta_R(1-\omega)\|ab\|$. The edges of this star subdivide L into wedges. We define the upper wedge to be the region above both the highest edge and the x-axis. The lowest wedge is defined accordingly. For each such wedge W (except the last) we take the lower point of R defining W and connect it to all points in $L \cap W$. The lowest point in R also connects to the points in the lowest wedge of L (green edges). Each of these new edges has weight at least $(1-2\omega)\|ab\|$.

Now we connect the points in M. The edges of the tree so far subdivide M into quadrilateral regions, which are defined by two edges of the tree. We again want to connect the vertices in such a subregion in a star like fashion. From the interior of such a subregion at least one boundary edge between a point from L and a point from R is fully visible. For every subregion we pick the better of the two stars centered at the two endpoints of such an edge (red edges). By Lemma 2.1 this yields a total additional weight of at least $0.5 \cdot \beta_M(1-2\omega)\|ab\|$.

Recall that $\alpha + \beta_L + \beta_M + \beta_R = 1$. By bounding the maximum by the average, we get

$$\begin{aligned}\max\{|B_{ab}|, |B_{ba}|\} &\geq \frac{n\|ab\|}{2}((\beta_L+\beta_R)(2-3\omega)+\beta_M(1-2\omega)) \\ &= \frac{n\|ab\|}{2}((1-\alpha)(2-3\omega)-\beta_M(1-\omega)) \\ &\geq \frac{n\|ab\|}{2}((1-\hat{\alpha})(2-3\omega)-\hat{\beta}(1-\omega)). \\ \frac{\max\{|B_{ab}|, |B_{ba}|\}}{|T_{\text{OPT}}|} &\geq \frac{\frac{n\|ab\|}{2}((1-\hat{\alpha})(2-3\omega)-\hat{\beta}(1-\omega))}{\|ab\|n} = \delta.\end{aligned}$$ ◀

▶ **Theorem 3.5.** *A $\delta = 0.512$-approximation for the longest noncrossing Euclidean spanning tree can be computed in polynomial time.*

Proof. We compute S_p for each $p \in P$. Additionally, for each pair a, b with $\|ab\| > d = 1/(2\delta)$, we compute B_{ab} and B_{ba}. Let T_{ALG} be the largest of these structures.

By the exhaustive case distinction in Lemmas 3.1 to 3.4, for the pair a, b which leads to the longest edge in T_{OPT} this leads to a $\delta = 0.512$-approximation. ◀

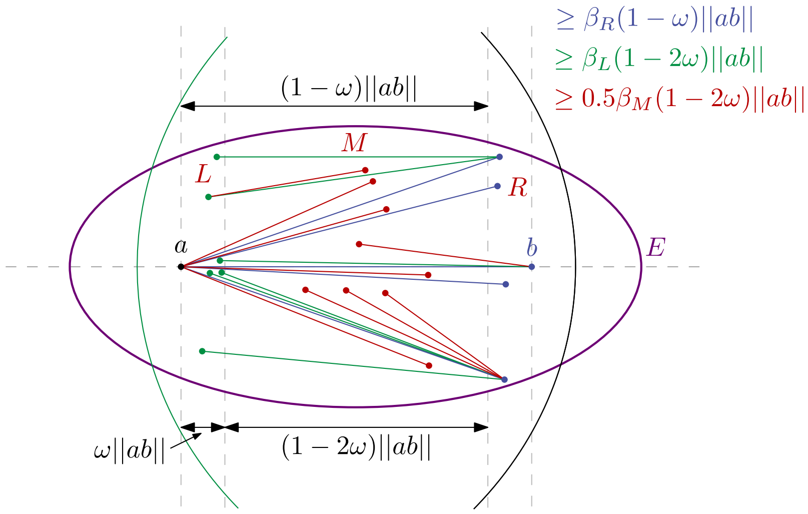

Figure 5 Structure B_{ab}. The edges of each stage of the construction have a different color.

4 Improved approximation factor for thin point sets

In this section we present stronger bounds for thin point sets. Given $\sigma > 0$, we say that P is (at most) *σ-thick* if there exists a diameter of P such that all points in P have distance at most σ from this diameter. Moreover, let T_{CR} be the longest (possibly crossing) tree on P.

▶ **Theorem 4.1.** *There is a polynomial-time algorithm that, given a σ-thick point set P with $\sigma \le \frac{1}{3}$, constructs a planar spanning tree T_{ALG} with*

$$|T_{\text{ALG}}| \ge f(\sigma) \cdot |T_{\text{CR}}| \ge f(\sigma) \cdot |T_{\text{OPT}}|,$$

where $f(\sigma)$ is given by

$$f(\sigma) = \frac{2}{3} \cdot \sqrt{\frac{1 + 4\sigma^2}{5 - 4\sqrt{1-\sigma^2} + 4\sigma^2}}.$$

Inspecting the function $f(\sigma)$, we get, e.g., $f(0.3) \ge 0.516$ and $f(0.1) \ge 0.636$. Also, in the limit $d \to 0$ we get $f(\sigma) \to 2/3$. The constant $2/3$ here is tight: There exist perturbations of point sets lying on a segment for which the longest planar trees have length arbitrarily close to $2/3$ of the length of the longest (possibly crossing) tree (see Figure 6).

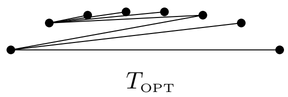

Figure 6 A thin convex set consisting of $n+1$ points with equally spaced x-coordinates $0, 1, \dots, n$. For large n, the length of any longest planar tree is $1 + 2 + \cdots + n \approx \frac{1}{2}n^2$, whereas the length of the longest (possibly crossing) tree is roughly $2 \cdot (n/2 + \cdots + n) \approx \frac{3}{4}n^2$. Thus, as $n \to \infty$, we obtain $|T_{\text{OPT}}|/|T_{\text{CR}}| \to \frac{2}{3}$.

Proof. [of Theorem 4.1] Fix P and $\sigma \le \frac{1}{3}$. Denote the relevant diameter of P by pq, and without loss of generality place it as $p = (0,0)$, $q = (1,0)$. Divide $P \setminus \{p, q\}$ by a vertical line ℓ into a set P_p of points closer to p and a set P_q of points closer to q (see Figure 7(a)).

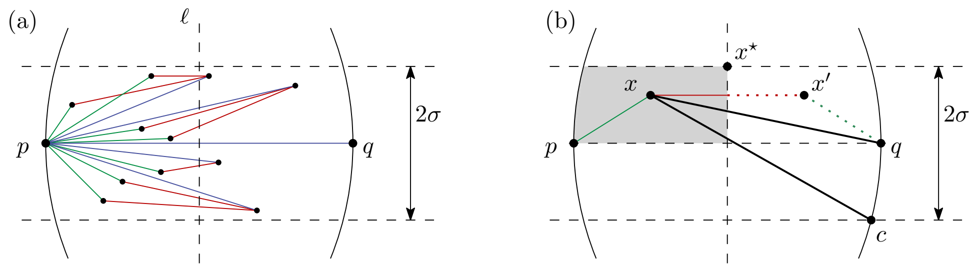

Figure 7 (a) We star the points in the right half from p (blue) and then either star the points in the left half from p too (yielding S_p, blue and green) or connect them to points in the right half (yielding T_{pq}, blue and red). (b) With the shown notation we have $f(\sigma) = 2\|x^\star q\|/(3\|x^\star c\|)$.

We construct a tree T_{pq} as follows: Connect p to all points in $P_q \cup \{q\}$. This splits P_p into wedges with apex p. For each wedge, connect all its points in P_p to the endpoint of its upper side in P_q (use the lower side for the uppermost wedge). Note that T_{pq} is planar. We construct T_{qp} in a symmetric fashion and set T_{ALG} to be the longest of T_{pq}, T_{qp}, S_p, S_q.

Next we argue that T_{ALG} satisfies $|T_{\text{ALG}}| \ge f(\sigma) \cdot |T_{\text{CR}}|$. It suffices to show

$$\frac{|S_p| + 2|T_{pq}| + 2|T_{qp}| + |S_q|}{6} \ge f(\sigma) \cdot |T_{\text{CR}}|.$$

Note that all four trees on the left-hand side include edge pq and since pq is a diameter, we can without loss of generality assume that T_{CR} contains it too. Direct all other edges of those five trees towards pq. Fix a point $x \in P \setminus \{p, q\}$ and let $x_{\text{CR}}, x_{pq}, x_{qp}$ be the other endpoints of the edges pointing from x in $T_{\text{CR}}, T_{pq}, T_{qp}$, respectively. (Note that in S_p all edges point towards p, similarly for S_q and q.) It suffices to prove that

$$\frac{\|xp\| + 2\|xx_{pq}\| + 2\|xx_{qp}\| + \|xq\|}{6 \cdot \|xx_{\text{CR}}\|} \geq f(\sigma)$$

Without loss of generality, suppose that x belongs to P_p and lies above pq. Let x' be the reflection of x about ℓ and c the furthest point from x within the intersection of unit disks centered at p and q. Using the triangle inequality in $\triangle pxx'$, the left-hand side is at least

$$\frac{\|xp\| + \|xx'\| + 2\|xq\| + \|xq\|}{6\|xc\|} \geq \frac{\|px'\| + 3\|xq\|}{6\|xc\|} = \frac{2}{3} \cdot \frac{\|xq\|}{\|xc\|}.$$

Since $\sigma \leq \frac{1}{3}$, the ratio $\|xq\|/\|xc\|$ is minimized when $x = x^\star$ lies on ℓ with distance σ from pq (see Figure 7(b)). Since $\|pc\| = 1$, using the Pythagorean theorem, we easily compute

$$\|x^\star c\| = \sqrt{\left(\sqrt{1-\sigma^2} - 1/2\right)^2 + (2\sigma)^2} \quad \text{and} \quad \|x^\star q\| = \sqrt{(1/2)^2 + \sigma^2},$$

which matches the desired expression $f(\sigma)$. ◀

5 Conclusion

We showed that it is possible to significantly increase the approximation factor from 0.503 to 0.512 in the general case and even towards 2/3, when the point set is σ-thick for $\sigma \to 0$.

The improvement in the approximation factor relies in one case on the planarity of the optimum tree. Without further analysis this does not yield a better approximation factor with regard to the longest crossing tree.

In future work, we aim to further reduce the running time and the approximation factor. For the latter we plan to build on the fact that T_{OPT} is noncrossing, which can lead to further advances. The last open problem would be to settle the question of NP-hardness.

References

1 Noga Alon, Sridhar Rajagopalan, and Subhash Suri. Long non-crossing configurations in the plane. *Fundam. Inform.*, 22(4):385–394, 1995. doi:10.3233/FI-1995-2245.

2 Ahmad Biniaz, Prosenjit Bose, Kimberly Crosbie, Jean-Lou De Carufel, David Eppstein, Anil Maheshwari, and Michiel Smid. Maximum Plane Trees in Multipartite Geometric Graphs. *Algorithmica*, 81(4):1512–1534, 2019. URL: http://link.springer.com/10.1007/s00453-018-0482-x, doi:10.1007/s00453-018-0482-x.

3 Adrian Dumitrescu and Csaba D. Tóth. Long non-crossing configurations in the plane. *Discrete Comput. Geom.*, 44(4):727–752, 2010. doi:10.1007/s00454-010-9277-9.